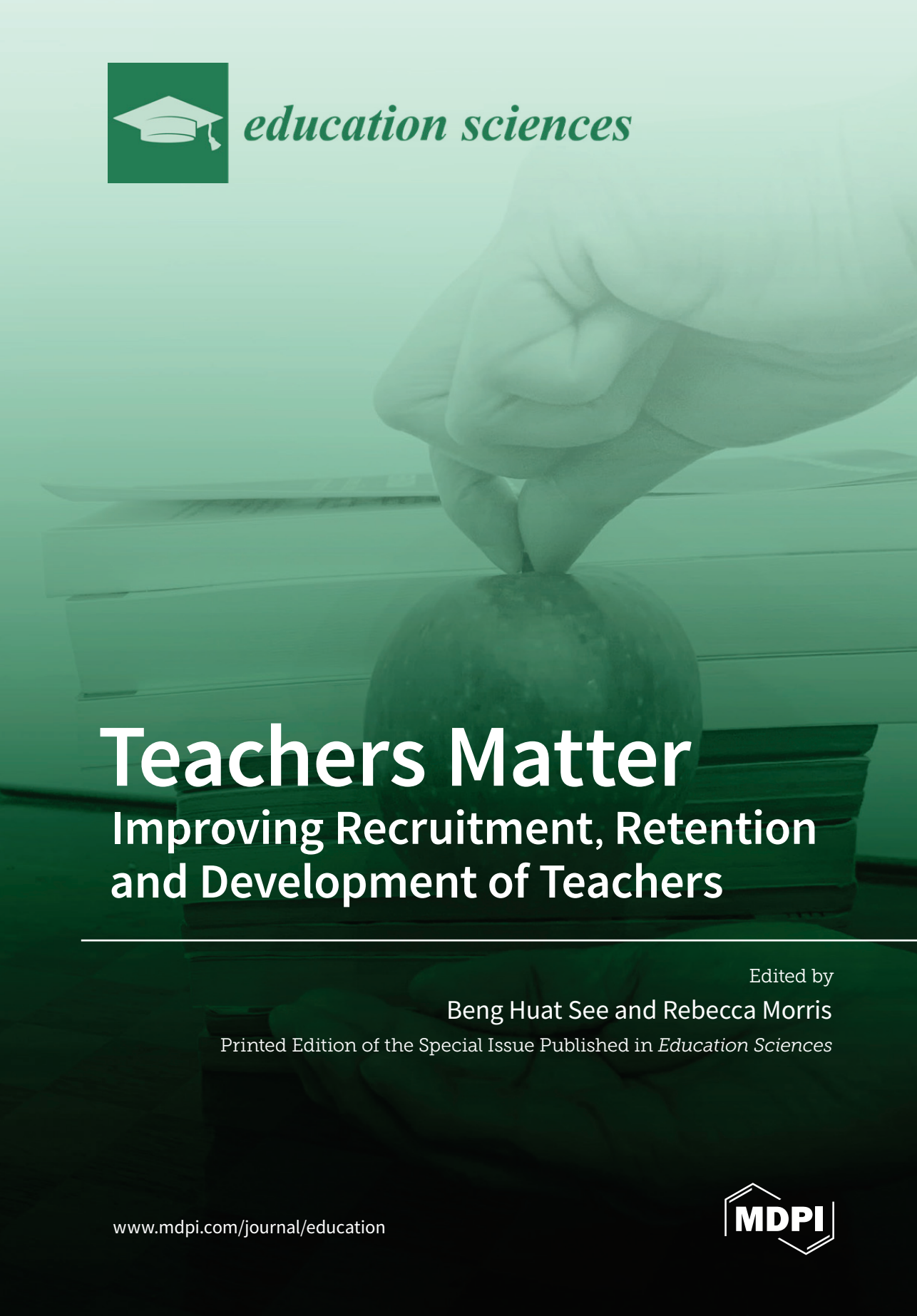


education sciences

Teachers Matter

Improving Recruitment, Retention and Development of Teachers

Edited by

Beng Huat See and Rebecca Morris

Printed Edition of the Special Issue Published in *Education Sciences*

www.mdpi.com/journal/education

MDPI

Teachers Matter—Improving Recruitment, Retention and Development of Teachers

Teachers Matter—Improving Recruitment, Retention and Development of Teachers

Editors

Beng Huat See
Rebecca Morris

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Beng Huat See
School of Education,
Durham University,
Durham DH1 1TA, UK

Rebecca Morris
Department for Education Studies,
University of Warwick,
Warwick, UK

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Education Sciences* (ISSN 2227-7102) (available at: https://www.mdpi.com/journal/education/special_issues/teacher_education).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-5445-7 (Hbk)
ISBN 978-3-0365-5446-4 (PDF)

Cover image courtesy of Beng Hua See

Contents

About the Editors

Beng Huat See

Beng Huat See is a Professor of Education Research at the Durham University Evidence Centre for Education. Her work largely involves reviews and synthesis of evidence in education, evaluation of large-scale randomised controlled trials and re-analysis of secondary data. Her research interests are understanding the determinants of teacher demand and supply, enhancing critical thinking in young people, parental involvement, teacher development and teacher effectiveness, and improving the learning and wider outcomes of children from disadvantaged backgrounds.

Rebecca Morris

Dr Rebecca Morris is Associate Professor in the Department of Education Studies at the Warwick University. Her research interests include: education and social policy, inequalities in education, the teaching workforce, widening participation and literacy.

education sciences

MDPI

Editorial

Editorial Introduction to the Special Issue "Teachers Matter—Improving Recruitment, Retention and Development of Teachers"

Beng Huat See [1,*] and Rebecca Morris [2]

[1] School of Education, Durham University, Durham DH1 1TA, UK
[2] Department of Education Studies, Warwick University, Coventry CV4 7AL, UK; rebecca.e.morris@warwick.ac.uk
* Correspondence: b.h.see@durham.ac.uk

Citation: See, B.H.; Morris, R. Editorial Introduction to the Special Issue "Teachers Matter—Improving Recruitment, Retention and Development of Teachers". *Educ. Sci.* **2021**, *11*, 528. https://doi.org/10.3390/educsci11090528

Received: 3 September 2021
Accepted: 8 September 2021
Published: 9 September 2021

This Special Issue of Education Sciences on the theme of Education Matters was commissioned in recognition of the important role that teachers play in the development of students' learning and wider outcomes. Ensuring an adequate supply of appropriate, suitably qualified teachers is a major policy concern facing many education systems around the world. Huge investments have been made by governments to overcome shortages of teachers, in regions or subjects, through offers of financial incentives, raising teacher salaries, improving teachers' working conditions, providing support through professional development, and the induction and mentoring of novice teachers. Despite these efforts, challenges to improve the quantity and quality of the teaching workforce remain.

In this issue we have an impressive collection of papers from around the world, all of which attempt to help readers understand more about researching and addressing these challenges. Contributions came from Estonia, England, the US, Pakistan, Italy, China, Turkey, South Africa and South Korea. These provide readers with an international perspective of the challenges and opportunities in research in this field.

Ingersoll et al.'s longitudinal analysis of the School and Staffing Survey data in the US shows that the teaching population has transformed over the last three decades. Teachers are now likely to be older but less experienced, more likely to be female and from academically prestigious universities, and they are more ethnically diverse, than thirty years ago. These demographic transformations indicate that the teaching population is dynamic and constantly growing and being replaced. This suggests that the movement of people in and out of teaching is not unusual and challenges the view that teacher turnover is a problem to be solved. We think that such transformations should be expected as a sign of a healthy profession and managed accordingly.

Carver's paper also suggests that teacher turnover is to be expected and is not always the result of dissatisfaction with the profession. He argues for alternative measures that can reflect changes in teachers' career intentions and risks of attrition. Adapting questions from the OECD's TALIS and the US' Beginning Teacher Longitudinal Study he illustrates how these measures can provide a better understanding of teacher attrition.

See et al.'s comprehensive systematic review of international evidence on how to increase teacher supply suggests that monetary inducements may be effective in getting people into teaching initially. However, they do not appear to have a lasting impact. Those attracted by financial incentives are less likely to stay in the profession, and financial inducements to retain teachers are only effective as long as these incentives are available. Once removed, attrition rates return to their previous levels. Money, by itself, does not seem to be the solution, but the evidence for the impact of other approaches is minimal and is usually not robust enough to inform strong policy recommendations.

Siddiqui and Shaukat's survey of over 1000 teachers in the Punjab area of Pakistan found that salary and teachers' workload are also not important determining factors for

teacher mobility. Teachers who moved between schools generally reported dissatisfaction with the working environment and the conditions of the school as the reasons, rather than better pay. Again, this suggests that policies to improve retention that are based on financial incentives may not be effective. In the long-term, continuing professional development (CPD) and support for novice teachers and better working conditions may be more useful.

The paper by Booth et al. cautions that the kinds of professional development offered to teachers should be tailored to their needs. They argue, for example, that mid-career teachers have different needs to those of early career teachers and as a result, professional development offered to teachers in schools should be more personalised and differentiated. They also found that teachers who are satisfied with their working environment are more likely to report positive experiences of professional development. While the correlational analysis is unable to show the direction of the causation, it is possible that positive experiences of professional development may help teachers cope with their working conditions better. The study also found that mid-career teachers are more likely to report increased barriers to professional development than early career teachers. This has important implications for initiatives to retain teachers in the profession. It suggests that we need to consider the needs of mid-career teachers and address the barriers to their professional development to support retention.

Two studies in this Special Issue describe how supporting teachers through a professional development programme can help develop teachers' competency and self-efficacy, which in turn, leads to improvement in student learning outcomes. Wu's randomised controlled trial, which tests the impact of training teachers in China on the delivery of a dialogic critical thinking curriculum, shows that children who were taught by teachers who had been trained did better on a critical thinking skills test than children of teachers who were not trained. Ibbotson and See's paper evaluates a collaborative partnership approach to training non-specialist teachers, using a Kodály-inspired pedagogy to teaching music. Pre-post comparisons show positive changes in teachers' pedagogical skills, self-efficacy and competence, and in children's self-confidence and their disposition for learning.

The paper by Baccaglini-Frank and colleagues demonstrated, through design-based action research, how the use of a battery-operated minirobot (the Geombot) for training teachers can enhance teachers' understanding of their own perspectives to the teaching of geometry. The authors use their findings to point towards emerging new learning opportunities for both teachers and students as a result of the implementation of the Geombot. They suggest that further research exploring development opportunities such as this would be valuable for teachers and for promoting dialogue across the academic and practitioner communities.

While developing teacher competency is beneficial, students' perception of teachers' competency is also important. The paper by Shin and Shim reports that students' perceptions of their teachers' professional competence in South Korea are associated with students' academic performance. These findings suggest that developing teachers' professional competence may have a positive influence on students' attainment and engagement, although further work is needed to understand if there is a direct causal relationship there.

Nakidien, Singh and Sayed's empirical work reminds us that professional development of teachers is not just about enhancing skills and competency. Professional development is also crucial in preparing teachers for curriculum reforms and new challenges. In their paper, they highlight the need to prepare teachers for the post-apartheid curriculum in South Africa, which emphasises equality, inclusivity, and recognition of all cultures and religions. This is timely and relevant not only in South Africa but also elsewhere in the world. There is an increasing need for recognition of diversity and inclusion in education. The recent call for decolonising the curriculum in the UK means teachers need support in order to effectively deliver an amended curriculum. Teacher training and professional development, however, is an area too often overlooked in the literature on decolonisation and would be a fruitful area for new research.

The paper by Leibur and colleagues explores the experiences and perceptions of teachers engaging in a professional development course in Estonia. The findings of this study foreground the importance of collaboration and support in the implementation of effective professional development. Without this, teachers may experience a range of barriers that inhibit their ability to successfully apply, participate in and complete professional development opportunities.

When the coronavirus pandemic happened, schools had to rely on new ways of delivering lessons. The paper here by Perry, Findon and Cordingley is timely. It is a review of the evidence on the efficacy of blended and remote approaches to teacher education and the limitation and affordances of these approaches. It reminds us of the importance of teacher training in preparing teachers for the digital age and also highlights some of the challenges associated with this area of training and development.

Aslantas' paper examines the challenges of measuring teacher quality or effectiveness by looking closely at the widespread use of value-added models. The study confirmed that most students' performance at secondary schools in Turkey can be predicted by their primary school grades. Therefore, systems that assess teachers on the basis of their students' performance are misguided. The contribution that teachers make to differential student outcomes is small after other key factors have been considered.

All of these studies show that developing teachers to help them meet the needs and new demands in teaching is useful in its own right. Moreover, there is some evidence which points towards potential benefits for their students too. We need to be careful not to use promotion or pay rises as an incentive for professional development, as is sometimes used in some countries. And we certainly should not use teachers' professional development status to differentiate "effective" from "ineffective" teachers. Teachers should be routinely kept abreast with changing curriculum and policy reforms and should be supported to pursue areas of interest and expertise which will enhance their role within the profession. Professional development, as long as it is based on sound evidence, can also potentially contribute to the retention of teachers in teaching, but teachers need to be properly supported to participate in and utilise these kinds of development opportunities.

The papers within this Special Issue all signal the high value that teachers hold within our education systems while simultaneously pointing towards the barriers and challenges that policymakers and practitioners face in maintaining an effective workforce. Yet, there is still so much more that we need to know about the issues of teacher recruitment, retention and development. We believe high quality research, such as the work reported here, is a step forward to a better understanding of researching and addressing the challenges and barriers faced in supporting our valuable teacher workforce.

Author Contributions: B.H.S. and R.M. contributed to writing the editorial. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

education sciences

MDPI

Article

The Demographic Transformation of the Teaching Force in the United States †

Richard Ingersoll [1,*], Elizabeth Merrill [2], Daniel Stuckey [3], Gregory Collins [4] and Brandon Harrison [1]

[1] Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104, USA; bjh@upenn.edu
[2] Research Alliance, New York City Schools, New York University, New York, NY 10012, USA; lisa.merrill@nyu.edu
[3] Relay Graduate School of Education, New York, NY 10004, USA; dstuckey@relay.edu
[4] Department of Accounting, Business, Economics, and Finance, Muhlenberg College, Allentown, PA 18104, USA; gcollins@muhlenberg.edu
* Correspondence: rmi@upenn.edu
† This article draws from, expands, and updates, earlier reports of this ongoing study (Ingersoll and Merrill 2017; Ingersoll, Merrill & Stuckey, 2018; Ingersoll, Merrill, Stuckey, & Collins, (2018)).

Abstract: This article summarizes the results of an exploratory research project that investigated what demographic trends and changes have, or have not, occurred in the elementary and secondary teaching force in the U.S. over the past three decades, from 1987 to 2018. Our main data source was the Schools and Staffing Survey and its successor, the National Teacher Principal Survey, collectively the largest and most comprehensive source of data on teachers available in the U.S. These surveys are conducted by the National Center for Education Statistics (NCES), the statistical arm of the U.S. Department of Education. The results show that the teaching force has been, and is, greatly changing; yet, even the most dramatic trends appear to have been little noticed or understood by researchers, policy makers, and the public. This article summarizes seven of the most prominent trends and changes that we found. The U.S. teaching force is: larger; older; less experienced; more female; more diverse, by race/ethnicity; consistent in academic ability; unstable. For each of the trends, we explore two broad questions: 1. What are the reasons for and sources of the trend? 2. What are the implications and consequences of the trend?

Keywords: teachers; teaching force; demographic characteristics

Citation: Ingersoll, R.; Merrill, E.; Stuckey, D.; Collins, G.; Harrison, B. The Demographic Transformation of the Teaching Force in the United States. *Educ. Sci.* **2021**, *11*, 234. https://doi.org/10.3390/educsci11050234

Academic Editor: Beng Huat See and Rebecca Morris

Received: 15 February 2021
Accepted: 3 May 2021
Published: 14 May 2021

1. Introduction

Has the elementary and secondary teaching force in the United States changed in recent decades? And, if so, how? Have the types and kinds of individuals going into teaching changed? Have the demographic characteristics of those working in classrooms altered? To answer these questions, we embarked on an exploratory research project to try to discover what trends and changes have, or have not, occurred in the teaching force over recent decades. We were surprised by what we found. We discovered that the teaching force has been, and is, greatly changing; yet, even the most dramatic trends appear to have been little noticed by researchers, policy makers, and the public. To explore these questions, we undertook analyses of nationally representative data on teachers in the U.S. collected over the three-decade period from 1987 to 2018. In this article, we summarize seven of the most prominent trends and changes; we found the U.S. teaching force to be:

1. Larger;
2. Older;
3. Less experienced;
4. More female;
5. More diverse, by race/ethnicity;
6. Consistent in academic ability;

7. Unstable.

For each of the trends, two significant questions arise:

1. What are the reasons for and sources of the trend?
2. What are the implications and consequences of the trend?

We will offer some possible answers to these questions. But our objective is not to provide an in-depth investigation of any of the seven trends. We do not seek to provide a full account of the sources behind these trends, nor do we seek to assess the many possible benefits and costs, advantages, and disadvantages of each trend. That is, our intent here is not to arrive at closure on the above two questions; that would require far more extensive analyses.

Our intent is to provide a broad and system-level "bird's eye" portrait of some key recent developments in the U.S. teaching force. Rather than explanatory and evaluative, our work here is largely exploratory and suggestive. In short, we ask more questions than we are able to answer. We believe that further research is both necessary and warranted to rectify that.

Moreover, we do not attempt to provide an exhaustive review of the research and policy literatures that pertain to each of the trends we highlight. Given our wide purview to describe seven very different trends, and to raise questions about both the sources and consequences of each trend, our findings touch upon a very wide range of different theories, empirical studies, policies and reforms regarding teachers and education. While we attempt to make a number of key connections, addressing the full range of relevant literatures and reforms is beyond the scope of this article.

Finally, our purpose is neither normative nor prescriptive; while we personally feel strongly about some of these trends and if and how they might be addressed, our purpose here is descriptive and analytic. That is, our objective is to describe and analyze seven major demographic trends that have happened to the U.S. teaching force in recent decades.

This article builds on, updates and expands earlier reports on our ongoing study of demographic changes in the teaching force. The first report of our study, released in 2012, presented data up to 2007–2008—the most current data then available. Since then, we have released several updated reports of our study as newer cycles of the data became available [1–3]. This new 2021 article updates our prior work by presenting the results of our analyses of newly released national data, and also expands on our prior work by introducing new types of data relevant to several of the trends. A key finding of this article is that the most recent data document that the trends we originally uncovered continue to hold true—even after 3 decades.

2. Materials and Methods

Our study entailed secondary analysis of large-scale nationally representative data on teachers in the United States. Our primary source of data is the Schools and Staffing Survey (SASS), which is the largest and most comprehensive source of data on teachers available in the U.S. These data are collected by the National Center for Education Statistics (NCES), the statistical arm of the U.S. Department of Education. The NCES has administered nine cycles of this survey over a 30 year period—1987–1988, 1990–1991, 1993–1994, 1999–2000, 2003–2004, 2007–2008, 2011–2012, 2015–2016 and 2017–2018. The most recent two cycles, administered in the periods 2015–2016 and 2017–2018, were renamed the National Teacher Principal Survey (NTPS).

In each cycle, NCES administers questionnaires to a nationally representative sample of 40,000 to 50,000 teachers, and 9000 to 11,000 school-level administrators, collecting an unusually rich array of information on teachers, their students, and their schools. The data represent all teachers for grades prekindergarten through 12, part time and full time, and from all types of schools, including public, charter, and private. This analysis uses data from all cycles of the SASS/NTPS available, over the three-decade period from 1987 to 2018, but, in particular, focuses on the most recent data cycle—the 2017–2018 NTPS. For information

on the SASS/NTPS, see references [4,5]. With the addition of the 2017–2018 NTPS data, we have been able to update almost all of our findings for this new 2021 edition of our study.

In addition, all those teachers in the SASS sample who departed from their schools in the year subsequent to the administration of the initial SASS survey questionnaire were contacted to obtain information on their departures. This nationally representative supplemental sample—the Teacher Follow-Up Survey (TFS)—contains approximately 7000 teachers. The TFS captures all teacher departures from schools, including both major components of total turnover—migration (teachers who move between local school districts and schools) and attrition (those who leave teaching altogether). We analyze TFS data items on the rates, variations and reasons for, teacher turnover. The analysis uses data from all seven cycles of the TFS available, but in particular from the most recent TFS, administered in the period 2012–2013, which only included public school teachers (for more information on the 2012–2013 TFS, see references [6].

We took advantage of both the depth and duration of the SASS/NTPS/TFS data to explore what changes have taken place in the teaching force and teaching occupation over the three decades from 1987 to 2018. We primarily used basic statistical analytic techniques to generate descriptive data estimates. This analysis uses data weighted to compensate for the over- and under-sampling of the complex stratified survey design. Each observation is weighted by the inverse of its probability of selection in order to obtain unbiased estimates of the national population of schools and teachers in the U.S. in the year of the survey.

3. Results

3.1. Trend 1: Larger

The elementary and secondary teaching force in the U.S. has increased in size—a trend we refer to as "ballooning." The U.S. Census Bureau indicates that PreK-12 teachers form one of the largest occupational groups in the nation [7], and the teaching force is growing even larger. Growth in the numbers of students and teachers is not new. The numbers of both students and teachers grew throughout the 20th century, and the rate of growth for both groups began to soar in the late 1940s with the post–World War II baby boom. Student enrollment peaked by 1970 and then declined until the mid-1980s. During this period, the numbers of teachers also peaked, and then leveled off. In the mid-1980s, elementary and secondary student enrollment again began to grow. Since then, the teaching force has also been increasing in size (see Figure 1).

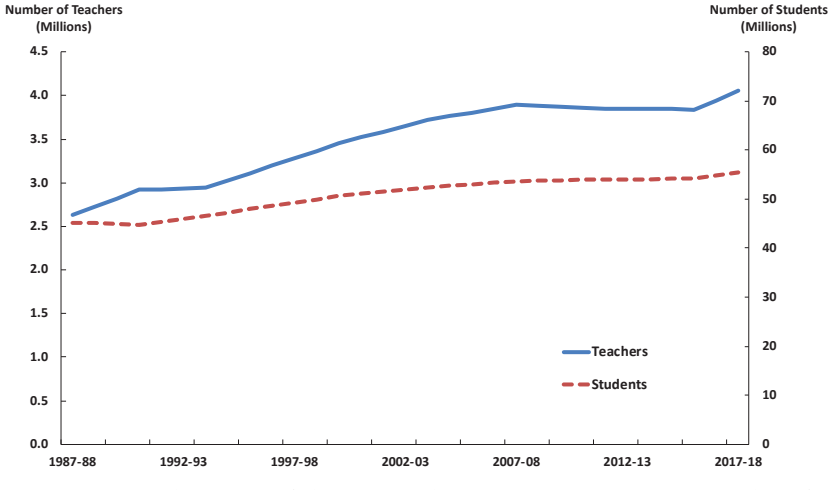


Figure 1. Trends in the Number of Elementary and Secondary School Teachers and Students, from 1987–1988 to 2017–2018.

The rate of these increases has not matched the magnitude of the baby boom years—with one large difference. In recent decades, the rate of increase for teachers has far outpaced the rate of increase for students—that is, the number of teachers has been going up far faster than the number of students.

As illustrated in Figure 1, from the late 1980s to 2008, the teaching force increased at a more rapid pace than the student population. Then during the economic downturn after 2008, growth in the teaching force leveled off. Between 2007–2008 and 2015–2016, while the student population slightly increased (by 1 percent), the teaching force slightly decreased (by approximately 1.6 percent). It is unclear how much of this decrease in teachers was due to layoffs or to hiring freezes combined with teacher attrition. After 2016, growth picked up again.

As the top of Figure 2 summarizes, from 1987–1988 to 2017–2018 total K-12 student enrollment in all U.S. schools went up by 22 percent. During the same period, the teaching force employed in schools increased at over double that rate, by 54 percent. This resulted in a sharp decrease in the overall pupil–teacher ratio in schools.

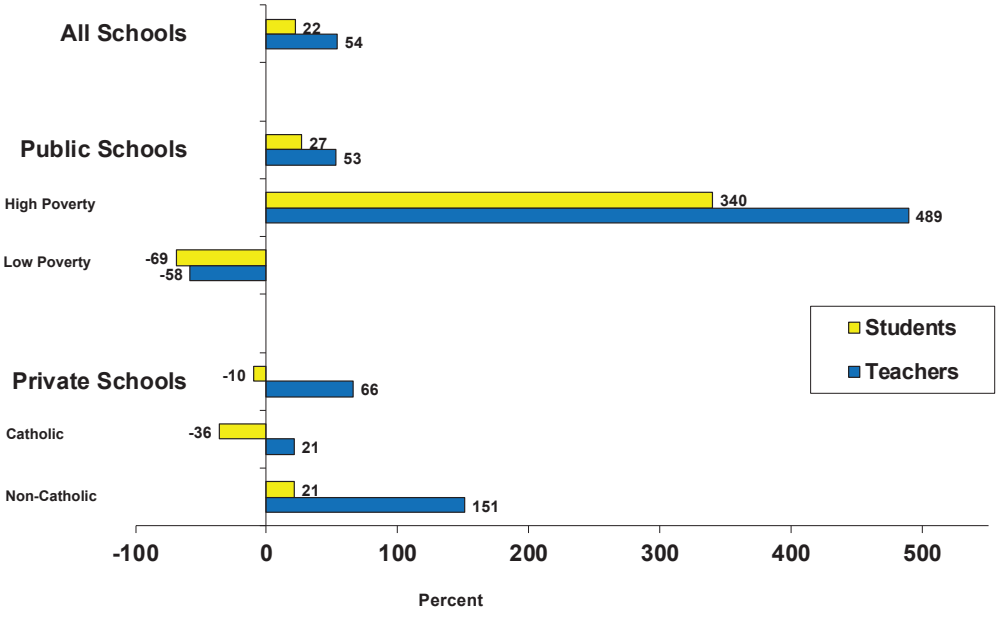


Figure 2. Percent Change in Students and Teachers, by School Type, from 1987–1988 to 2017–2018.

What accounts for this ballooning of the teaching force between the late 1980s and 2018?

As shown in Figure 2, the growth in the number of both students and teachers has not been evenly distributed across schools. In the U.S., approximately one-fifth of elementary and secondary schools are privately owned and operated institutions. Approximately one-third of private schools are non-sectarian and the remainder are operated by religious groups. Among the latter, Catholic schools comprise the largest group. Interestingly, the number of teachers employed in private schools as a whole has increased at a faster rate than in public schools relative to the student population (bottom of Figure 2). But, surprisingly, while the total number of teachers in private schools has increased, the number of students in private schools has decreased. This also varies by type of private school. The three decades from 1987 to 2018 saw a 21 percent increase in the number of teachers employed in Catholic schools, while there was a 27 percent decrease in the number of Catholic schools and a 36 percent decrease in the total number of students enrolled in them. On the other hand, there were increases in the total number of schools,

students, and teachers in the non-Catholic religious private school sector and in the non-sectarian private school sector. The overall result has been a sharp decrease in the average pupil–teacher ratio and average class sizes in private schools as a whole, which were already lower than in public schools. However, this increase in private sector teachers and reduction in the student-to-teacher load in private schools does not account for much of the overall ballooning because private schools account only for a small portion of the student population (approximately 8 percent) and of the teaching force (approximately 13 percent).

Growth in the number of teachers was also not even among public schools. The number of students from poverty-level families, and hence who qualify for the National School Lunch Program (NSLP) has dramatically increased over the past three decades (middle of Figure 2). (The National School Lunch Program is a federal program that offers free and reduced-price school lunches to school-age children from poverty-level families and it is widely used as a measure of the poverty level of students in schools in the U.S.). It is unclear whether this is due to increases in overall poverty, increases in poverty among families with school-age children, or changes in either student applications or the eligibility requirements for the NSLP. But the result is that there have been large increases in the number of high-poverty public schools (defined as those in which three-quarters or more of the students were eligible for the lunch program) and in the numbers of students and teachers in such schools. Over half of the total increase in the number of teachers in public schools during this period took place in high-poverty schools. This group went from 10 percent of all public schools in the period 1987–1988 to over one-third of all public schools by 2017–2018. On the other hand, there have been large decreases in the number of lower-poverty public schools (defined as those in which less than 15% of the students were eligible for the lunch program) and in the numbers of students and teachers in such schools.

Charter schools are publicly funded schools that are granted a greater degree of autonomy from governmental regulations than traditional public schools. They are a relatively new model of schooling in the U.S that have dramatically grown in number over the past couple of decades. But they account for only a small portion of the ballooning of the teaching force because they represent a small segment of all schools (approximately 7.9 percent in the period 2017–2018) and of the teaching force (approximately 5.8 percent).

Another possible explanation for the ballooning is that a reduction in teachers' workloads—class sizes, hours worked, or classes taught per day—necessitated an increase in the number of teachers employed. For instance, some states, such as California, implemented class size reduction reforms to great fanfare, leading to a demand for more teachers.

On close examination, this explanation does account for part of the ballooning of teachers, but not as much as one might expect. Elementary-level class size did decrease by 18 percent during this period (late 1980s to 2017–2018), from an average of 25.6 to 21 students per general elementary school classroom. Accordingly, the number of general elementary school teachers increased, and because elementary teachers comprise the largest field in teaching—almost one-third of the entire teaching force—their increase explains approximately 27 percent of the ballooning in the school system.

However, in contrast to elementary classrooms, typical subject-area teachers at middle and secondary schools experienced, if anything, increases in their workloads. Average class sizes at these levels went up slightly during this period. The average number of classes taught per day changed little, and, at all grade levels, the average number of instructional hours that teachers work per week slightly increased.

As shown in Figure 3, there have also been large differences in the growth rates of different teaching fields. In particular, there have been dramatic increases in the number of teachers whose main field was bilingual or English as a second language (ESL), and those whose main assignment was elementary enrichment/elementary subject specialist (these are teachers who teach only one subject, such as art, music, physical education, computer science, or mathematics, to different classes of students in an elementary school). (The increase in teachers in the field of elementary enrichment/subject specialist is from 1990–1991 to 2017–2018. This question was not included in the 1987–1988 SASS.)

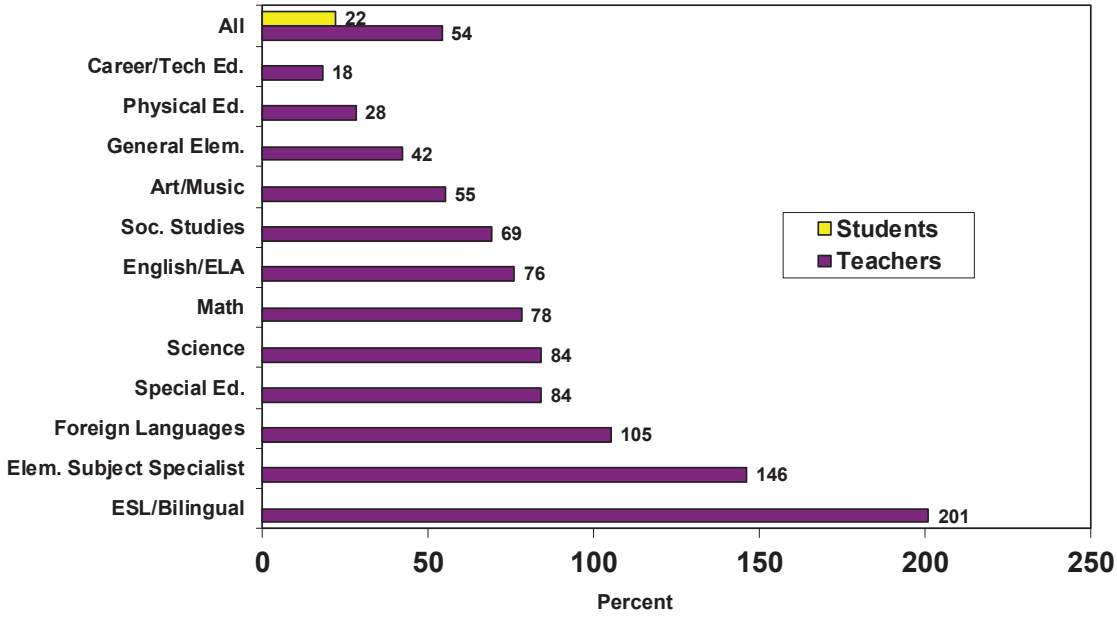


Figure 3. Percent Increase in Students, and Teachers by Field, from 1987–1988 to 2017–2018.

It is important to also note that these data on percentage increases in fields do not take into account the relative size of fields and can be misinterpreted if a large percentage increase occurs in a small field, or vice versa. In the case of the above two fields of teaching (bilingual/ESL and elementary subject specialists), while each is undergoing dramatic growth, combined, they remain a small segment of the teaching force, and hence, their rapid increases together account for only approximately 16 percent of the increase in teachers during this three-decade period.

The data also indicate that a significant source of the ballooning has been the growth of the field of special education, likely linked to changes in the Individuals with Disabilities Education Act, the main U.S. special education legislation. As Figure 3 shows, the number of teachers whose main field was special education increased by 84 percent, compared to 42 percent for general elementary school teachers. Special education classes average approximately half the size of typical classes in elementary and secondary schools, and special education is a relatively large field (11% of all teachers). Hence, the increase in special education teachers alone accounts for approximately 15 percent of the increase in the teaching force.

As the teaching force has grown, it has also experienced large shifts at the middle and secondary levels during this period. Overall, the number of typical subject-area teachers at the middle and secondary school level has increased by 59 percent. But there has also been a large redistribution of these teachers across fields, with some growing far faster than others. Among those growing the slowest from the late 1980s to 2018 were career/technical education (CTE), art, music, and physical education. Among those growing the fastest, besides special education and bilingual education, were the core academic subjects of English/language arts, foreign languages, mathematics, and science. For example, the number of teachers whose main field was English/English language arts increased by 76 percent. This is one of the largest fields and comprises a variety of subfields, such as literature, composition, reading, and language arts. The number of teachers whose main field was reading increased by 80 percent during this period.

The number of teachers of mathematics went up by 78 percent. The number of teachers of science went up by 84 percent. Although there are two and a half times as many general elementary teachers as mathematics and science teachers, the increase in math and science teachers accounts for approximately 20 percent of the overall ballooning in public schools. A major factor in the growth in the number of mathematics and science teachers appears to be changes in secondary school graduation requirements across the nation. While the number of courses required for graduation went up slightly for English, social studies, and foreign languages, they increased far more for mathematics and science during this period. This change meant that students took more mathematics and science courses. The data show that the number of 9th- through 12th-grade students enrolled in mathematics and science went up dramatically, in turn driving the large increase in the employment of teachers qualified in those subjects during this period.

However, we have not yet uncovered all of the reasons for, and sources behind, the ballooning of the teaching force. One possible set of factors behind the ballooning could be ongoing increases in the number and range of programs, courses, and curricula that schools are required to offer, especially at the secondary level. Educational historians tell us that programmatic expansion has been going on for a century, as schools have continually been asked to take on more and more goals and tasks that were once the responsibility of parents, families, and communities, and as our schools are continually asked to address larger problems and changes of our society and economy [8,9](Kirst, 1984; Vollmer, 2010). Indeed, it is important to note that the hiring of more math, science, special education, ESL/bilingual, foreign language, reading, and elementary enrichment/specialist teachers is simply in response to what the public wants. These are fields that have been in high demand.

There are, however, good reasons to further investigate the sources of the rapid growth in the teaching force, given the broad implications. For instance, the expansion of the teaching force, while in response to public demand, and to many of course entirely beneficial, has not been cost free, especially considering that teacher salaries are the largest item in school budgets. To illustrate the overall cost of teacher salaries, we aggregated the base salaries paid to teachers in the 2017–2018 NTPS teacher sample, by teachers' fields. To give several examples, the total of the base salaries paid to mathematics teachers in the U.S. in the period 2017–2018 was almost $29 billion; for special education teachers, it was just over $26 billion; for physical and health education teachers, it was just over $10 billion; for career/technical education, it was just over $8 billion; and for ESL/bilingual teachers, it was just over $4 billion.

How much has the ballooning itself cost the U.S. in additional teacher salaries? In other words, how much has it cost to hire teachers at a rate beyond what would have been necessary to have simply kept pace with student enrollment increases?

To answer these questions, we undertook an analysis to estimate the increase in aggregate salaries for teachers across the nation due to the ballooning in the number of teachers. Our approximate and conservative estimate is that in the almost three-decade period between 1987–1988 and 2015–2016 the additional aggregate cost of salaries due to hiring more teachers, beyond what would have been necessary to have kept pace with student enrollment increases, was over 40 billion dollars.

Our point is not that ballooning has been unnecessary or excessive or that the additional funds were not well spent. Indeed, having more teachers per students appears to be what the public wants. Our point is simply that the ballooning trend also leaves us with sobering questions: how have school systems been able to cope with such an increase in their largest budget item, who has been paying for it and what will this ballooning mean for the possibility of future increases in teachers' salaries—as proposed by reformers?

The ballooning of the teaching force also raises related questions regarding the performance and cost-effectiveness of the school system. Economists have long pointed out that, through technological advances, employees in many industries and occupations have become far more productive. A key example is the remarkable changes in agricultural

production over the past century—far fewer farmers produce far more produce. Our data raise the question of whether the opposite has been true for teachers. That is, does a decrease in the pupil–teacher ratio mean that teachers have become less "productive"?

A closer look suggests that more teachers per student does not necessarily mean that there has been a decrease in teacher "output". For instance, a portion of the ballooning is accounted for by the increased demand for, and employment of, bilingual/ESL and special education teachers. These fields appear to require a more intensive teaching process and smaller class sizes. Hence, fewer students per teacher in these fields does not necessarily mean that there has been a decline in the output of teachers.

Moreover, it is important to note that teaching is not the only occupation in which the number of practitioners has been increasing at a faster rate than the client base. For example, data from the U.S. Bureau of Labor Statistics (2020) show that the number of nurses, dentists, and pharmacists have all increased in recent decades at a faster rate than the populations they serve [7]. It is unclear why this has happened, but such increases do not necessarily mean that nurses, pharmacists and dentists are less productive than in the past.

Another implication of the ballooning is for the much-heralded teacher shortages, especially in mathematics and science. We have explored this issue in depth elsewhere [10–14]. As shown in Figure 3, these have been relatively high-growth fields. Among other findings from our research on shortages, our data analyses show that, contrary to conventional wisdom, the growth in the new supply and employment of qualified mathematics and science teachers has not only more than kept pace with increases in mathematics and science student enrollments, but also with mathematics and science teacher retirement increases—a point we address in Trend 2.

The significant increase in teachers over the past few decades has often been missed or misunderstood by commentators on both sides of the ideological and political spectrum. For instance, some liberal-left economists argued that the economic recession that began in 2008 could have been kick-started into recovery if those teachers who were laid off beginning in 2008 had simply been rehired [15]. Rehiring teachers as a means to stimulate the economy during the economic recession may well have been a worthwhile initiative. But it is important to recognize that the decline in the size of the teaching force from 2007–2008 to 2015–2016 (see Figure 1) was modest compared to the ballooning of previous decades. Our data show that the teaching force increased by approximately 1.3 million from 1987–1988 to 2007–2008, but only declined by approximately 64,000 teachers between 2007–2008 and 2015–2016, over half of which were from private schools. In other words, put into a historical context, reductions in the teaching force were neither large nor severe.

On the right side of the political spectrum, some advocates of education reform have cited the ballooning of the teaching force as evidence of "bureaucratic bloat" and "negative productivity" in the public sector [16,17]. This view holds that ballooning is evidence of a reduction in teacher productivity—more teachers per students, coupled with little change in test scores. The solution this view espouses is school choice and enhanced competition or privatization, based on the argument that private schools (and charter schools) are in a competitive, market environment, are leaner and more efficient, and hence marked by greater teacher productivity—in other words, they get better student achievement with fewer teachers.

However, these critics overlook the larger ballooning of the teacher force in private schools. Private schools have long had lower pupil–teacher ratios and smaller average class sizes—no doubt in response to what their market and clients want. But, as we show in Figure 2, it is important to recognize that not only has the teaching force grown faster in private schools than in public schools since the late 1980s, this has happened while the overall number of students in private schools has decreased. In other words, overall, the private school sector has been employing more teachers to serve fewer students. Not surprising, average full tuition at private schools rose 137 percent (after adjusting for inflation) during the 24 year period from 1987–1988 to 2011–2012.

The ballooning of the U.S. teaching force is a dramatic trend, and it is no surprise that a variety of commentators have noticed this trend and have offered explanations for it. However, the reasons for, and implications of, this dramatic growth are still unclear. We hope to address these questions with further research.

3.2. Trend 2: Older

The U.S. teaching force has been getting older. We have often heard about this trend because of its link to teacher shortages. Since the mid-1980s, numerous highly publicized reports have warned of a coming educational crisis caused by severe teacher shortages in elementary and secondary schools [18–27]. These reports predicted a dramatic increase in the demand for new teachers, primarily as a result of two converging demographic trends—increasing student enrollments and increasing teacher retirements due to a "graying" teaching force. Shortfalls of teachers, the argument surmised, would force many school systems to resort to lowering standards to fill teaching openings, inevitably resulting in high numbers of underqualified teachers and low school performance.

Our data confirm this demographic trend: the teaching force has gotten older, and teacher retirements have steadily increased. But our analyses also show that the aging trend is largely over. As Figure 4 shows, in the period 1987–1988 the age distribution of teachers was shaped like a tall peak. The modal, or most common, age was 41. As the years went by this group continued to age, and by 2007–2008 the modal age of teachers rose to 55. However, by 2017–2018 the most common age of teachers had spread out—with the modal age ranging from the low 30s to the high 40s.

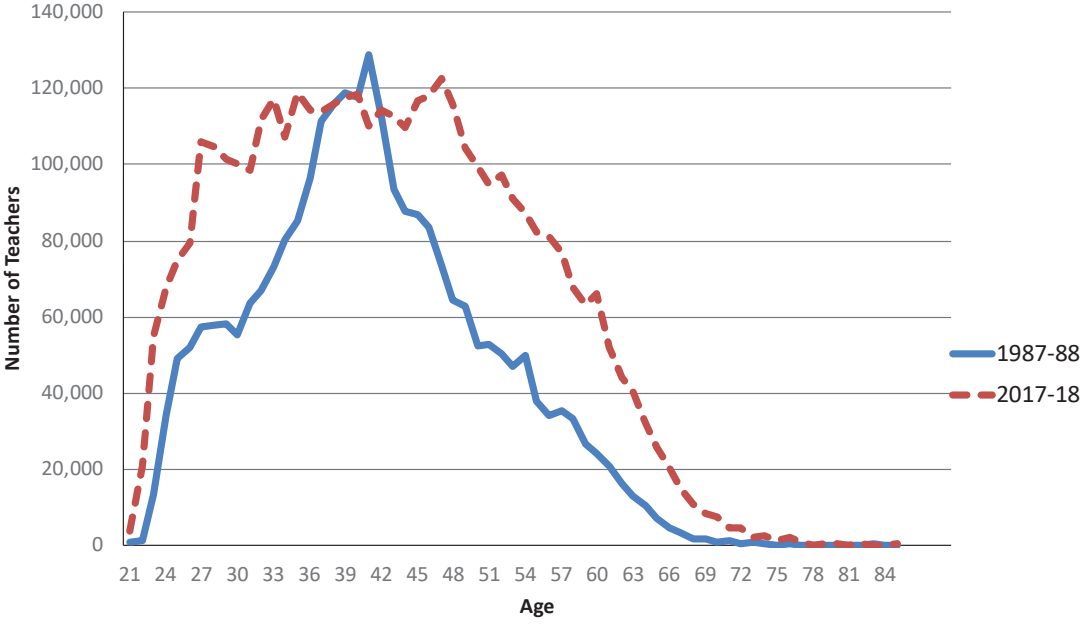


Figure 4. Age of Teachers, 1987–1988 and 2017–2018.

Likewise, the number of teachers 50 years or older increased, from approximately 529,000 in 1988 to a peak of 1,271,000 in 2008. However, by 2017–2018 the number of teachers 50 or older had decreased to approximately 996,000 (see Figure 5 for percentage changes).

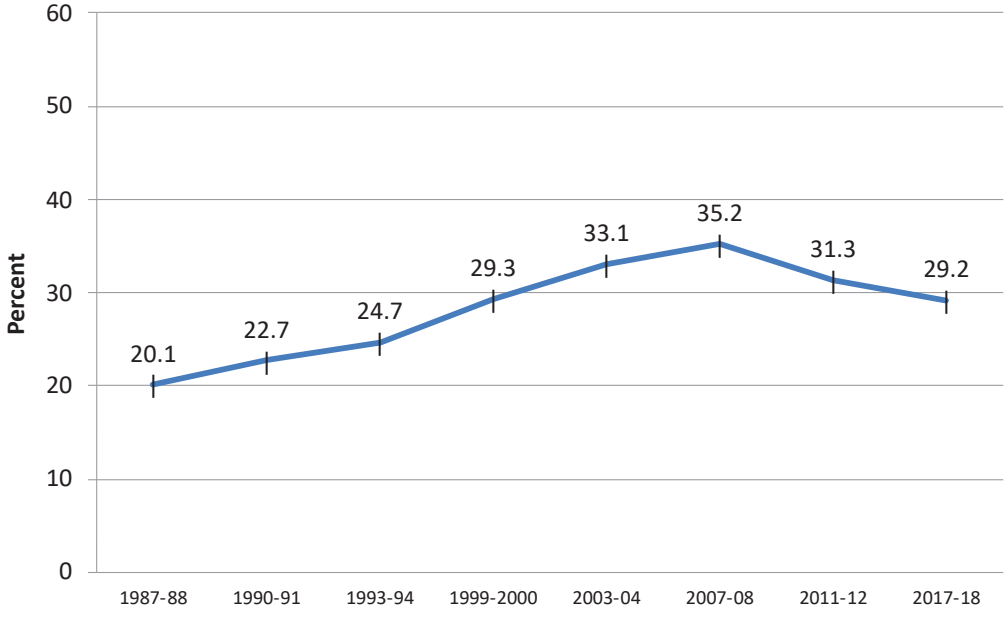


Figure 5. Trends in the Percent of Teachers Age 50 and Over, from 1987–1988 to 2017–2018.

What are the implications of this trend?

The aging of the teaching force has had large cost implications for both school budgets and for state pension systems—an issue that has received much media and policy attention in recent decades. Veteran teachers earn higher salaries, which, in turn, can strain school and district budgets. Increases in the number of retirees mean larger outlays from state pension plans [28,29].

But in discussions of the dire future for pension systems, another factor has been underemphasized: if schools replace retirees with new teachers, who earn lower salaries and who also pay into state pension plans, these additional costs could be lessened. As we discuss in Trend 3, not only have retirees been replaced with newcomers, but the flow of newcomers has become a flood.

Another implication of aging is its impact on the supply of teachers. Conventional wisdom has long held that retirements are a major factor behind teacher shortages. But teacher retirements have always represented only a small portion of all of those leaving teaching—less than one-third in recent years. And, if we examine all departures of teachers from schools (both those moving between schools and those leaving teaching altogether), retirement is only approximately 14 percent of the total outflow. In our research on the math and science teacher shortage, as mentioned above, we have found that, contrary to the conventional wisdom, the new supply of qualified math and science teachers has been more than sufficient to cover student enrollment increases and teacher retirement increases in these subjects. In contrast, the main, but under-recognized, source of mathematics and science teacher staffing problems is pre-retirement voluntary turnover [10,12,13], an issue to which we return in Trend 7.

3.3. Trend 3: Less Experienced

Graying is not the only change in the age and experience of the U.S. teaching force. Another opposite and unrecognized trend has occurred simultaneously. As Figure 4 shows, by 2017–2018, the teacher age distribution had spread out, with multiple peaks. As the number of older, veteran teachers increased, so has the number of beginning teachers. The

increase in beginning teachers is largely driven by the ballooning trend, that is, by the huge increase in new hires.

Most of these new hires are young, recent college graduates; however, a significant number are older, but inexperienced, beginning teachers. For instance, in the period 2017–2018, approximately 36 percent of new teachers in schools were over age 29, and approximately 15 percent were over 40—the phenomenon often referred to as mid-career switching. This has been fostered by recruitment programs such as Troops-to-Teachers. But mid-career switching into teaching is neither new nor an upward trend. Indeed, despite an increase in the number of older new hires, the current proportion of new hires age 29 or older is little different from that in the late 1980s.

Regardless of their age, these many new hires have resulted in a third trend—a dramatic increase in the number of teachers who are beginners—which we have labeled the "greening" of the teaching force. This trend is illustrated by the distribution of teachers by their years of teaching experience. In the period 1987–1988, the modal, or most common, teacher had 15 years of teaching experience under his or her belt, and the shape of the distribution was a single peak, as shown in Figure 6. By 2007–2008, the modal teacher was not a gray-haired veteran; he or she was a beginner in his or her first year of teaching. With the advent of the economic downturn beginning in the period 2007–2008 and the subsequent decrease in hiring, which was accompanied by layoffs—usually of beginners [30,31]—this greening of the teaching force slowed down. Hence, by 2011–2012, the modal teacher was someone in his or her fifth year. However, with the pickup in hiring, by 2017–2018, the modal teacher was again a beginner—in their first year.

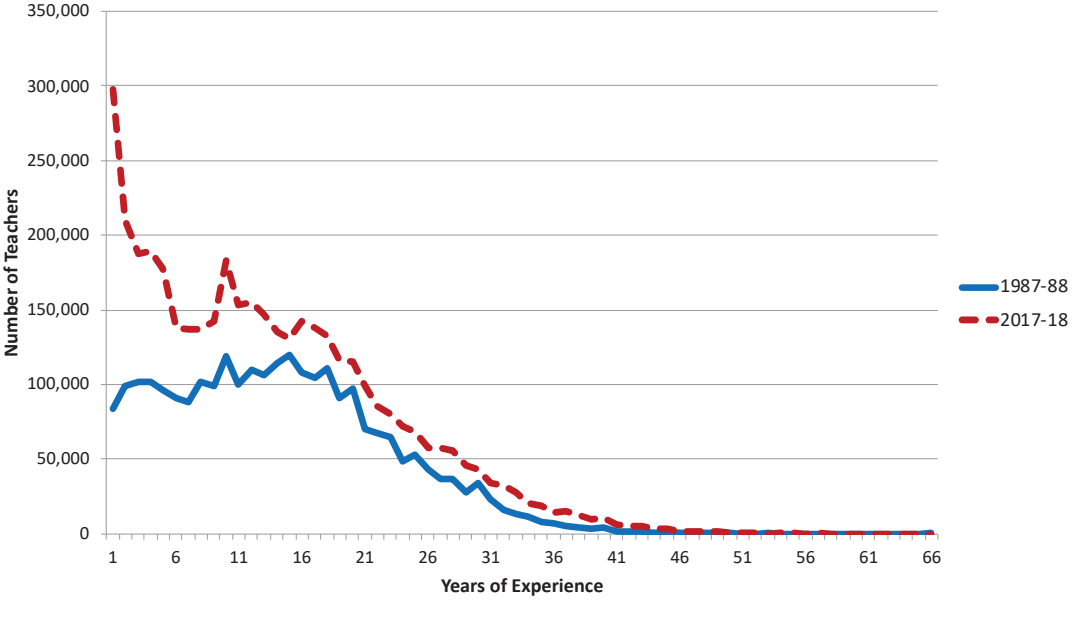


Figure 6. Teaching Experience of Teachers, 1987–1988 and 2017–2018.

There are, of course, still large numbers of veteran teachers—in the period 2017–2018, approximately one-quarter of all teachers had 20 years or more of teaching experience. But it is useful to recognize that the percentages on greening included above do not take into account the ballooning of the teaching force. Because the teaching force has dramatically grown, numerically there are far more beginners than before. For example, in the period 1987–1988, there were approximately 84,000 first-year teachers; by 2017–2018, there were approximately 300,000 first-year teachers. Similarly, in the period 1987–1988,

approximately 1 million teachers (approximately 37 percent of all teachers) had 10 or fewer years of teaching experience; in the period 2017–2018, there were over 1.8 million teachers (approximately 44 percent of the teaching force) with 10 or fewer years of experience.

What are the implications of this trend?

New teachers can be a source of fresh ideas and energy, and it can be beneficial to have new faculty coming into schools. On the other hand, having an increasing number of beginners, along with a decreasing number of veterans, in a school could also have a negative impact. Being taught by more experienced teachers, for example, can make a positive difference for students' academic achievement. A growing number of empirical studies document what is common sense among those who have taught—that teachers' effectiveness at improving their students' test scores usually increases significantly through their first several years on the job [32,33]. Beyond academic instruction, as they collect more experience, teachers also have more opportunity to develop many other skills crucial to teaching, such as how to deal with student behavior problems, how to teach students with diverse backgrounds and abilities, how to work and communicate with parents, how to best promote good work habits in students, and how to nurture students' self-esteem. Having sufficient numbers of veteran teachers in a school can also make a positive difference for beginning teachers. A solid body of empirical research documents that support, including mentoring by veteran teachers, has a positive effect on beginning teachers' quality of instruction, retention, and capacity to improve their students' academic achievement [34].

Greening also has large financial implications for several reasons. First, greening has implications for overall teacher salary costs. A teaching force with an increasingly large portion of beginners at the low end of the pay scale is less expensive, and this could ameliorate some of the increased payroll costs of the ballooning trend mentioned above. For instance, our analyses of the NTPS data show that in the period 2015–2016, the average starting salary for public school teachers with a college degree and no teaching experience was approximately $38,800; while that same year, the average salary for veteran public-school teachers, with over 15 years of experience and a master's degree, was approximately $66,500.

Second, greening has implications for pension systems. Greening may defray some of the increased pension costs resulting from the graying trend. Economic analysts have been arguing that there has been an alarming decrease in the ratio of new employees who pay into pension systems and Social Security, compared to retired employees who withdraw from pension systems and Social Security [28,29]. This imbalance does not appear to be the case for teaching, which is one of the largest occupational groups in the nation. As Figures 4 and 6 illustrate, both the proportion and the numbers of younger and less experienced teachers have increased, not decreased.

Moreover, as we will show in Trend 7, early attrition has remained high among this growing number of beginners, meaning that a significant number will never withdraw all of the funds accrued in their school system's pension plan. In many states it can take 5 to 10 years for a teacher to become fully vested, and hence eligible, upon leaving a school system, to receive any funds contributed by their employers to their pension plan [35–37]. In addition, school system pension plans are sometimes backloaded: pension payout levels do not increase evenly with each year of increased classroom experience. Rather, payouts are often relatively small until a teacher has accumulated two or three decades of classroom experience, after which they suddenly jump. High attrition combined with delayed vesting and pension backloading could result in a reduction in payouts from state pension plans, for more detailed discussion, see references [28,29].

In sum, greening (along with high attrition) means more of the teaching force is less expensive and more teachers are paying into pension plans, while fewer of them will fully withdraw from pension systems.

3.4. Trend 4: More Female

Educational historians and scholars have documented that when the public school system was created in the U.S. at the end of the 19th century, teaching was intentionally

designed to be a predominantly female occupation [38–42]. There were a number of reasons for this, among them were that, since women had fewer other employment options compared to men, and generally paid less, they were less expensive for school districts to employ than men. Moreover, teaching seemed to fit the traditional gender-role stereotype—that women are assumed more naturally suited to child rearing and caregiving than men [43]. As a result, in this view, historically, in the U.S., K-12 teaching became stereotyped as lower-paid "women's work".

For the same reasons, both financial and social, these scholars have held, teaching became less attractive to males. Unlike women, males had more and better-paid occupational options, hence becoming a teacher entailed greater financial "wage sacrifice", "alternatives foregone", and "opportunity costs" for males [44]. Hence, males were less likely to both enter and to stay in the teaching occupation. Indeed, some have concluded that women essentially subsidized the education system for most of the previous century because they were relatively well-qualified employees working for relatively low wages [43,45,46].

However, in recent decades, educational scholars have predicted that this would change as gender stereotypes softened, many occupations became less sex-segregated, and women's employment options and career opportunities expanded [47–49]. Over the past four decades, many occupations and professions that traditionally have been predominantly male have opened up to women. For instance, data from the Bureau of Labor Statistics show that in 1972, only 10 percent of physicians, 4 percent of lawyers, 2 percent of dentists, and 13 percent of pharmacists were female; by 2019, these proportions had risen to 41 percent (physicians), 37 percent (lawyers), 36 percent (dentists), and 63 percent (pharmacists) [7]. With career and employment alternatives increasingly available, especially in higher-paid, and/or more prestigious occupations and professions, some have predicted that fewer women would enter lines of work that traditionally have been predominantly female.

Perhaps surprisingly, this has not happened for teaching in the U.S. Both the number of women entering teaching and the proportion of teachers who are female have gone up. The SASS data, along with other NCES data, show that since the early 1980s there has been a slow but steady increase in the proportion of teachers who are female, from 67 percent in the period 1980–1981 to over 76 percent in the period 2017–2018 (see Figure 7).

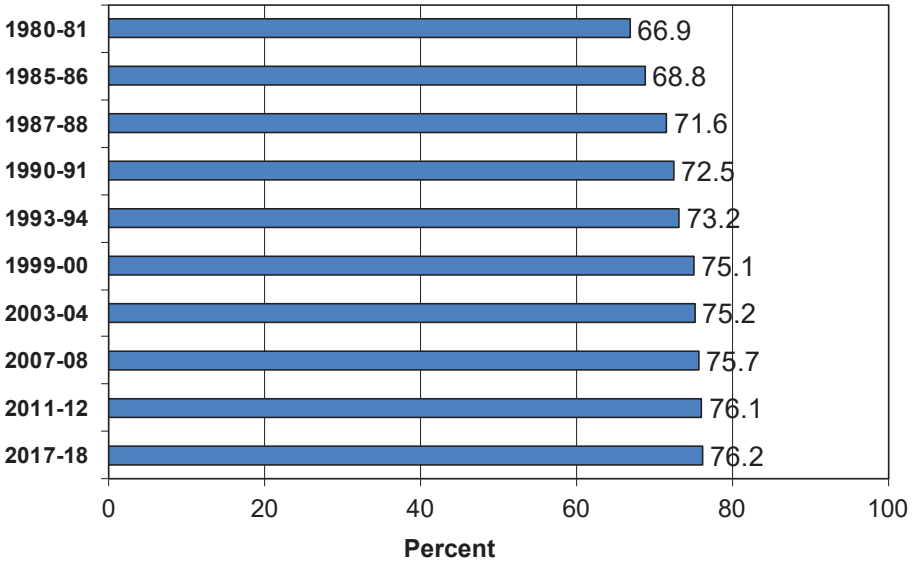


Figure 7. Percent Female Teachers, from 1980–1981 to 2017–2018.

It is unclear why this increase has occurred. Here we will offer four possible hypotheses for this increase in the female-to-male ratio in teaching. These hypotheses are neither exhaustive nor mutually exclusive.

(1). One possible hypothesis for the increase in the female-to-male ratio in teaching is that teaching has continued to be less attractive and less welcoming to men than women because of the persistence of traditional gender roles and segregation as well as issues with financial rewards, occupational status, and social pressures associated with these roles.

For example, some analysts have argued that males continue to be discouraged from entering teaching due to the persistence of a "wage sacrifice"—as men continue to be able to earn substantially more in other lines of work. For an insightful summary, see references [49,50].

Along with financial disincentives, researchers have held that males continue to be faced with a "prestige penalty" if they enter teaching [51]. The occupational status and stature of teaching has long been a source of both hope and frustration for educational reformers. Historically, during the development of the school system, elementary and secondary teaching was viewed as requiring less ability, skill and training, and in turn, was afforded less pay and less occupational prestige than the well-established professions, such as medicine, law, academia, engineering, dentistry, and accounting. For example, in his classic work, *The Sociology of Teaching*, published in 1932, Willard Waller, noted that, "The difficulties of the teacher . . . are greatly increased by the low social standing of the teaching profession and its' general disrepute in the community at large . . . Concerning the low social standing of teachers much has been written. The teacher in our culture has always been among the persons of little importance and his place has not changed for the better in the last few decades" [52] (pp. 11, 58). Similarly, C. Wright Mills (1951, pp. 129) in his classic study *White Collar* published in 1951, referred to schoolteachers as the "proletarians of the professions" in the U.S. [53] (p. 129).

Not surprisingly, in response, educational reformers have repeatedly sought to promote the view that elementary and secondary teaching is in reality a highly complex kind of work, requiring specialized knowledge and skill, and deserving of the same status, pay and standing as highly regarded professions, like law, academia and medicine. These aspirations have yielded mixed results. Scholars have found that, as far as prestige and status, elementary and secondary teaching in the U.S. lies in the middle—compared with a wide cross-section of occupations and professions. As documented in the General Social Surveys over the past half century the prestige and stature of K-12 teaching has been well below that of professionals, such as doctors, lawyers, professors, dentists and engineers, has been slightly higher than that of the skilled trades, such as plumbers, electricians, and well above that of the lowest status types of work, such as secretaries, waiters and barbers [54,55]. However, when compared solely to the subset of white-collar occupations and professions, teaching lies near the bottom, at a similar level as nurses, police, librarians and social workers.

International comparisons of the stature of teachers have also confirmed this finding. In the recent Global Teacher Status Index, teachers were ranked against 14 other white-collar occupations in 35 nations [56]. On average, teachers' status was ranked near the bottom of the 14, well below doctors, engineers, lawyers and accountants and only above librarians, social workers and web designers. There were also large differences across nations; China, Malaysia, Taiwan and Indonesia respect their teachers more than other countries, Brazil and Israel were at the lower end of the Teacher Status Index, and the U.S was in the middle.

Given this relative lack of professional legitimacy and prestige, social scientists have argued that the status and social role of teaching has been akin to that of a "semi-profession" in the U.S. [57–59]. The persistence of this view may discourage some males from entering teaching.

In addition to financial and status factors, the persistence of social pressures surrounding the gender roles could also be a factor discouraging males from becoming teachers—especially at the elementary school level. An insightful field study conducted

over three decades ago found that enduring gender stereotypes made it very difficult for males to teach in elementary schools, despite a strong desire from the public to employ more male teachers in elementary schools [60]. Male elementary school teachers faced suspicion and distrust, especially from female teachers, who felt that men were simply not suited to work with young children. It is unclear whether such attitudes still persist and it might be illuminating if this study were to be updated and replicated in today's schools. Interestingly, some recent research has found that socialization factors such as parental influence may be partly responsible for the continued gender homogeneity of the teaching profession, as the children of teachers are disproportionately likely to also become teachers [61].

However, it is possible the influence of gender roles and their associated financial, status, and social pressures, in dissuading males from entering teaching, has diminished over time. The data show that the increase in the female-to-male ratio in teaching is not due to a decrease in males entering, or an increase in males exiting, the occupation. While the data show that the elementary school teaching force remains overwhelmingly female (89% in 2017–2018), there has been a strong increase (57%) in the number of male teachers employed in elementary schools in the past three decades (Figure 8). Many of these males are employed as elementary enrichment/subject specialists, a high-growth field (see Figure 3). Moreover, the TFS data show that there is little difference in the turnover rates of male and female teachers. The number of male teachers employed in schools has grown overall by 29 percent since the late 1980s, which is also faster than the rate of increase in the student population (22%). The increase in the female-to-male ratio in teaching is because the number of females in teaching has increased at over twice the rate of males (see Figure 8). Rather than males being increasingly "pushed back", or "pushed out", it may be that females are increasingly being "pulled into" teaching.

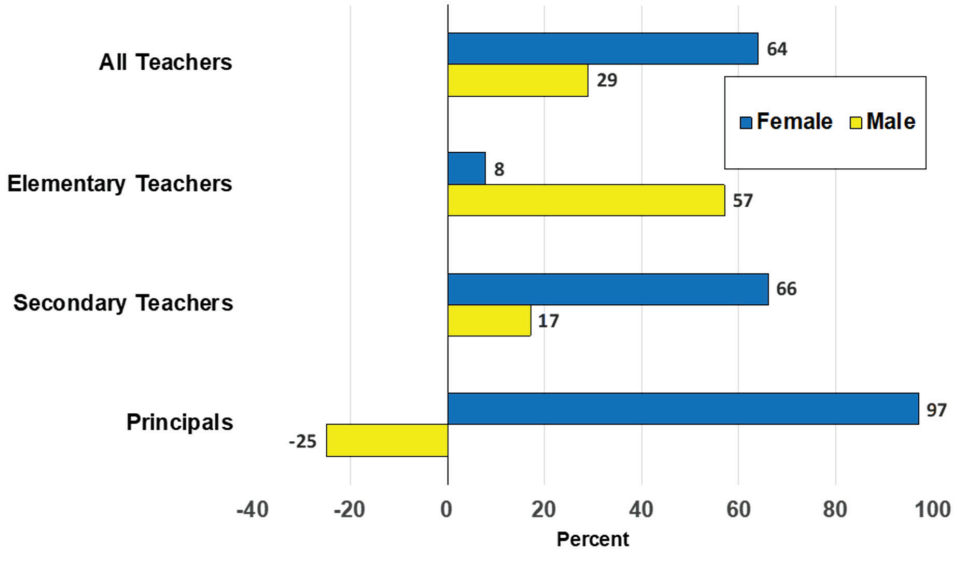


Figure 8. Percent Change in Teachers and Principals, by Sex, from 1987–1988 to 2017–2018.

(2). A second possible hypothesis for the increase in the female-to-male ratio in teaching could be a variant of the above mentioned increasing-career-opportunities hypothesis—females have other employment options in general, but also growing opportunities in the educational sector, both at the secondary level and in leadership. The increase in female teachers is not spread evenly across teaching fields, levels and roles within schools. As shown in Figure 8, increases in the proportion of female teachers have been concentrated

at the secondary level, where the majority of teachers were male until the late 1970s. The largest increase in the representation of females in particular fields has been in mathematics and science. In the period 1987–1988, these were both male-dominated fields. By 2017–2018 approximately two-thirds of math and science teachers were female. Moreover, as shown in Figure 3, these have both been high-growth fields.

As also shown in Figure 8, there have been far sharper increases in the proportion of female school heads—principals. By 2017–2018, over half of all school principals were female, up from 31 percent in the period 1987–1988, although the distribution of female principals varies by school level, as it does for teachers (see Figure 9). There has also been a similar increase in females at the position at top of the hierarchy in local educational agencies—school superintendents. Data from the American Association of School Administrators show that, while superintendents remain predominantly male, there has been an increase in the numbers of females holding these positions over the past five decades—from 1.2 percent in 1982 to 27 percent in 2020 [62].

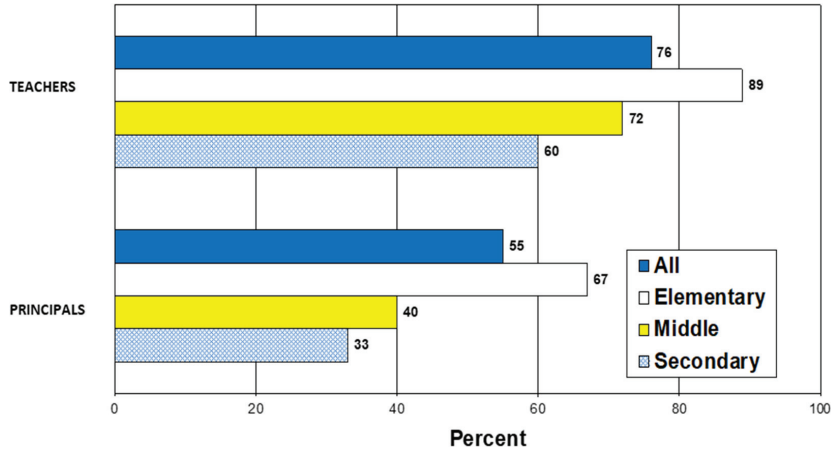


Figure 9. Percent Female Teachers and Principals, by School Level, 2017–2018.

The trend in female school leaders could be a factor in the recruitment, as well as the retention, of females, including those of high academic ability—an issue we address in Trend 6. Historically, while teaching was designed as a predominantly female occupation, educational administration was designed to be "men's work" [38,40]. Part of the rationale was that the recruitment and retention of capable males in schools required a career ladder with opportunities for advancement and enhancement in status, pay, and authority. Likewise, the contemporary opening up of educational management and leadership to women—demonstrated by the rapid growth in the numbers of female school leaders—could be one possible explanation for the continuing attraction of teaching and education careers for women, despite the growth of other employment opportunities.

(3). A third hypothesis to explain the increase in the female-to-male ratio in teaching might be that the overall proportion of adult women obtaining college degrees, and entering the paid workforce, has dramatically increased over the past four decades. Hence, while women have more job choices than in the past, the large overall increase in women with college degrees and women seeking employment may be partly responsible for the increase in females entering teaching.

Data from the NCES' Integrated Postsecondary Education Data System show that since the late 1990s over half of college Bachelor's degrees in any given year are awarded to women [63]. This is especially the case for degrees in education. In 2018, approximately 43 percent of all Bachelor's degrees awarded in the U.S. went to males, while only approximately 18 percent of Bachelor's degrees in education awarded went to males. Moreover,

the above cited data from the Bureau of Labor Statistics indicate that in a number of specific occupations and professions the number of women has gone up at a faster rate than in teaching [7].

But the data also show that the number of women in teaching has continued to increase at a rapid rate. The number of women employed in the U.S. labor force overall increased by 36 percent between 1988 and 2016, from 55 million to 74 million. However, the number of women in K-12 public school teaching increased by more than twice that rate during the same period. The proportion of all employed females who were teachers rose from 3 percent to 3.9 percent during the same period. That is, teaching's share of employed women has gone up, not down, during the same time that the number of women dramatically increased in many male-dominated occupations and professions. It appears that the increase in women in teaching is more than simply a result of more women obtaining college degrees and entering the labor force.

(4). Finally, a fourth possible hypothesis for the increasing proportion of female teachers might have to do with negotiating the dual roles of homemaker and breadwinner—the fit between work-life and family life. Scholars have long held that an important factor behind the high proportion of women in teaching over the past century was the perception of a relatively workable fit between the job of teaching and the work of child rearing [38–41,49]. Compared to many lines of work, teaching has long been characterized as a relatively flexible, easy entry/easy exit/easy re-entry occupation, with fewer career and financial penalties for temporary leaves, such as for raising children [44,48,64]. Moreover, given the perception of a shorter workday and summers off, caring for family was thought to be more manageable for teachers than for those in many other jobs and careers. Given the continuing disproportionate responsibility on the part of women for the domestic labor and childrearing in families [50,65], this perceived workday structure may still be disproportionately attracting women to teaching and hence a factor in the increase in the female-to-male ratio. But it is also important to recognize that teaching may not be as amenable to family life as this perception and stereotype holds, given the SASS/NTPS data documenting that teachers in the U.S. typically work more than a 40 hour week and many seek additional employment during the school year and in the summer.

What are the implications of this trend?

One set of questions raised by the data concerns the possible implications for students of the increasing female-male ratio in teaching. If the trend continues, 8 of 10 teachers in the U.S. may soon be female. A large percentage of elementary schools will have few, if any, male teachers. An increasing number of students may encounter few male teachers during their time in either elementary or secondary school. Given the importance of teachers as role models, and even as surrogate parents for some students, certainly some may see this trend as a problem and a policy concern.

Another set of questions raised by the data concerns the implications of an increase in the female-male ratio for the stereotype of teaching as "women's work" and for the future status and standing of teaching as an occupation. The direction of causality and of the relationships between gender, pay and occupational status are a source of debate, but as discussed above, scholars have long documented that female-dominated occupations are more likely to have lower salaries and more likely to be held in lower esteem than many male-dominated well-established professions [45,46,50,51,59,61]. This raises a troubling question: if teaching continues to become an increasingly female-dominated occupation, would this lead to a further solidification of the stereotype of teaching as "women's work", and in turn, make it harder in the future for teaching to improve its pay and its' status as a respected profession in the U.S.?

3.5. Trend 5: More Diverse, by Race/Ethnicity

While the teaching force is becoming more homogenous gender-wise, the opposite is true for the race/ethnicity of teachers. At first this finding may also seem surprising. For several decades, shortages of minority teachers have been a major issue for U.S

schools. It is widely held that, as the nation's population and students have grown more racially/ethnically diverse, the teaching force has not kept pace. The result, in this view, is that minority students in the nation's schools increasingly lack minority adult role models, lack contact with teachers who understand their racial and cultural background, and often lack access to qualified teachers of any background, because white teachers eschew schools with large percentages of minorities [66–73]. This minority teacher shortage, in turn, is widely viewed as one of a number of key factors behind the minority achievement gap and, ultimately, unequal occupational and life outcomes for minority students [74–77]. In response, in recent decades, numerous government and nongovernment organizations have instituted and funded a variety of programs and initiatives to recruit minorities into teaching. Moreover, often these efforts seek to place these new teachers in schools serving disadvantaged and minority student populations. By 2008, over half of the states had some kind of minority teacher recruitment policies or programs in place. (Throughout this report, we use U.S. Census Bureau classifications of race/ethnicity: "Minority" includes those identified as: Black/African American; native Hawaiian/Pacific/Islander and Asian; Native American/Indian/Alaska Native; Hispanic/Latino; and those of multiple races. "Hispanic/Latino" refers to ethnicity and includes those of all races. Over half of those identifying as Hispanic also identify as White. Hence, the term "person of color" is not synonymous with minority, and, following Census Bureau usage, we will not use the former term.)

However, our extensive analysis of national data show that this portrait is changing [78,79]. Our data analyses confirm that teachers remain a primarily White, non-Hispanic workforce and that a parity gap continues to persist between the percentage of minority students and the percentage of minority teachers in U.S. schools. For instance, in the 2017–2018 school year, approximately 40 percent of the nation's population belonged to minority groups, 51 percent of all elementary and secondary school students were minority, but only 20 percent of all elementary and secondary school teachers were minority (see Figure 10). Moreover, this gap between percentages of students and teachers holds for each of the main minority groups—Black, Hispanic, Asian, and Native American.

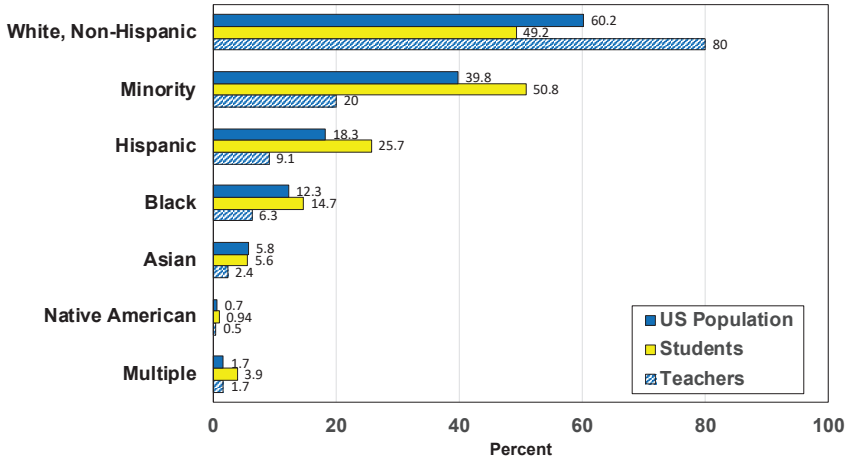


Figure 10. Percent US Population, Students and Teachers, by Race/Ethnicity, 2017–2018.

But the data also show that this parity gap is not due to a failure to recruit minority teachers. The gap has persisted in recent years largely because the number of white, non-Hispanic students has decreased, while the number of minority students has increased. The percent of all teachers who belonged to minority groups increased from 12.5 percent in the period 1987–1988 to 20 percent in the period 2017–2018. Growth in the number of

minority teachers outpaced growth in the number of minority students and was more than three times the growth rate of white teachers (see Figure 11). Moreover, during this period, increases in teachers outpaced increases in students for Blacks, Hispanics, and Asians. However, the data also show large differences in the pace of these teacher increases by group. For instance, the number of Black teachers increased by 29 percent, the number of Asian teachers increased by 263 percent, while the number of Hispanic teachers increased by 373 percent [78,79].

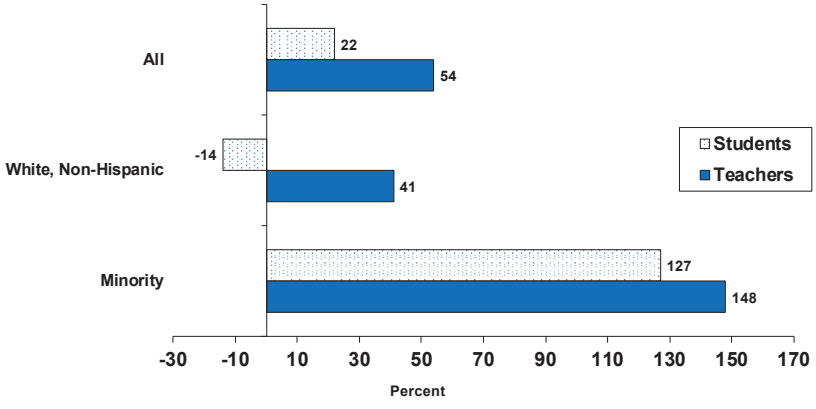


Figure 11. Percent Change in the Number of Students and Teachers, by Race/Ethnicity, from 1987–1988 to 2017–2018.

Moreover, these percentages do not take into account the ballooning of the teaching force. Since the teaching force has dramatically grown, numerically there are far more minority teachers than before. In the period 1987–1988, there were approximately 327,000 minority teachers employed in public and private schools; by 2017–2018, there were over 810,000 minority teachers. Our data also show that the increase in the number of minority teachers has not been even across different types of schools. Most of the increase has been in higher-poverty public schools [1]. Minority teachers are two to three times more likely than white teachers to work in hard-to-staff schools serving high-poverty, high-minority, and urban communities.

This increase in the number of minority teachers has been something of an unheralded victory. Commentators and researchers have tended to discuss the minority teacher shortage and the outcome of minority recruitment efforts in dire and pessimistic terms. However, although the proportion of minority students in schools is still far greater than the proportion of minority teachers (Figure 10), the teaching force has in fact grown more diverse. While we have not evaluated the efficacy of particular programs, our data suggest that the widespread efforts over recent decades to recruit more minority teachers and place them in schools serving disadvantaged and minority student populations appear to have made a difference. Moreover, given the continued parity gap, the data suggest that it would be worthwhile to continue such efforts.

However, while minority candidates have entered teaching at higher rates than whites in recent decades, the data also show that the rates at which minority teachers depart from schools is higher, at a statistically significant level, than that of white teachers, and has also been increasing. In the decades from the late 1980s to 2012–2013, the annual rate of minority teacher turnover from public schools increased by 45 percent, undermining minority teacher recruitment efforts [78,79]. Indeed, the diversification of the teaching force is all the more remarkable because it has occurred in spite of the high turnover rate among minority teachers. Our data suggest that ameliorating the minority teacher shortage will require focusing on both minority teacher recruitment and teacher retention. We return to the issue of minority teacher turnover in Trend 7.

3.6. Trend 6: Consistent Academic Ability

It has long been held that higher academic achieving college students—"the best and brightest"—are likely to find elementary and secondary teaching less attractive than other career and job options in the U.S. [19,44,47], as well as England [80]. Over the years, data from different sources have seemed to confirm this. For instance, based on the assumption that academic ability is accurately captured by standardized tests, a number of analyses have shown that the Scholastic Assessment Test (SAT)/American Collegiate Test (ACT) and other standardized college entrance exam scores of college graduates going into teaching have long been well below the average for college graduates [81]. In our own analyses of national data on college seniors from the NCES' Baccalaureate and Beyond Survey (in both 1999–2000 and 2007–2008), we found that this was especially true for those majoring in Education, who tended to have among the lowest average SAT scores. Moreover, within most fields and majors, we found that those who became teachers had lower SAT scores than those in the same field/major who did not go into teaching.

Not only do teachers tend to have below-average academic test scores, some researchers and commentators maintain that the academic ability of teachers has been declining over time—and that gender is at the root of the issue. While the number of women going into teaching has increased, as discussed in Trend 4, proponents of this view have argued that the academic quality of women who choose to go into teaching has gone down. With alternative careers and jobs increasingly available, this view holds that the "best and brightest" women have decreasingly entered traditionally female-dominated occupations and professions, such as teaching [82,83]. As mentioned, some have argued that women essentially subsidized the U.S. education system for most of the previous century because they were relatively high-ability employees working for relatively low wages. But, the argument continues, this subsidy has stopped, and as a result, the academic caliber of the female portion of the teaching force has declined in recent years [47,48].

In the data, however, support for this proposition appears to be mixed. For instance, using longitudinal survey data of birth cohorts from the 1940s to 1990, Bacolod (2007) found that females within the top quintile of standardized test results increased as a share of professionals while decreasing as a share of teachers. Similarly, another study looking at trends in female standardized test scores from the 1960s to 2000 found a decline in the proportion of female teachers who scored in the high deciles [83]. But another study with data from the Educational Testing Service (ETS) showed no decline in SAT scores of teachers as a whole from the mid-1990s to the mid-2000's [84].

Of course, we cannot assume that either the "best and brightest," or those scoring higher on standardized tests, are the best or the most effective teachers. Accurately measuring both academic ability and teaching quality are subjects of controversy. Moreover, the relationship between them is unclear. But academic ability is often assumed to be an important indicator of both the caliber of employees in any line of work and the attractiveness of an occupation or profession.

We examined these trends using another possible measure of academic ability—the selectivity or competitiveness of one's undergraduate institution, which is no doubt correlated with SAT/ACT and other standardized college entrance exam scores. The measure we used is Barron's six-category ranking of colleges and universities: most competitive, highly competitive, very competitive, competitive, less competitive, not competitive [85]. The top two categories have accounted for approximately 14 percent of institutions and approximately 21 percent of undergraduates. The bottom two categories have accounted for approximately 19 percent of institutions and approximately 13 percent of undergraduates.

What did we find?

Under one tenth of newly hired first-year public school teachers come from the top two categories of higher education institutions. Because in some of the data cycles there is a large amount of missing data on undergraduate institutions for private teachers, for this trend we only report data for public school teachers. Approximately one-quarter come from the bottom two categories. Approximately two-thirds of first-year teachers come from

middle-level institutions. From 1988 to 2018 these proportions tended to fluctuate both up and down slightly, but over this three decade period, it appears that there has been a slight overall increase in the proportion of new teachers from the least selective colleges and a slight overall decrease in the proportion from the most selective schools (see Figure 12).

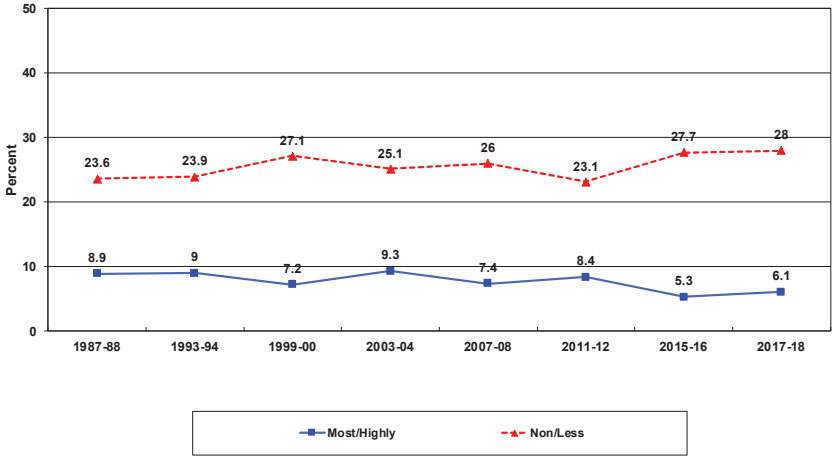


Figure 12. Percent First-Year Public School Teachers, by Selectivity of their Undergraduate College/University, 1987–1988 to 2017–2018.

There are, however, differences in college selectivity by gender. Interestingly, beginning male teachers in public schools have often been slightly more likely to come from top-ranked institutions than have female teachers (see Figure 13). And, while the levels fluctuated between 1987–1988 and 2017–2018, there appears to have been an overall decrease in the proportion of first-year male teachers in public schools coming from the top two ranks of institutions—from 13 percent in the period 1987–1988, to 10 percent in the period 2011–2012, and to 7 percent in the period 2017–2018.

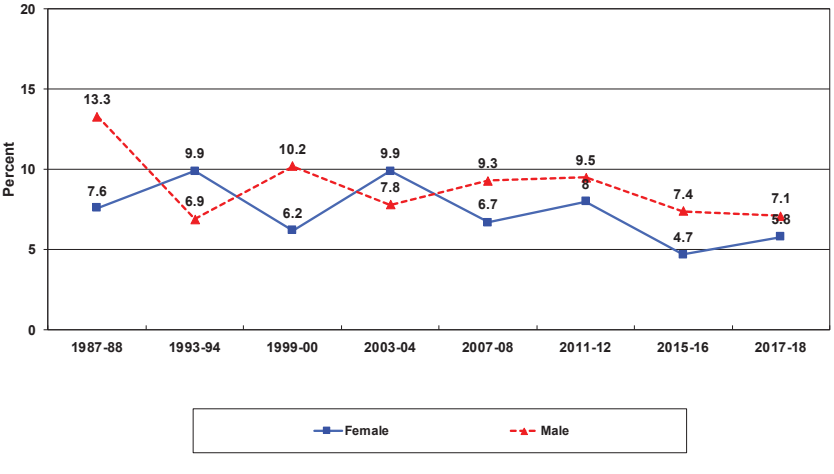


Figure 13. Percent First-Year Public School Teachers, with Undergraduate Degrees from the Most and Highly Selective Colleges/Universities, by Gender, 1987–1988 to 2017–2018.

For first-year female teachers in public schools the proportion coming from the top two categories of institutions has also slightly fluctuated from year to year. There also appears to have been a decrease in the proportion of first-year female teachers in public schools from the top two ranks of institutions—from 2011–2012 to 2017–2018. However, this decrease—from 8 to 6 percent—has been less than for male teachers.

Once again, however, these percentages do not tell the whole story. Although the percentage of female teachers from top institutions has not changed much since the late 1980s, because the teaching force has ballooned (Trend 1) and has also become more female (Trend 4), numerically teaching has been employing far more female candidates from all of higher education, including top colleges and universities.

Hence, assuming our college selectivity measure of academic ability is valid, our data show that there has been a decrease in the proportion of male teachers from top institutions since the late 1980s. But these data also show that this trend has been less true of female teachers. Perhaps we should call the latter a non-trend. So, contrary to the view that there has been a sharp decline in the academic caliber of female teachers, our data suggest that this has not been true in the past three decades.

Along with the increase in the numbers and proportions of female teachers, we do not know the reasons for the apparent stability in the academic ability of females entering teaching in recent decades. As we suggested in Trend 4, perhaps the increased opportunities for women in school leadership and positions in secondary schools (see Figure 8) have been attractive incentives for able and ambitious females to enter education.

3.7. Trend 7: Unstable

Elementary and secondary teaching in the U.S. has long been marked by relatively high rates of annual departures of teachers from schools and from teaching altogether [38,44,86], and especially in the past three decades, there has been a burgeoning of empirical research on a wide variety of aspects of teacher turnover, mobility, and attrition [10,87–91]. Here we will draw from our own extensive research findings on teacher turnover—especially focusing on trends for beginning teachers.

To contextualize the stability of teachers we empirically investigated how attrition for teachers compares to that of employees in other lines of work, by analyzing national data from the Baccalaureate and Beyond survey on cross-occupational attrition rates of college graduates through their first 10 years after graduation from college. We found that teaching does have less attrition—those leaving the occupation entirely—than some other occupations, such as the child-care, secretarial, and paralegal fields (see Figure 14). Attrition rates are similar for teachers and police officers. Perhaps surprisingly, teacher attrition is higher than nursing attrition, and teachers have far higher attrition than traditionally highly respected professions, such as law, engineering, and architecture.

But these overall figures mask large differences in departure rates among different types of teachers and different locales, revealing the need to disaggregate our data. The flow of teachers out of schools is not equally distributed across states, regions, and school districts. The largest variations in teacher departures by location, however, are those between different schools, even within the same district. This includes both major components of total turnover—migration (teachers who move between districts and schools) and attrition (those who leave teaching altogether). The data show that almost half of all public-school teacher turnover takes place in just one-quarter of the population of public schools. The data show that high-poverty, high-minority, urban, and rural public schools have among the highest rates of turnover.

Moreover, by analyzing the SASS/TFS data on the demographic characteristics of both the schools migrating teachers move from and also those they move to for the following school year, we have been able to compare the flows of teachers between different kinds of schools. For instance, we found that of teachers who migrated between the 2004 and 2005 school years, over four times as many went from high-poverty schools to low-poverty schools as in reverse. Likewise, of teachers who moved, over three times as many went from

urban to suburban schools, as in reverse [13]. The net result is a large annual asymmetric reshuffling within the school system of a significant portion of the teaching force, with a net loss on the part of poor, minority and urban schools and a net gain to non-poor, non-minority, and suburban schools. For instance, Lankford, Loeb, and Wyckoff offer an insightful examination of teacher migration to and from New York City schools and the suburbs [92].

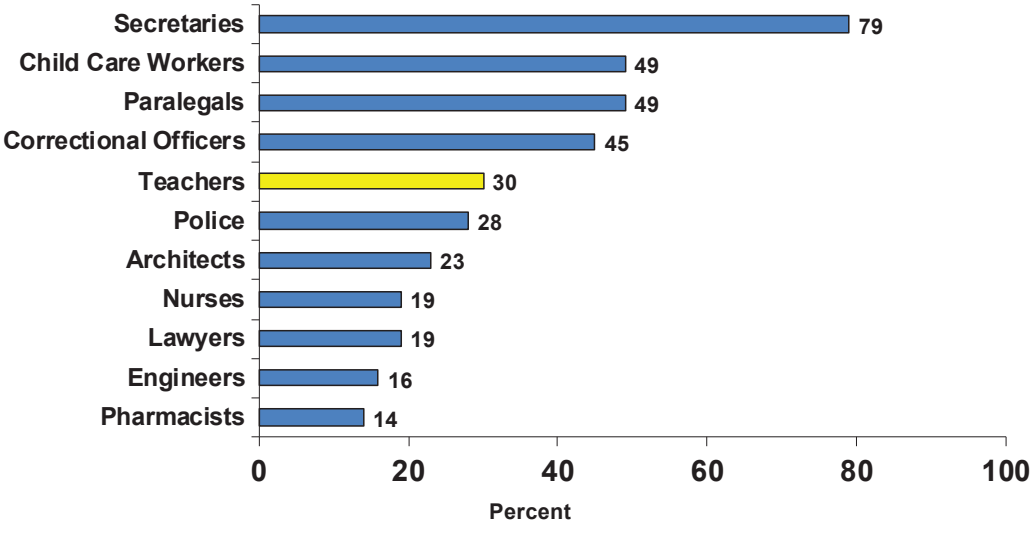


Figure 14. Among 1993 College Grads Who Entered Selected Occupations by 1997, Percent Who Left Occupation by 2003.

The data also show that rates of both moving between schools and leaving teaching altogether differ by the race/ethnicity of the teacher. As mentioned in Trend 5, over the past couple of decades, minority teachers have had significantly higher rates of turnover than white teachers. Moreover, the gap has widened in recent years. Why is this? Strikingly, while the demographic characteristics of schools appear to be highly important to minority teachers' initial decisions as to where to teach, this appears to be less the case for their later decisions about whether to stay or depart. What does impact their decisions, our analyses show, are school working conditions, in particular the degree of autonomy and discretion teachers are allowed over issues that arise in their classrooms, and the level of collective faculty influence over school- wide decisions that affect teachers' jobs. The same difficult-to-staff schools that are more likely to employ minority teachers are also more likely to offer less-than-desirable working conditions, according to our data, and these conditions account for the higher rates of minority teacher turnover. These high levels of turnover, of course, undermine efforts to diversify the teaching force [78,79].

Beginning teachers, regardless of their race, have among the highest rates of turnover of any group of teachers. Almost two decades ago we estimated that between 40 to 50 percent of those who enter teaching leave teaching within 5 years [11]. This figure has been widely reported since, but it was only a rough estimate using cross-sectional national data. More recently, using national longitudinal data from the Baccalaureate and Beyond survey, we were able to more accurately document rates of cumulative beginning attrition (see Figure 15). We found that more than 44 percent of new teachers leave teaching within 5 years of entry.

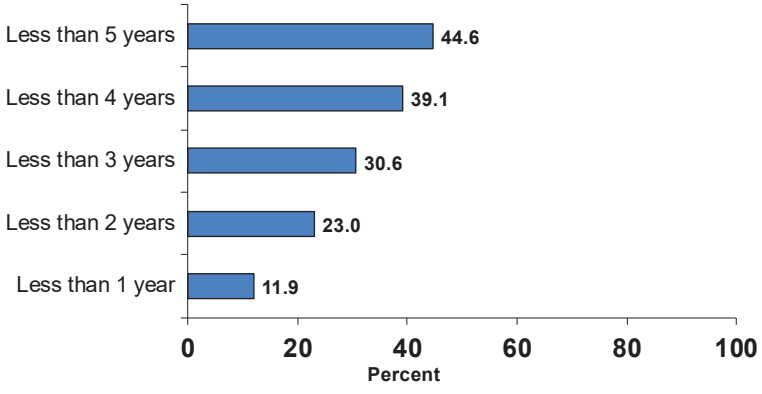


Figure 15. Cumulative Percent Attrition of Beginning Teachers, by Years of Experience, 1993–2003.

Moreover, we have also found, despite a temporary dip after the 2008 recession, that high levels of attrition among beginning teachers have been holding steady or even slightly increasing since the late 1980s up to 2012–2013 (Figure 16). Note: the last cycle of the TFS to be administered was in the period 2012–2013 and it only included public school teachers. Hence, we present turnover data for public school teachers only in Figures 16–18.

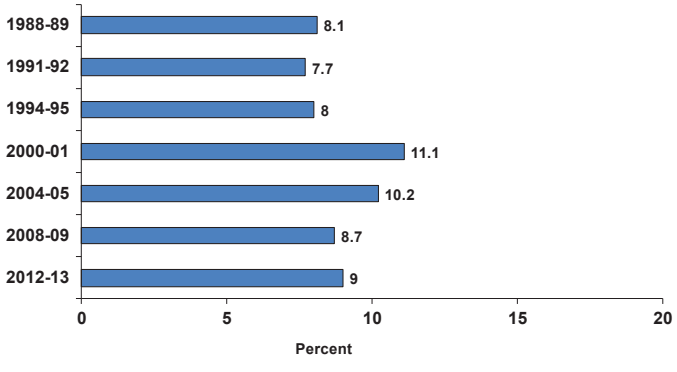


Figure 16. Percent Annual First-Year Public School Teacher Attrition, from 1988–89 to 2012–2013.

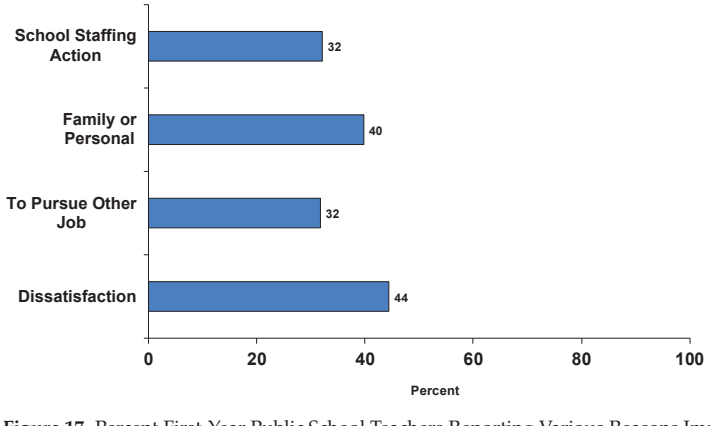


Figure 17. Percent First-Year Public School Teachers Reporting Various Reasons Important for Their Turnover, 2012–2013.

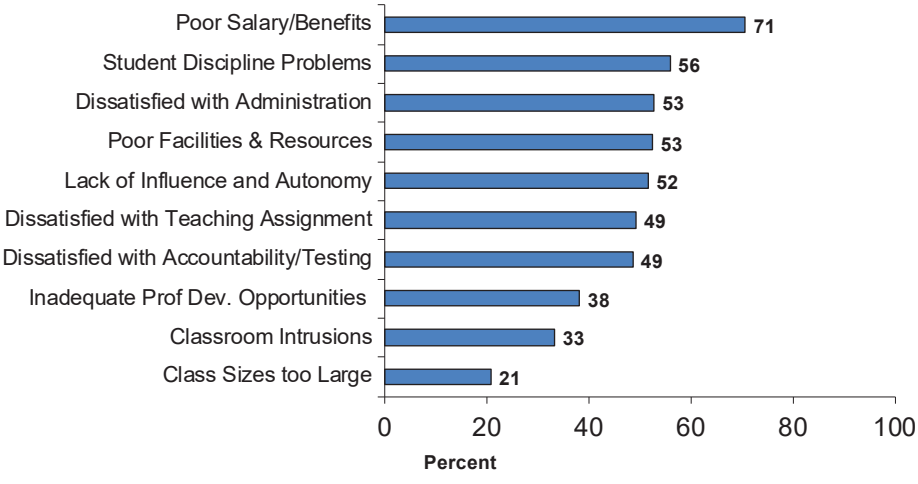


Figure 18. Percent First-Year Public School Teachers Reporting Dissatisfaction-Related Reasons Important for Their Turnover, 2012–2013.

Again, however, an increase in the annual percentage does not tell the whole story. Because the teaching force has grown dramatically larger, numerically there are more beginners than before (Trend 3), and hence the actual numbers of teachers who quit the occupation after their first year on the job has also increased. Soon after the 1987–1988 school year, approximately 7500 first-year public school teachers left teaching, while just after the 2012–2013 school year, approximately 11,000 first-year public school teachers left the occupation. There are more beginners in the teaching force, and these beginners are less likely to stay in teaching than others.

Along with analyzing TFS data on the rates and cross-school variations of turnover, we also analyzed data drawn from an additional set of items in the TFS questionnaire that asked teacher-respondents to indicate the reasons for their departures from a list in the survey questionnaire. Of first-year teachers who departed from their school after the end of the 2011–2012 school year (including both movers and leavers), one-third indicated that they had been involuntarily transferred, laid off, or terminated (see Figure 17). This included for both budgetary and performance reasons. A larger portion indicated that family or personal issues played an important part in their decision to depart. This set included reasons of health, pregnancy, a residence move, and caring for family members. Another one-third indicated that they departed to pursue further education or another career. Finally, the most frequently cited set of reasons concerned dissatisfaction. Of those who indicated that dissatisfaction was a major reason for departure, most cited a variety of school and working conditions, including salaries, classroom resources, student misbehavior, accountability, opportunities for development, input into decision making, and school leadership (Figure 18).

In sum, beginners—the largest group within one of the largest occupations in the nation—have been leaving at relatively high rates, and these rates have held steady or even increased, in recent decades. Together, ballooning (Trend 1) and turnover (Trend 7) indicate a growing flux and instability in the teaching occupation, as both the large numbers of those entering teaching and the large numbers of those leaving teaching have been increasing in recent years.

These changes have large implications. Employee turnover in any occupation has pros and cons, costs and benefits. On the one hand, some degree of employee turnover, with the accompanying job and career changes, is normal, inevitable, and can be efficacious for individuals, for organizations, and for the economic system as a whole. Too little turnover of employees is tied to stagnancy in organizations; effective organizations usually

both promote and benefit from a limited degree of turnover by eliminating low-caliber performers and bringing in "new blood" to promote innovation.

On the other hand, high levels of employee departures are consequential not only because they can be a symptom of underlying problems in how well organizations function, but also because departures can entail costs and other negative effects for organizations and for the larger system These include large financial costs [93–95], and decreases in student achievement [96–98].

As mentioned earlier, we have found that one negative consequence of teacher turnover is its important but often overlooked role in teacher shortages [10–14]. Turnover is a major factor behind the problems that many schools have staffing their classrooms with qualified mathematics, science, and other teachers. Moreover, increases in turnover among minority teachers, especially in disadvantaged schools, undermine efforts to recruit new teachers in hard-to-staff schools and to diversify the teaching force. Additionally, research using longitudinal administrative data shows that teacher turnover disproportionately impacts some schools and districts more than others [99,100].

As mentioned in Trend 3, a possible financial consequence of high levels of teacher attrition among beginning teachers is a reduction in payouts from state pension plans, because increasing numbers of those who paid into systems remain in teaching long enough to reap full pension benefits.

A further negative consequence of high levels of beginning teacher attrition is the loss of newcomers before they are able to fully develop their skills. As mentioned earlier, a number of studies have documented the reasonable proposition that teachers' academic effectiveness—as measured by gains in their students' test scores—increases significantly with additional experience for the first several years in teaching [32,33].

4. Discussion

Has the elementary and secondary teaching force in the United States changed in recent decades? The answer is most certainly yes—and in a number of important ways. It has become far larger. It has simultaneously become older, younger, and far less experienced. It has simultaneously become less diverse, by gender, and more diverse, by race/ethnicity. It does not appear to be suffering from a decline in the academic ability of females entering teaching; indeed, the numbers of new teacher hires coming from the top-ranked colleges and universities has greatly increased. Finally, it remains unstable.

For each of these trends, significant questions immediately arise. What are the reasons for, and sources of, the trend? Will the trend continue, and what impact will it have? In this report, we have offered some possible answers to these questions.

It is also striking that while these trends raise important questions, until recently we have seen little awareness or discussion of many of them or their implications—whether by researchers, by policy makers, by educators, or by the public. But there are good reasons to investigate the sources and continuation of these changes—because if these trends do indeed continue, there will be large implications, with serious financial, structural, and educational consequences for America's educational system.

For instance, will the teaching force continue to outgrow the student population it serves, and, if so, why? If the teaching force does continue to balloon in size, the expense to local school districts could become unsustainable, and without an increase in funds, districts may increasingly turn to cutting teacher salary levels.

Will the hiring, and thus the greening trend, continue? In turn, will an increasing number of new hires decide not to stay in teaching, making teaching increasingly an occupation practiced by the young and inexperienced, and if so, why? If this trend continues, the expense to local school districts could become more sustainable, because of lower overall average salary costs per employee. In other words, will there effectively be a financial tradeoff between the numbers of teachers and their experience? On the other hand, as the older portion of the teaching force finishes retiring, will a large portion of the newcomers decide to stay with teaching to become the next generation of veterans?

Will the teaching force continue to become more female, and if so, why? If the teaching force does continue to become even more female-dominated, with the presence of male role models a rarity for an increasing number of students in their K-12 school years, will there be negative implications for students, both male and female?

Perhaps there is an irony in these changes. Historians tell us that when the public school system was invented a century ago, the teaching force was transformed into a mass occupation that was relatively low-paying, temporary, and designed predominantly for young, inexperienced women, prior to starting their "real" career of child rearing [38,44]. Perhaps the changes that we have traced represent not an entirely new face, but a return to the old face of the American teaching force.

A return to an earlier composition could have serious implications for the future status of elementary and secondary teaching in the United States. Professionalization has long been a source of both hope and frustration for teachers. Since early in the 20th century, educators have repeatedly sought to upend the notion that teaching is akin to lower-skill industrial work where teachers are interchangeable and easily replaced, and they have sought to promote the view that teaching is highly complex work, requiring specialized knowledge and skills, and deserving of the same status as well-established professions, like law, medicine, engineering, and academia. These efforts to enhance the professional status of teaching have also long met with limited success. And if teaching becomes an even larger, lower-paying line of work, predominantly employing young, inexperienced women, who stay for limited periods, it does not suggest optimism for the aspirations to promote the image of teaching as a respected profession.

At the same time, these possible future trajectories, and similarities between the contemporary transformation of the teaching force and its previous incarnation, are strictly speculative on our part. Nothing in our data analyses so far can be considered conclusive evidence that the teaching force is, or will be, "better" or "worse" in one way or another. As we indicated at the beginning of this report, thus far our objective has been exploratory and suggestive. At this point, we have more questions than answers.

What is clear is that large-scale changes are happening to one of the largest occupational groups in the U.S. Immediately after World War II and before the post-war baby boom, there were just over three-quarters of a million elementary and secondary teachers in the United States. By 2017–2018, there were more than five times as many—over 4 million elementary and secondary teachers. In the 2017–2018 school year alone, over 150,000 newcomers entered teaching. These data suggest a very large opportunity—one of the largest occupations in the U.S. is being expanded, replaced, and re-made. Who will our new teachers be and what implications will this have? Further research is warranted to answer this question.

Author Contributions: Conceptualization, R.I.; methodology, R.I., E.M., D.S., G.C.; software, All; validation, All; formal analysis, All; investigation, All; resources, All; data curation, All; writing—original draft preparation, R.I.; writing—All; Visualization, All; supervision, All; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ingersoll, R.; Merrill, E. *A Quarter Century of Changes in the Elementary and Secondary Teaching Force: From 1987 to 2012*; Statistical Analysis Report (NCES 2015-076); National Center for Education Statistics: Washington, DC, USA, 2017.
2. Ingersoll, R.; Merrill, E.; Stuckey, D. The changing face of teaching. *Educ. Leadership.* **2018**, *75*, 44–49.
3. Ingersoll, R.; Merrill, E.; Stuckey, D.; Collins, G. *Seven Trends: The Transformation of the Teaching Force*; Consortium for Policy Research in Education, University of Pennsylvania: Philadelphia, PA, USA, 2018.
4. National Center for Education Statistics. *Schools and Staffing Survey (SASS) and Teacher Followup Survey (TFS)*; U.S. Department of Education: Washington, DC, USA, 2013.
5. Taie, S.; Goldring, R. *Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results From the 2017–18 National Teacher and Principal Survey First Look*; U.S. Department of Education: Washington, DC, USA, 2020.

6. Goldring, R.; Taie, S.; Riddles, M. *Teacher Attrition and Mobility: Results From the 2012–2013 Teacher Follow-Up Survey First Look*; National Center for Education Statistics: Washington, DC, USA, 2014.
7. Bureau of Labor Statistics. Labor Force Statistics from the Current Population Survey. 2020. Available online: https://www.bls.gov/cps/cpsaat11.htm (accessed on 8 May 2021).
8. Kirst, M. *Who Controls Our Schools*; W.H. Freeman and Co.: New York, NY, USA, 1984.
9. Vollmer, J.R. *Schools Cannot Do It Alone: Building Public Support for America's Public Schools*; Enlightenment Press: Fairfield, IA, USA, 2010.
10. Ingersoll, R. Teacher turnover and teacher shortages: An organizational analysis. *Am. Educ. Res. J.* **2001**, *3*, 499–534. [CrossRef]
11. Ingersoll, R. *Is There Really a Teacher Shortage*; University of Pennsylvania: Philadelphia, PA, USA, 2003.
12. Ingersoll, R.; Perda, D. Is the supply of mathematics and science teachers sufficient? *Am. Educ. Res. J.* **2010**, *20*, 1–32. [CrossRef]
13. Ingersoll, R.; May, H. The magnitude, destinations and determinants of mathematics and science teacher turnover. *Educ. Eval. Policy Anal.* **2012**, *34*, 435–464. [CrossRef]
14. Ingersoll, R. *Teacher Supply, Demand, Shortages and Turnover in Rural Schools: What the National Data Tell Us*, 2021; in press.
15. Krugman, P.R. *End This Depression Now*; W.W. Norton & Co.: New York, NY, USA, 2012.
16. Moore, S. We've become a nation of takers, not makers. *Wall Street J.* **2011**, A15.
17. Scafidi, B. *The School Staffing Surge: Decades of Employment Growth in America's Schools*; Friedman Foundation: Indianapolis, IN, USA, 2013.
18. National Academy of Sciences. *Toward Understanding Teacher Supply and Demand*; National Academies Press: Washington, DC, USA, 1987.
19. National Commission on Excellence in Education. *A Nation at Risk: The Imperative for Educational Reform*; Government Printing Office: Washington, DC, USA, 1983.
20. National Commission on Teaching and America's Future. *What Matters Most: Teaching for America's Future*; National Commission on Teaching and America's Future: New York, NY, USA, 1996.
21. National Commission on Teaching and America's Future. *Doing What Matters Most: Investing in Quality Teaching*; National Commission on Teaching and America's Future: New York, NY, USA, 1997.
22. National Commission on Mathematics and Science Teaching for the 21st Century (the John Glenn Commission). *Before It's Too Late*; Government Printing Office: Washington, DC, USA, 2000.
23. National Research Council. *Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. Schools*; National Academies Press: Washington, DC, USA, 2002.
24. U.S. Department of Education. *Meeting the Highly Qualified Teachers Challenge: The Secretary's Annual Report on Teacher Quality*; U.S. Department of Education: Washington, DC, USA, 2002.
25. U.S. Department of Education. *Teacher Shortage Areas Nationwide Listing 1990–1991 to 2009–2010*; U.S. Department of Education: Washington, DC, USA, 2009.
26. National Academy of Sciences. *Rising above the Gathering Storm*; National Academies Press: Washington, DC, USA, 2007.
27. Garcia, E.; Weiss, E. The Teacher Shortage Is Real, Large and Growing, and Worse than We Thought. 2019. Available online: https://www.epi.org/files/pdf/163651.pdf (accessed on 8 May 2021).
28. Aldeman, C.; Rotherham, A. *Teacher Pension Plans: How They Work, and How They Affect Recruitment, Retention, and Equity*; Bellweather Education Partners: Washington, DC, USA, 2019.
29. Boivie, I. *The Three Rs of Teacher Pension Plans: Recruitment, Retention, and Retirement*; The National Institute on Retirement Security: Washington, DC, USA, 2011.
30. Ingersoll, R.; Merrill, L. Review of "Teacher Layoffs: Rethinking "Last Hired, First Fired" Policies". 2010. Available online: https://repository.upenn.edu/gse_pubs/218 (accessed on 8 May 2021).
31. Thomsen, J. Teacher Performance Plays a Growing Role in Employment Decisions. Education Commission of the States. 2014. Available online: http://www.ecs.org/clearinghouse/01/12/42/11242.pdf (accessed on 8 May 2021).
32. Henry, G.; Fortner, K.; Bastian, K. The effects of experience and attrition for novice high school science and mathematics teachers. *Science* **2012**, *335*, 1118–1121. [CrossRef] [PubMed]
33. Kane, T.J.; Rockoff, J.E.; Staiger, D.O. *What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City*; National Bureau of Economic Research: Cambridge, MA, USA, 2006.
34. Ingersoll, R.; Strong, M. The impact of induction and mentoring for beginning teachers: A critical review of the research. *Rev. Educ. Res.* **2011**, *81*, 201–233. [CrossRef]
35. Will, M.; Sawchuk, S. Teacher Pay: How Salaries, Pensions, and Benefits Work in Schools. 2018. Available online: https://www.edweek.org/teaching-learning/teacher-pay-how-salaries-pensions-and-benefits-work-in-schools/2018/03 (accessed on 8 May 2021).
36. Snell, R.K. Pensions and Retirement Plan Enactments in 2011 State Legislatures. National Conference of State Legislatures. 2012. Available online: https://www.ncsl.org/documents/employ/2011EnactmentsFinalReport.pdf (accessed on 8 May 2021).
37. Lohman, J. 50-State Survey of Certain Teacher Retirement Plan Provisions. 2000. Available online: https://www.cga.ct.gov/2000/rpt/2000-R-1110.htm (accessed on 31 March 2021).
38. Tyack, D. *The One Best System*; Harvard University Press: Cambridge, MA, USA, 1974.
39. Reich, A. Teaching is a Good Profession for A Woman. In *And Jill Came Tumbling After*; Stacey, J., Bereaud, S., Daniels, J., Eds.; Dell: New York, NY, USA, 1974; pp. 337–343.
40. Strober, M.; Tyack, D. Why do women teach and men manage? *Signs* **1980**, *5*, 499–500. [CrossRef]
41. *Hoffman, N Women's "True" Profession: Voices from the History of Teaching*; Harvard Education Press: Cambridge, MA, USA, 2003.

42. Clifford, G. *Those Good Gertrudes: A Social History of Women Teachers in America*; Johns Hopkins University Press: Baltimore, MD, USA, 2016.
43. Perlmann, J.; Margo, R.A. *Women's Work?: American Schoolteachers, 1650–1920*; The University of Chicago Press: Chicago, IL, USA, 2010.
44. Lortie, D. *School Teacher*; University of Chicago Press: Chicago, IL, USA, 1975.
45. Goldstein, D. *The Teacher Wars*; Anchor Books: New York, NY, USA, 2014.
46. Moulthrop, D.; Calegari, N.; Eggers, D. *Teachers Have It Easy: The Big Sacrifices and Small Salaries of America's Teachers*; The New Press: New York, NY, USA, 2005.
47. Grant, G. The Teacher's Predicament. *Teach. Coll. Rec.* **1983**, *84*, 593–609.
48. Flyer, F.; Rosen, S. The New Economics of Teachers and Education. *J. Labor Econ.* **1997**, *15*, S104–S139. [CrossRef]
49. Shelton, J. Women's Work, Women's Walkouts. 2018. Available online: https://genderpolicyreport.umn.edu/womens-work-womens-walkouts/ (accessed on 8 May 2021).
50. Reskin, B.; Roos, P. *Job Queues, Gender Queues*; Temple University Press: Philadelphia, PA, USA, 1990.
51. Jacobs, J. Men in Female-Dominated Fields. In *Doing Women's Work: Men in Non-Traditional Occupations*; Williams, C., Ed.; Sage: Newbury Park, CA, USA, 1993; pp. 49–63.
52. Waller, W. *The Sociology of Teaching*; Wiley: New York, NY, USA, 1932.
53. Mills, C.W. *White Collar*; Oxford: New York, NY, USA, 1951.
54. Davis, J.; Smith, T. *General Social Surveys, 1972–1996: Cumulative Codebook*; National Opinion Research Center: Chicago, IL, USA, 1996.
55. Smith, T.; Son, J. *Measuring Occupational Prestige on the 2012 General Social Survey*; University of Chicago: Chicago, IL, USA, 2014.
56. Dolton, P.; Marcenaro-Gutierrez, O.D.; De Vries, R.; She, P.-W. *An International Index of Teacher Status*; The Varkey GEMS Foundation: London, UK, 2018.
57. Lortie, D. The balance of control and autonomy in elementary school teaching. In *The Semi-Professions and Their Organization*; Etzioni, A., Ed.; Free Press: New York, NY, USA, 1969.
58. Etzioni, A. *Semi-Professions and Their Organizations: Teachers, Nurses and Social Workers*; Free Press: New York, NY, USA, 1969.
59. Ingersoll, R.; Collins, G. The Status of Teaching as a Profession. In *Schools and Society: A Sociological Approach to Education*, 6th ed.; Ballantine, J., Spade, J., Stuber, J., Eds.; Sage: Berkeley, CA, USA, 2018; pp. 199–213.
60. Allan, J. Male Elementary Teachers. In *Doing Women's Work: Men in Non-Traditional Occupations*; Williams, C., Ed.; Sage: Newbury Park, CA, USA, 1993; pp. 113–127.
61. Jacinto, A.; Gershenson, S. The intergenerational transmission of teaching. *Am. Educ. Res. J.* **2020**. [CrossRef]
62. American Association of School Administrators. *The American School Superintendent: 2020 Decennial Study*; American Association of School Administrators: Alexandria, VA, USA, 2021.
63. U.S. Department of Education. *Digest of Education Statistics*; National Center for Education Statistics: Washington, DC, USA, 2019.
64. Grissom, J.A.; Reininger, M. Who comes back? A longitudinal analysis of the reentry behavior of exiting teachers. *Educ. Financ. Policy* **2012**, *7*, 425–454. [CrossRef]
65. Pietrykowski, B. The Return to Caring Skills: Gender, Class, and Occupational Wages in the US. *Fem. Econ.* **2017**, *23*, 1–30. [CrossRef]
66. Achinstein, B.; Aguirre, J. Cultural match or cultural suspect: How new teachers of color negotiate socio-cultural challenges in the classroom. *Teach. Coll. Rec.* **2008**, *110*, 1505–1540.
67. Carver-Thomas, D. *Diversifying the Teaching Profession: How to Recruit and Retain Teachers of Color*; Learning Policy Institute: Palo Alto, CA, USA, 2018.
68. Campoli, A.K. Supportive principals and black teacher turnover: ESSA as an opportunity to improve retention. *J. Sch. Leadersh.* **2017**, *27*, 675–700. [CrossRef]
69. Irvine, J.J. An analysis of the problem of the disappearing Black educator. *Elem. Sch. J.* **1988**, *88*, 503–514. [CrossRef]
70. Ladson-Billings, G. Toward a theory of culturally relevant pedagogy. *Am. Educ. Res. J.* **1995**, *32*, 465–491. [CrossRef]
71. Lewis, C.W.; Toldson, I. *Black Male Teachers: Diversifying the United States' Teacher Workforce*; Emerald Group: Bingley, UK, 2013. [CrossRef]
72. Rogers-Ard, R.; Knaus, C.B.; Epstein, K.K.; Mayfield, K. Racial diversity sounds nice; systems transformation? Not so much: Developing urban teachers of color. *Urban Educ.* **2013**, *48*, 451–479. [CrossRef]
73. Villegas, A.M.; Strom, K.; Lucas, T. Closing the racial-ethnic gap between students of color and their teachers: An elusive goal. *Equity Excell. Educ.* **2012**, *45*, 283–301. [CrossRef]
74. Albert Shanker Institute. *The State of Teacher Diversity in American Education*; Albert Shanker Institute: Washington, DC, USA, 2015.
75. Torres, J.; Santos, J.; Peck, N.L.; Cortes, L. *Minority Teacher Recruitment, Development, and Retention*; Brown University, Educational Alliance: Providence, RI, USA, 2004.
76. Villegas, A.M.; Lucas, T. Diversifying the teacher workforce: A retrospective and prospective analysis. In *Developing the Teacher Workforce (103rd Yearbook of the National Society for the Study of Education, Part 1)*; Smylie, M.A., Miretzky, D., Eds.; University of Chicago Press: Chicago, IL, USA, 2004; pp. 70–104.
77. Zumwalt, K.; Craig, E. Teachers' characteristics: Research on the demographic profile. In *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*; Cochran-Smith, M., Zeichner, K.M., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2005; pp. 111–156.
78. Ingersoll, R.; May, H.; Collins, G. Recruitment, employment, retention and the minority teacher shortage. *Educ. Policy Anal. Arch.* **2018**, 27. [CrossRef]

79. Ingersoll, R.; May, H.; Collins, G.; Fletcher, T. Trends in the Recruitment, Employment and Retention of Teachers from Under-Represented Racial-Ethnic Groups, 1987 to 2016. In *AERA Handbook of Research on Teachers of Color and Indigenous Teachers*; Bristol, T., Gist, C., Eds.; American Educational Research Association: Washington, DC, USA, 2021.
80. Gorard, S.; Maria Ventista, O.; Morris, R. Who Wants To Be a Teacher? Findings from a Survey of Undergraduates in England. 2021. Available online: https://www.tandfonline.com/doi/full/10.1080/03055698.2021.1915751 (accessed on 8 May 2021).
81. Atkin, M. Who Will Teach in in High School? *Daedalus* **1981**, *110*, 91–104. [CrossRef]
82. Bacolod, M.P. Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *Rev. Econ. Stat.* **2007**, *89*, 737–751. [CrossRef]
83. Corcoran, S.; Evans, W.; Schwab, R. Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *Pap. Proc. Am. Econ. Rev.* **2004**, *94*, 230–235. [CrossRef]
84. Gitomer, D.H. *Teacher Quality in a Changing Policy Landscape: Improvements in the Teacher Pool*; Educational Testing Service: Princeton, NJ, USA, 2007.
85. Barron. *Barron's Profiles of American Colleges*, 28th ed.; Barron: Hauppauge, NY, USA, 2009.
86. Murnane, R.; Singer, J.; Willett, J.; Kemple, J.; Olsen, R. (Eds.) *Who Will Teach? Policies That Matter*; Harvard University Press: Cambridge, MA, USA, 1991.
87. Borman, G.D.; Dowling, N.M. Teacher attrition and retention: A meta-analytic and narrative review of the research. *Rev. Educ. Res.* **2008**, *78*, 367–409. [CrossRef]
88. Carver-Thomas, D.; Darling-Hammond, L. The trouble with teacher turnover: How teacher attrition affects students and schools. *Ed. Policy Anal. Arch.* **2019**, 27. [CrossRef]
89. Guarino, C.; Santibanez, L.; Daley, G. Teacher recruitment and retention: A review of the recent empirical literature. *Rev. Educ. Res.* **2006**, *76*, 173–208. [CrossRef]
90. Johnson, S.M.; Berg, J.; Donaldson, M. *Who Stays in Teaching and Why: A Review of the Literature on Teacher Retention*; Harvard Graduate School of Education: Cambridge, MA, USA, 2005.
91. See, B.H.; Morris, R.; Gorard, S.; Kokotsaki, D.; Abdi, S. Teacher Recruitment and Retention: A Critical Review of International Evidence of Most Promising Interventions. *Educ. Sci.* **2020**, *10*, 262. [CrossRef]
92. Lankford, H.; Loeb, S.; Wyckoff, J. Teacher sorting and the plight of urban schools: A descriptive analysis. *Educ. Eval. Policy Anal.* **2002**, *24*, 37–62. [CrossRef]
93. Barnes, G.; Crowe, E.; Schaefer, B. *The Cost of Teacher Turnover in Five School Districts*; National Commission on Teaching and America's Future: Washington, DC, USA, 2007.
94. Levy, A.J.; Ellis, P.; Joy, L.; Jablonski, E.; Karelitz, T. *The Causes and Costs of Turnover for Science and Other Teachers*; American Educational Research Association: Denver, CO, USA, 2010.
95. Milanowski, A.; Odden, A. *A New Approach to the Cost of Teacher Turnover*; Center on Reinventing Public Education: Seattle, WA, USA, 2007.
96. Merrill, E. An Analysis Of The Teacher Turnover Policy Problem In New York City's Middle Schools. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2014.
97. Ronfeldt, M.; Loeb, S.; Wykoff, J. How teacher turnover harms student achievement. *Am. Educ. Res. J.* **2013**, *50*, 4–36. [CrossRef]
98. Ingersoll, R.; Perda, D. *How High Is Teacher Turnover and Is It a Problem*; University of Pennsylvania: Philadelphia, PA, USA, Forthcoming.
99. Holme, J.J.; Jabbar, H.; Germain, E.; Dinning, J. Rethinking Teacher Turnover: Longitudinal Measures of Instability in Schools. *Educ. Res.* **2018**, *47*, 62–75. [CrossRef]
100. Papay, J.P.; Bacher-Hicks, A.; Page, L.C.; Marinell, W.H. The Challenge of Teacher Retention in Urban Schools: Evidence of Variation from a Cross-Site Analysis. *Educ. Res.* **2017**, *46*, 434–448. [CrossRef]

education sciences

MDPI

Article

Survey Methods to Identify Risk of Attrition: Measures of Career Intention and Regret

Mark Carver

School of Education, University of Strathclyde, Glasgow G1 1XQ, UK; mark.carver@strath.ac.uk

Abstract: The common measure of teacher retention as snapshots of those employed in state-funded schools may overestimate attrition by failing to consider a desire for flexibility in contemporary teaching careers. When used as a measure of the effectiveness of teacher education, an over-emphasis on classroom teacher supply may also narrow the curriculum to teacher training rather than the more expansive 'learning teaching'. This paper discusses two 'softer' measures of retention, career intention and training regret, to give a more general sense of how contemporary teachers see their career development as relating to their initial teacher education and professional learning. These measures are generated by adapting survey questions from the OECD's TALIS and the US' Beginning Teacher Longitudinal Study, simply asking teachers where they see themselves in five years' time and if they would still choose to become a teacher if they could go back to before they began training. Surveys were administered annually to two cohorts of recent graduates as part of the Measuring Quality in Initial Teacher Education project—three data captures for 2018 graduates, two for 2019 graduates. It is shown how these measures help to mitigate declines in survey response and can give some helpful estimates of teacher attrition with respect to sex, ethnicity, school type, and degree type. The alternative measures are also argued to give helpful indicators of attrition risk before it happens, allowing discussion of how teachers' career intentions change during their early careers. In particular, it is found that leaving the classroom is a fairly common expectation, but not necessarily because of teacher burnout. It is suggested that asking what teachers can imagine themselves doing is an effective measure for engaging with issues around vocational choice and teaching as a lifelong profession, with implications for how careers in education are conceptualised in initial teacher education programmes.

Keywords: teacher retention; wastage rate; attrition; teachers' careers; early career teachers; initial teacher education

Citation: Carver, M. Survey Methods to Identify Risk of Attrition: Measures of Career Intention and Regret. *Educ. Sci.* **2021**, *11*, 617. https://doi.org/10.3390/educsci11100617

Academic Editors: See Beng Huat, Rebecca Morris and Albright Jim

Received: 11 August 2021
Accepted: 1 October 2021
Published: 9 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The supply of enough high-quality teachers is regarded as a perennial issue "in all industrialised countries" [1] (p. 202), with regular shortages of teachers in particular subject areas, with certain skills, in rural areas, or because of a need to maintain target class sizes as birth rates fluctuate [2]. In part, this is an issue of forward planning and financing: the lead-in time for educating teachers means that changing priorities can take a while to address, while maintaining class sizes during peaks in birth or immigration rates can mean having too many teachers when the surge passes and risk teacher unemployment.

One of the most common uses of retention figures is to help policymakers estimate the numbers and types of available teachers to determine the allocation of training places, incentives for new or serving teachers, and related policies around class size or teacher workload. Some of these measures can also be used to judge the quality of initial teacher education (ITE), particularly with notions such as early career teacher resiliency being used to compare different ITE providers or routes. While a range of retention measures can be used, each with their own emphasis—retention, attrition, wastage rate, vacancy rate—these tend to see teacher supply at the system level rather than considering what

they tell us about the needs and desires of individual teachers. This is most apparent in the 'wastage rate' measure, which counts any teacher not in the national state-funded school sector (except for short-term or maternity leave) as 'wasted', a term imbued with covert meaning and intent.

One helpful refinement to the standard headcount of teachers is to analyse vacancy rates. Analysis of vacancies in England and Wales has shown that there is not a teacher supply crisis in general, but rather the problem is a mismatch of supply and demand and a need to better predict demand in subject areas or geographical regions using a broader range of measures [3]. Analysis across OECD countries [4], likewise, found that teacher supply was much more about a qualitative shortage: while some countries actually had an oversupply of qualified teachers, all countries recorded shortages in key areas and high attrition among early career teachers. Working with vacancies at subject level, Sprigade [5] shows how adverts in the Times Educational Supplement related to the number of training places on offer in England creates an estimate of 'trainees per post': 0.59 for English indicating that more trainees should be recruited, while 3.58 for citizenship suggests an oversupply of teachers who may struggle to find jobs within their subject specialism. Similarly, measuring teachers in precarious employment or being underemployed (e.g., as teaching assistants or cover supervisors) can highlight the damaging impact of over-supply in some areas, including an analysis of career paths of new teachers in Scotland showing how teacher precarity can undermine professional learning, resulting in fragmented and less satisfying careers, or even questioning the notion that teaching *is* a career [6]. Interpretation of retention measures, therefore, needs to distinguish whether problems are related with not having enough teachers willing to perform the work, or enough schools willing to employ the available teachers. This is further complicated since retention tends to assume that teaching is a traditional 'job for life', creating a questionable interpretation that must be a negative reason for teachers leaving the profession before retirement [7].

Conceptualising teacher attrition in such broad terms also makes the scale of the retention problem appear dramatic. Recent figures for England, for example, suggest that as many as 250,000 teachers under the age of 60 are lost to the system: more than half the number of current practicing teachers [8]. However, this takes a particularly narrow view of what it means to not be in teaching since it only ignores those who retire, die, are on maternity leave, or who are barred from teaching. After that, it becomes a simple binary categorisation of whether a teacher is in a state-funded mainstream classroom: those who become teacher educators or teach abroad, in the private sector, in state schools, in pupil referral units, or in other educational institutions are all classified as 'out of service', giving little indication of how, where, or even if this is a problematic situation.

Aside from the issue of underemployed teachers, who are still working in schools, there appears to be little interest in what teachers are doing when they leave the profession, or even if they intended to leave or simply take a break. For instance, requesting data from the Scottish Teachers' Superannuation Scheme showed that 3126 new teachers registered with the scheme from April 2012 to March 2013, of whom 2194 remained in 2019, with a further 426 transferring to England's equivalent scheme. This data is not routinely published, but there is valuable insight in knowing that nearly half of the teachers leaving Scotland's classrooms are still doing the same jobs south of the border rather than being 'wasted'. Indeed, if teacher retention was measured UK-wide rather than at national level, five-year retention would be around 80%, a far less alarming figure than is generally reported. The number may indeed be higher if including those teaching in other UK nations, overseas, or in the private sector.

One notable area of progress is the changed methodology of Scotland's teacher census [9], which signals that the work of teachers is valued beyond their service in mainstream classrooms as the census includes all teachers in publicly funded primary, secondary, special schools, and early learning and childcare settings, expanding on typical teacher counts which only look at primary and secondary schools. Scotland's local authorities centrally employed teachers are also counted, which includes those in a general local authority

non-teaching role, short- and long-term supply teachers, peripatetic/visiting specialists, and those employed as home visiting tutors or in the hospital teaching service [10]. Such improvements to counting teachers give a clearer view of teachers working in the educational system, more broadly defined, and so gives a better estimate of figures such as teacher: pupil ratio and overall teacher retention across the profession.

Such added detail can generally be regarded as an improvement in workforce tracking, even if it makes comparisons with other data sources and other countries problematic as what is counted as a teacher can vary. However, the variety of ways to count teachers can lead to misuse as the issue becomes politicised. Weldon [11] makes an incisive critique of the commonly cited figure of 50% teacher attrition in the first five years, picking apart a trail of guesses, circular references, and unchecked assumptions in governmental reports and the academic literature to show that there is little robust data to support claims of a teacher retention crisis. Nevertheless, maintaining the narrative of crisis seems to serve neoliberal policies and reform in the name of efficiency as the concept of attrition is related to value-for-money in ITE.

This narrative can also include teacher preparation programme completion as well as in-service attrition. Such a measure of overall retention—from the start of ITE through to the early career—may have some value in showing how many of those who started a programme then remained as teachers (and thus giving an inference about the level of "teacher resiliency" developed on these programmes). However, competition between ITE providers illustrates the potential to use retention figures to mislead. For instance, Teach First claims a programme completion rate of 95%, which it reports as "uniquely high" compared with a PGCE average of 86% [12], although this is called into question by UK Government statistics showing the rate of qualified teacher status awarded to graduates of these programmes which show Teach First at 91%, university-led postgraduate programmes at 90%, and university-led undergraduate programmes at 93%. Teach First confusingly call this 'wastage' instead of 'non-completion', which a cynic may see as trying to present themselves favourably against the common assumption of 30–50% wastage [11]. The confusion increases as retention is reported in different ways: 42% "long-term retention", 54% who "remain teaching in the UK", by which they seem to mean England and Wales, 57% "still in teaching 5 years" after training, or 68% who "remain employed in education" [12]. This echoes earlier research finding that "fewer than half" of those admitted to teacher education programmes made it to their fifth year of teaching [2] (p. 25), but the range of measures perhaps says more about the political value of these figures rather than the realities of teachers in the labour market.

On a pragmatic level, having such a range of measures with confusingly similar names can obscure the issues and make it difficult for policymakers to find suitable measures. On a more philosophical level, the use of retention as a measure of the quality of ITE can more problematically lead to a narrative of pounds spent per year of a teacher being in a classroom. Thus, ITE is seen to be failing if the number of those remaining employed in schools falls and its primary purpose is assumed to be providing schools with teachers on a regular enough basis to mitigate seemingly inevitable attrition.

As well as adding nuance to the simple teacher headcount as in vacancy rate or teacher censuses, attempts have also been made to consider teachers' economic motivations. Sims shows how recruitment and retention must engage with a range of market conditions, showing a "perfect storm" [13] of a bulge in pupil numbers working its way through from 2015 to 2024, while teachers from the Baby Boom generation are retiring and the 30% increase in graduates training to be teachers, following the 2008 recession, potentially leaving teaching as the economy recovers.

Aside from economic motivations, individual teachers may also change careers due to "vocational maturity", in which "people become more clear about their assets and liabilities as well as about the opportunities and limitations of their job" [14] (p. 94). Rather than being 'wasted', these teachers experience attrition positively as they go on to new opportunities. For instance, while qualification level can be a crude proxy for teacher quality, US data

suggests that the most qualified teachers have "an 85% greater likelihood of leaving the profession than less selective graduates in the first three years of teaching" [15] (p. 648).

Along similar lines, analysis of working conditions, salaries, and the attraction of alternative employment points out a range of push and pull factors for teachers [16,17], where policy can be informed better by figures such as teacher turnover, resignation rates, or average years of service for teachers in a school or district. Other studies have invested in better tracking of teachers beyond the binary in/out of service, showing, for instance, that more than half of teachers leaving state schools actually stayed in the education sector, most commonly "teaching in private schools, becoming teaching assistants and taking up a non-teaching role in school" [18] (p. 10). Such reassurance echoes similar improvements in understanding of student retention in higher education more generally, where fears over high drop-out rates are greatly diminished when it is shown that the vast majority of students quickly return to study on a different programme [19], distinguishing the relatively low number who 'stop out' from those observed to 'drop out'.

Thus, the starting point for the current study was to seek a context-suitable measure [20] of the proportion of teachers who could be deemed to be putting their ITE to good use. The measure should consider value to the individual teacher rather than just to the state-school system as a measure of the quality of ITE provision. This is in line with Scottish Government funding of higher education, that study at a higher level is deemed a social good in general terms and should therefore not be treated in terms of narrow utility for a particular job. Specifically, the idea that teaching internationally should be counted reflects a broader commitment in national policy to internationalisation of the curriculum and diversity of the workforce while reflecting the argument, common in HE, that international work is considered an indicator of quality. Likewise, the concept of "vocational maturity" [14] (p. 94) was helpful in considering how teachers might see their capabilities as opening up a range of new career options, which again could be taken as a positive indication of ITE quality even if it led to higher teacher attrition figures. It was also desirable to find measures which could be used in surveys of teachers before they experience attrition. From discussing and comparing a range of possible measures in survey data, we conclude that two measures, teacher career intention and career regret, offer many advantages over current measures of retention and may usefully extend into measures of ITE quality.

2. Materials and Methods

Data comes from the Measuring Quality in Initial Teacher Education (MQuITE) Project, a six-year study funded by the Scottish Government and involving co-investigators from every ITE provider in Scotland. Currently in Year Four, the project has a broad remit to find and generate data suitable for measuring the quality of ITE in ways which suit the context of Scotland or identify where existing data already exists.

The MQuITE study began with a snapshot survey of 2018 graduates (n = 332) as they completed their ITE programmes, but later expanded to include 2019 graduates and a small group of 2020 graduates when two new universities were accredited for initial teacher education. From around 8000 survey invitations cascaded through institutional contacts and the General Teaching Council for Scotland, 197 teachers have remained in the study throughout every stage as of the most recent data capture in September 2021. A larger number of teachers have completed at least one survey, meaning that overall the study has gathered views from 572 teachers through 1551 individual survey completions. For the most recent data capture, the 197 responses came from 494 invitations sent to those who volunteered their email addresses at the end of a previous survey. A total 16 surveys were started and not completed, while 6 emails bounced. Response attrition from those invited is therefore 60%, although as a percentage of all those who started an earlier survey (n = 962) it is a much higher at 80%. This is a clear disadvantage to survey estimates of teacher retention, particularly when relying on cascade methods of final-year students who

may already be experiencing survey fatigue, compared with the much easier to collect headcount data from employment contracts.

Survey questions related to teacher retention were adapted from large-scale surveys, specifically the OECD's TALIS [21] and the US-based Beginning Teacher Longitudinal Study [22]. The OECD data is freely available online, so there was also the added benefit of enabling international comparisons. The questions included were:

- If you could go back to the start of university and start over again, would you become a teacher or not? [scale response]
- How long do you plan to remain in teaching? [multiple response checkbox options]
- At this point in time, where do you envisage yourself in 5 years' time? [multiple response checkbox options]

Other questions were helpful in making comparisons relevant to policy questions around retention, such as ethnicity and sex data being used to see if retention had an equalities impact and teachers' self-assessed competence levels and professional development needs to see if teacher learning or efficacy affected retention. Analysis used simple descriptive statistics to report percentages, with cross-tabulation identifying differences in means. Ethical approval was granted by the University of Edinburgh, where the project originated in 2017, with endorsement from each of the partner institutions and, most recently, refreshed approval from the University of Strathclyde in 2020 as it took overall responsibility for the project.

3. Results

This section reports on variables potentially related to teacher retention, including one—career regret—which enables international comparisons, before presenting overall responses to the career intention prompt. Career intention responses are then discussed with respect to how they might be used to consider the kinds of questions usually asked of headcount data.

3.1. Participant Responses

3.1.1. Estimates of Teacher Regret

While not a direct measure of teacher retention—indeed, it is a measure that gives no indication of a teacher's employment status—presentations of the emerging data at conferences and government committees have found that one measure resonates particularly well: "If you could go back to the start of university and start over again, would you become a teacher or not?" OECD's TALIS phrases this as "If I could decide again, I would still choose to work as a teacher" under the prompt "We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?", followed by a range of statements around the desire to change schools, whether teachers enjoy their work, etc. [21]. While it limits our ability to claim direct comparability, our slight rephrasing suited our context of a survey about initial teacher education, rather than about working conditions, as is the focus of TALIS. We refer to this affectionately as the 'do-over rate', turning the adjectival rating scale responses into simple percentages to give an indication of teacher career regret.

Across all survey cohorts at the point of graduation, only 6% of teachers indicated they would decide differently if given their choice again. This increases slightly, to 14%, when surveyed during their induction year, and then up again to 16% at the end of their first-year post-registration (i.e., their second year as a classroom teacher since graduating from university). Overall, this estimates early career regret at 13% and suggests a relatively positive view of the value of ITE programmes. By way of comparison, filtering the TALIS dataset to only include teachers at a similar career stage shows regret expressed by 27% of early career teachers across the OECD sample. The measure may even be rankable, showing some interaction between ITE quality and working conditions. For instance, Mexico leads with just 8% regretting their choice, Japan comes bottom at 45%, while

Scotland's neighbours are around the OECD mean: Finland with 21%, England 32%, Sweden 36%.

3.1.2. Intended Length of Service

Taking prompts from the BTLS, respondents were given a range of tick-box options to indicate how long they planned to remain in teaching. Response options included until retirement, qualifying for benefits, a better job, life events, as soon as possible, etc. While comparisons are not possible with BTLS in the same way as can be performed with TALIS data because BTLS is a restricted dataset, responses offer some insight into whether early career teachers share the view of retention statisticians that teaching is a lifelong career. For estimating intention to remain, responses were recoded into three main categories: until retirement/inability to work, until a life change or better job, leaving as soon as possible, and undecided. These summaries are presented in Table 1, below.

Table 1. Intention to remain by career stage.

Remain Until	On Graduation	During Induction	After Induction
Retirement or inability to continue work	48%	73%	73%
Until a life change or better job	2%	1%	3%
Leaving as soon as possible	2%	3%	2%
Undecided	3%	5%	3%

It, therefore, seems that the assumption underlying many retention measures, that teaching is a career through to retirement, is still widely held by early career teachers, although it is by no means a universal expectation. In particular, there appears to be some doubt upon graduation, although this rapidly improves once teachers start work. It also suggests that very few teachers see themselves as likely to be out of the classroom in the short-term, highlighting a need to better understand the gap between their expectations and what seems to be borne out in teacher census data.

3.1.3. Career Intention

Table 2, below, shows career intention by career stage. As with the 'intention to remain' options, these largely followed the BTLS prompts but were rephrased to use more locally common terms.

Table 2. Career intention by career stage.

Intention	On Graduation	During Induction	After Induction
(a) Class/subject teacher in Scotland	72%	67%	63%
(b) Middle leader	35%	33%	37%
(c) School leader/headteacher	4%	2%	1%
(d) Not teaching at all	7%	9%	10%
(e) Teaching out with Scotland	23%	18%	16%
(f) Working in further or higher education	7%	8%	7%
(g) Working in education but not teaching	11%	12%	11%
(h) Studying for, or having achieved, a Master's Degree in Education	32%	24%	23%
(i) Studying for, or having achieved, a doctorate in education, e.g., PhD or EdD	4%	3%	3%
(j) Studying for, or having achieved a master's or doctorate in a non-education-related field	3%	2%	2%
(k) Other	3%	5%	3%

The first response, those seeing themselves as teaching in classrooms in Scotland, matches closely with teacher retention figures, the teacher census, and the pension participation figures previously discussed. This suggests that the career intention question may be a helpful indicator of which teachers may indeed leave the classroom which, by virtue of being a survey item, may be beneficial in predicting attrition. It is also noteworthy to see the large number of respondents who thought they may be teaching out with Scotland, which in follow-up interviews seemed to include overseas (the Middle East being the most common), but in the most part referred to other UK nations. There are also small but significant numbers of EU citizens who take ITE programmes in Scotland which, prior to 2021, received a fee waiver from the Scottish Government, so they may be more likely to anticipate returning to teach in their own country, particularly following Brexit.

It can also be seen that the proportion of respondents who think they might not be teaching at all in five years' time is low, below 10%, but increases each year. Some response bias can be assumed for the survey, although it is issued to private email addresses so can include teachers who have left Scotland. There are also high numbers for those considering further study, and the 'other' responses included many education-related ideas (e.g., church minister, children's charity workers, etc.). Since the majority of response options might be considered relevant to an education-related career, aggregating responses is helpful for providing a summary. Ignoring 'other', since responses were so varied, responses were grouped into 'teaching related' (responses a, b, c, e, f), 'education-related' (g, h, i), and 'out of service' (responses d or j). These are summarized below in Table 3.

Table 3. Aggregated career intention responses.

Intention Category	On Graduation	During Induction	After Induction
I. Teaching related	95%	96%	87%
II. Education related	40%	26%	37%
[neither I nor II]	*3%*	*5%*	*7%*
III. Out of service	8%	22%	14%

Creating an additional 'neither teaching-related nor education-related' category to compare with 'out of service' helps to show how many teachers saw themselves as open to several options, which included some of those who also chose 'out of service' responses. That is to say, many of those teachers who could imagine a future where they would be counted as 'out of service' in census data could still see other options in teaching-related or education-related roles. While this number still increases year-on-year, it is again much smaller than might be assumed in the context of retention as it is commonly reported.

3.2. Relationships between Retention Measures

Aside from survey data helping to anticipate and perhaps mitigate attrition before it occurs, having survey questions that serve as reasonable proxies of attrition offers the potential for analysing relationships with other variables. For instance, policymakers may find it helpful to be able to identify teachers who are categorised as 'retained' but perhaps feel trapped and would prefer to leave if they felt they could. However, one major limitation in conducting such analysis is that the numbers expressing regret were so low that a much larger sample or a survey with better participant retention would be necessary for running any inferential statistics (e.g., Table 4).

Table 4. Example of cross-tabulation interpretation for ratings upon graduation (n = 154).

		Intend to Remain Teaching as Long as Able (At Graduation)	
		No (n = 30)	**Yes (n = 124)**
Regret becoming a teacher (at graduation)	Regrets (n = 9)	9 (30%)	0 (0%)
	No regrets (n = 145)	21 (70%)	124 (100%)

While such small numbers preclude inferential statistics, the crosstabulations between related items suggest that there is some nuance and therefore value in asking both questions. For instance, Table 4 shows that there were only nine respondents who expressed regret at becoming teachers, and that, unsurprisingly, none of these respondents thought they would remain in teaching for as long as they were able. However, these are not all the respondents who felt they would not remain in teaching, with a further 21 respondents choosing this option. This means that, even though 100% of those intending to remain as long as they were able also expressed no regrets at becoming teachers, both questions still have value in looking for nuances in the experiences of those small numbers of teachers not intending to remain or expressing career regret.

To help explore the career regret variable further, a new variable was created for any teachers who responded to any of the three annual surveys that they 'certainly would not' or 'probably would not' become a teacher if they had their time again. Of the 485 teachers who answered this question, 60 (12%) expressed regret, 2 (.5%) expressed regret in one data capture but not another, and 423 (87%) never expressed regret. The same can be conducted with the intention to remain until retirement or as long as they are able. Of the 458 responses to this question across all data captures, 268 (59%) said throughout that they intended to remain, with 190 (42%) saying that they did not. In this case, where the same teacher stayed in the sample from one year to the next, their response was unchanged. Collapsing the 'regret' variable into a dichotomous response allows the cross-tabulation shown in Table 5, below.

Table 5. Cross-tabulation of regret and intention to remain variables across all data captures (n = 329).

		Intend to Remain Teaching as Long as Able	
		No (n = 72)	**Yes (n = 257)**
Regret becoming a teacher	Regrets (n = 40)	33 (46%)	7 (3%)
	No regrets (n = 289)	39 (54%)	250 (97%)

Looking across all years of data capture shows some interesting nuances—while the vast majority of those who express regret also responded that they did not intend to remain teaching as long as they were able (33 of 40 respondents; 83%), there were some exceptions. Seven respondents who regretted becoming teachers nevertheless intended to remain in the profession, while 39 respondents who did not regret becoming teachers still did not intend to remain in the profession. There is perhaps some insight possible from the 'remain until a specific life event (e.g., parenthood, marriage)' response. Of the 330 who responded, just 37 teachers (11%) agreed with the statement, but these made up 30% of the 'regrets' category and only 9% of the 'no regrets' category. Remembering how early in their careers these respondents were sampled, this could suggest that they quickly found out that teaching was not compatible with their life plans.

Comparing the 'regret' rating with the career intentions responses is perhaps the most illuminating analysis. Table 6, below, shows that almost all teachers expect to still be in a teaching-related or education-related role in five years' time. However, almost all of those who do not (87%) are also those who expressed regret at becoming teachers (13, compared with 2), suggesting that this phrasing of the question could be useful for those who still feel that they are using their teaching skills but do not see their future in the classroom. As a measure of the value of a student's initial teacher education experience, it is perhaps the intersection between these two questions that is most revealing, with just 13 of the 329 respondents (4%) feeling that they both regretted becoming teachers and did not expect to remain in the education profession.

Table 6. Cross-tabulation of regret and expectation to be in a teaching- or education-related role across all data captures (n = 329).

		Expect to Be in a Teaching or Education-Related Role in Five Years' Time	
		No (n = 15)	**Yes (n = 314)**
Regret becoming a teacher	Regrets (n = 40)	13 (87%)	27 (9%)
	No regrets (n = 289)	2 (13%)	287 (91%)

Regretting their choice of becoming teachers seems to be a good indicator of teachers who expect to be in different careers in the near future, whether education-related or not. However, it is not the full story, and there are still sufficient numbers of teachers who do not regret taking their ITE programmes but nevertheless can see themselves being 'out of service' within the next five years. This illustrates the potential of using such survey questions to inform policy related to teachers' careers, or for making judgements about the quality of ITE.

3.3. Other Factors to Cconsider

Working with estimates of attrition risk in survey data lacks the clarity of census data, but it does allow exploration of questions around teacher efficacy, professional development needs, or feelings of preparedness To illustrate one way of working with these new estimates, a new 'at risk' variable is created for those responses where something is wrong. This includes any respondents at any stage who regretted becoming teachers (n = 62), those who intended to leave teaching 'as soon as possible' (n = 16), and those who anticipate being out of service within the next five years (positive responses to the 'out of service' prompt, n = 61, or negative responses to all the 'teaching-related' and 'education-related' prompts, n = 26). These new variables can be compared with teachers' self-reported efficacy, professional learning need, efficacy with particular reference to managing behaviour, and their overall rating of competence taken from elsewhere in the MQuITE surveys. Summarised below in Table 7 as a percentage increase or decrease in the risk category, it can be seen, for instance in the first row, that those student teachers expressing career regret also have 8% lower efficacy ratings than those who do not express regret, rising to 10% for efficacy ratings relating to classroom management or pupil behaviour, and a 15% lower rating of their own competence, but no difference in their level of perceived need for further professional learning.

Table 7. Self-reported teacher competency changes by risk category.

Student Teacher Risk Category	Self-Efficacy (Mean of All Measures)	Self-Efficacy (Behaviour)	CPD Need	Competence
Career regret	−8%	−10%	0%	−15%
Not expecting to be teaching in a classroom in Scotland	+1%	+1%	−2%	+1%
'At risk' composite	−6%	−9%	0%	−13%
Not expecting to be in teaching-related or education-related roles	−7%	−5%	+1%	−13%

4. Discussion

This study set out to find context-appropriate measures related to teacher retention. Specifically, the context was a state-funded higher education system broadening the concept of 'in service' to consider a wider sense of what it means to make good use of a teacher's initial teacher education. There was also a desire to find a suitable retention-related measure that could be used in surveys before attrition occurs. One immediate practical advantage is that it is easier to survey teachers while they are at risk of attrition instead of after they have left, meaning that the influence of a broader range of variables can be considered. Finding

a suitable measure to include in surveys would also assist in making inferences about the quality of initial teacher education programmes, possibly extending beyond surveys of early career teachers to include student teachers during their teacher preparation at risk of in-programme attrition and helping to identify those teachers in need of support.

Comparing four different ways that teachers might be identified as 'at risk' helps to indicate the value of adapting measures to suit local context. For example, if the survey only relied upon the 'Do you expect to be teaching in a classroom in Scotland in five years' time?' measure, there would be no meaningful difference in any of the efficacy, CPD, or competence ratings between those teachers who answered yes and those who answered no (Table 7). As the MQuITE project continues and the dataset grows, intersectionality between variables may be feasible to analyse, and longitudinal analysis tracking the cohorts will be possible. At the moment, for example, there are not enough teachers from BME backgrounds to draw any conclusions.

It has conversely been shown that too broad a measure of 'making good use of your ITE' is also limiting. Taking a positive interpretation of retention including all teaching-related or education-related roles means that fewer than 3% of teachers are identified as at risk of attrition, which could erroneously imply that there is no problem. This lack of discrimination in the question, with over 97% of teachers seeing themselves in at least one teaching-related or education-related role in five years' time, can be mitigated by a composite measure of risk which includes all respondents identified in any of the other measures. However, this requires asking several questions of participants, each with a range of responses, and so would have to justify the extra data collection. In particular, questions adapted from the BTLS around staying to retirement, benefits eligibility, life event, etc., seemed less meaningful in the Scottish context.

As a balance between ease of responding and enabling further analysis, the best option seems to be a pared-down version of the career intention question from BTLS [22] and the career regret question from TALIS [21]. Qualitative follow-up may help in suggesting context-meaningful response options that would better support inferences around whether teachers are anticipating going into teaching-related or education-related roles and how or if these expectations relate to their views on teachers' working conditions or the quality of their teacher education. Rather than being a checkbox, more sophisticated responses such as ranking may also be worth exploring. Adding in questions about teachers' longer-term intentions–such as their intention to return–could help when discussing how to categorise teachers who gain valuable experience, they can bring back to have direct benefits in the classroom. Both the career intention and career regret questions may also easily adapt to surveys of student teachers during their ITE programmes, helping to identify where these feelings may originate or strengthen. Such a view of career intention as student teachers move through their education and early career could combine particularly well with recent work looking at the views of those who consider but do not ultimately choose to enter teacher education programmes [23]. A measure of career intention and ITE regret may help to broaden this focus further, engaging with those who perhaps always saw a teaching career as part of something else.

As a summary of increased risk factors, looking at intention and regret together suggests a greater attrition risk for teachers who rate their competency lower, particularly in terms of self-efficacy abilities from the TALIS list and, even more specifically, in terms of managing pupil behaviour. It also appears that professional learning needs as surveyed using the TALIS list have no relation to retention risk factors, running contrary to expectations.

More broadly, it is hoped that this study opens discussion around what counts as retaining a teacher in the profession. There remains the practical need for teachers to meet teacher-student ratios and for there to be enough high-quality teachers (however defined) able to teach the full curriculum in all areas of the country and meet the diverse needs of learners. However, governments may also wish to consider the education profession as extending beyond this. There may be a broader view of how society benefits from

having teachers taking on different roles, consideration of how teachers may move flexibly between different jobs within an education career, and even a strategic view of how flexible teaching careers can be supported to manage the ebb and flow of births and immigration. For example, there is currently renewed interest in sabbaticals for teachers in Scotland, which could be a powerful policy lever for managing teacher employment rates, giving policymakers an extra tool beyond ITE recruitment targets. As Gorard et al. [24] point out, it hardly seems wasteful for teachers to move from the state to the independent sector, or to teach in further or higher education. Likewise, teachers entering other voluntary or public service sectors may continue to derive benefits from their teacher preparation and provide benefits back to the state. Examples from our free text responses included a teacher setting up a road safety education charity, writing textbooks, and becoming a teacher educator. These would be counted as 'out of service' in current measures, so the statistic needs to be treated with caution. Similarly, Scotland's teacher education has a strong social justice and diversification agenda, so the number of teachers going to teach overseas may represent important contributions on a global scale even if these are a short-term loss to Scottish schools.

The range of figures discussed above suggest that there may be a convergence around a common estimate of teacher retention, leading to a general impression that attrition and teachers' own expectations of attrition are around 20% over the first five years. There is also some support for the assumption that attrition is more acute in the earliest years of a teacher's career. However, it is difficult to claim that there is a 'true' figure for teacher retention since the measure has become so politicised. Instead of arguing for any one single approach, transparency in methods and reporting may be better to ensure that appropriate estimates of retention are used for different policy contexts. The range of measures here also warrants caution when using conventional estimates of teacher retention, since around 75% of teachers typically categorised as 'out of service' may still be in education-related roles and experiencing what they regard as successful careers.

Considering the extent to which ITE intends to prepare students for a teaching career or an education career, or indeed whether that is a helpful distinction to make [25], may be helpful in deciding specifically how retention should be measured for making inferences about the quality of ITE provision. Indeed, future research asking some form of these questions during ITE may be able to create combined estimates of programme non-completion and teacher attrition. Given the changes in career expectations from one year to the next, it may also be prudent to consider career planning and advice as part of teacher preparation or early career professional learning.

Finally, the study acknowledges a limitation in sampling bias, that those who have left teaching already are less likely to remain in the survey cohort. This will inevitably produce an overly optimistic view of intention to remain. Nevertheless, particularly when thinking of intention to leave or sense of regret, this sample may still help to indicate career dissatisfaction or early signs of burnout and includes teachers who may wish to leave but feel unable to find suitable work elsewhere or are otherwise trapped by circumstances. These attitudinal measures can therefore supplement the 'harder' data of teachers' employment status, better informing debate around what it means for a teacher to make good use of their professional preparation.

Funding: This research was funded by the Scottish Government through the General Teaching Council for Scotland as part of the Measuring Quality in Initial Teacher Education Project (www.mquite.scot).

Institutional Review Board Statement: The study was conducted according to the British Educational Research Association guidelines and approved by the Institutional Review Board of the University of Strathclyde (UEC20/34) in May 2021 further to initial annual approval from Moray House School of Education and Sport, University of Edinburgh, June 2017 through 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data collection instruments are available through the project website, www.mquite.scot.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ladd, H.F. Teacher labor markets in developed countries. *Futur. Child.* **2007**, *17*, 201–217. [CrossRef] [PubMed]
2. Mcintyre, J. Why They Stayed: A Study of the Working Lives of Long Serving Teachers in Inner City Schools. Ph.D. Thesis, University of Nottingham, Nottingham, UK, May 2010.
3. See, B.H.; Gorard, S.; White, P. Teacher demand: Crisis what crisis? *Cambridge J. Educ.* **2004**, *34*, 103–123. [CrossRef]
4. McKenzie, P.; Santiago, P. *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*; OECD: Paris, France, 2005.
5. Sprigade, A. The Labour Market for Teachers in England: Teachers and Recruitment Difficulties in Schools. In *Governance in der Lehrerausbildung: Analysen aus England und Deutschland: Governance in Initial Teacher Education: Perspectives on England and Germany*; Kuhlee, D., van Buer, J., Winch, C., Eds.; Springer: Wiesbaden, Germany, 2015; pp. 99–110.
6. Hulme, M.; Menter, I. New professionalism in austere times: The employment experiences of early career teachers in Scotland. *Teach. Teach. Theory Pract.* **2014**, *20*, 672–687. [CrossRef]
7. Sharp, S.; Draper, J. Leaving the register: Scottish teachers lost to the profession, 1997-98. *J. In-Service Educ.* **2000**, *26*, 247–266. [CrossRef]
8. DfE. The size of the schools' workforce. National Statistics School Workforce in England: November 2017; Department of Education: London, UK, 2018; p. 4.
9. Scottish Government Teacher census supplementary statistics. Available online: https://www.gov.scot/publications/teacher-census-supplementary-statistics/ (accessed on 6 October 2021).
10. Scottish Government Centrally employed teachers explanatory diagram. Available online: https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2019/07/scottish-exchange-of-data-staff-census/documents/centrally-employed-teachers-explanatory-diagram/centrally-employed-teachers-explanatory-diagram/govscot%3Adocument/Cen (accessed on 6 October 2021).
11. Weldon, P. Early career teacher attrition in Australia: Evidence, definition, classification and measurement. *Aust. J. Educ.* **2018**, *62*, 61–78. [CrossRef]
12. Parliament UK Education committee: Further written evidence submitted by Teach First. Available online: http://www.publications.parliament.uk/pa/cm201012/cmselect/cmeduc/1515/1515we31.htm (accessed on 6 October 2021).
13. Sims, S. Essays on the recruitment and retention of teachers. Ph.D. Thesis, University College London, London, UK, July 2018. Available online: https://core.ac.uk/download/pdf/195304798.pdf (accessed on 6 October 2021).
14. Chapman, D.W.; Hutcheson, S.M. Attrition from teaching careers: A discriminant analysis. *Am. Educ. Res. J.* **1982**, *19*, 93–105. [CrossRef]
15. Kelly, S.; Northrop, L. Early Career Outcomes for the "Best and the Brightest": Selectivity, Satisfaction, and Attrition in the Beginning Teacher Longitudinal Survey. *Am. Educ. Res. J.* **2015**, *52*, 624–656. [CrossRef]
16. Waghid, Y.; Divala, J. Teacher mobility, 'brain drain', labour markets and educational resources in the Commonwealth. *Comp. A J. Comp. Int. Educ.* **2010**, *40*, 257–258. [CrossRef]
17. Kelly, N.; Cespedes, M.; Clarà, M.; Danaher, P.A. Early career teachers' intentions to leave the profession: The complex relationships among preservice education, early career support, and job satisfaction. *Aust. J. Teach. Educ.* **2019**, *44*. [CrossRef]
18. Lynch, S.; Worth, J.; Bamford, S.; Wespieser, K. *Engaging Teachers: NFER Analysis of Teacher Retention*; National Foundation for Educational Research: Slough, UK, 2016.
19. Yorke, M.; Longden, B. *The First-Year Experience of Higher Education in the UK*; Higher Education Academy: York, UK, 2008.
20. Kennedy, A.; Adams, P.; Carver, M. Measuring Quality in Initial Teacher Education in Scotland: A context-specific endeavour. In *Teacher Education Policy and Research: Global Perspectives*; Mayer, D., Ed.; Springer: Singapore, 2021.
21. OECD. TALIS teacher questionnaire. Available online: http://www.oecd.org/edu/school/TALIS-2018-MS-Teacher-Questionnaire-ENG.pdf (accessed on 6 October 2021).
22. Beginning Teacher Longitudinal Study. Available online: https://nces.ed.gov/surveys/btls/ (accessed on 6 October 2021).
23. Gorard, S.; Maria Ventista, O.; Morris, R.; See, B.H. Who wants to be a teacher? Findings from a survey of undergraduates in England. *Educ. Stud.* **2021**, 1–23. [CrossRef]
24. Gorard, S.; See, B.H.; Smith, E.; White, P. *Teacher supply: The key issues*; Continuum: London, UK, 2006.
25. Rippon, J.H. Re-defining careers in education. *Career Dev. Int.* **2005**, *10*, 275–292. [CrossRef]

education sciences

MDPI

Article

Teacher Recruitment and Retention: A Critical Review of International Evidence of Most Promising Interventions

Beng Huat See [1,*], Rebecca Morris [2], Stephen Gorard [1], Dimitra Kokotsaki [1] and Sophia Abdi [3]

[1] School of Education, Durham University, Durham DH1 1TA, UK; s.a.c.gorard@durham.ac.uk (S.G.); Dimitra.kokotsaki@durham.ac.uk (D.K.)
[2] Department for Education Studies, Warwick University, Coventry CV4 7AL, UK; rebecca.e.morris@warwick.ac.uk
[3] School of Education, University of Birmingham, Birmingham B15 2TT, UK; sophia46@hotmail.com
* Correspondence: b.h.see@durham.ac.uk

Received: 13 August 2020; Accepted: 17 September 2020; Published: 23 September 2020

Abstract: Background: A raft of initiatives and reforms have been introduced in many countries to attract and recruit school teachers, many of which do not have a clear evidence base, so their effectiveness remains unclear. Prior research has been largely correlational in design. This paper describes a rigorous and comprehensive review of international evidence, synthesising the findings of some of the strongest empirical work so far. Methods: The review synthesises a total of 120 pieces of research from 13 electronic databases, Google/Google scholar and other sources. Each study is weighted by strength of evidence. Results: The strongest evidence suggests that targeted money can encourage people into teaching but does not necessarily keep them in the teaching profession. The money needs to be large enough to compensate for the disadvantages of working in certain schools and areas, and competitive enough to offset the opportunity costs of not being in more lucrative occupations, and its effect is only short-term. Conclusions: Continuing professional development (CPD) and early career support could be promising approaches for retaining teachers in the profession, but the evidence for them is weak. There is no evidence that any other approaches work, largely because of the lack of robust studies.

Keywords: teacher recruitment; teacher retention; systematic review; causal evidence; interventions

1. Introduction

This paper provides a comprehensive review of international evidence synthesising the findings of some of the strongest empirical work on improving the recruitment and retention of school teachers.

Attracting and retaining qualified teachers is a persistent problem that has plagued many countries for decades. The trends in recruitment to initial teacher training are often associated with the national labour market condition and the relative attractiveness of other occupations. Shortages are more severe in some subjects especially for maths and physics where there is a high demand for graduates in these fields. Compounding this is the growing pupil population. In England, teacher demand has consistently outstripped supply [1]. Reportedly, more people are leaving teaching than ever before. Only 60% of teachers remained in state schools after five years and for 'high-priority' subjects like physics and maths, this five-year retention dropped to just 50% [2].

An adequate supply of qualified teachers is important for the provision of an effective education system. A shortage of teachers can have a detrimental effect on the life chances of children [3–7]. Many countries in Europe have reported a widespread shortage of teachers [8]. Teacher supply continues to be a challenge in Australia and New Zealand and, in England and the US, the teacher

shortage is predicted to get worse as the pupil population rises. Across each of these contexts, the extent of the shortages can vary depending on geographical region, subject area, age of student and school types.

There are a number of factors which may influence a shortage of people being recruited into the teaching profession. These predominantly relate to people recognising and opting for what they perceive to be more favourable alternative career options. From an individual's perspective, these decisions may be influenced by the financial rewards available (e.g., salary, prospect of bonuses) or by their understanding of what the role entails (e.g., required tasks, working conditions, level of autonomy). Economic and employment cycles can have an impact on the number of people choosing teaching as a career, with more people seeing it as an attractive option during times of economic uncertainty [9–13]. It is also the case that government policies could influence teacher recruitment too, e.g., through funding and allocation of training places, the development of training routes, or marketing strategies [14].

The shortage of teachers is reportedly also partly the result of people leaving the profession prematurely. Teaching has often been characterised as an occupation with a high level of turnover especially among new teachers [15–17]. While all occupations experience some degree of turnover and career change, turnover in teaching is considered high particularly in the first few years compared to many other professions, such as lawyers, engineers, architects and professors [18,19]. In the US, it has been reported that around 40 to 50 percent of new teachers leave within the first five years of entry into teaching [20]. In England, the attrition rates are similar, particularly in maths, science and languages subjects [21]. Among the secondary teachers who qualified in 2010–2012 around 66% stayed on in state-funded schools in the fifth year [2] (Table 8). Government data show that the odds of leaving are higher for newly qualified teachers (NQTs) and those with stronger academic backgrounds [22].

In countries experiencing teacher shortages, numerous policy initiatives have been introduced in an attempt to address recruitment and retention issues and the factors which contribute to them. Many strategies involve financial incentives such as increased pay for teachers (e.g., for those teaching certain subjects or in particular areas) as well as bursaries or scholarships designed to attract more people into the profession, or to keep them there once they have qualified. In England, for example, there is a long history of providing tax-free bursaries and maintenance grants/loans for those entering training, and additional 'early career' payments for those continuing in the role after completing their NQT year. The amount available is, in theory, associated with the level of shortage. Similar approaches can be found in other countries experiencing teacher shortages. In the US, there have also been more widespread interest in variable salaries for those working in certain geographical areas, or in schools with higher proportions of disadvantaged pupils, as well as some initiatives which seek to reward teachers financially based upon their performance [23–25].

In more recent years, some policymakers have moved beyond financial incentives for retaining teachers. A growing awareness of the often challenging working conditions associated with teaching [26,27] has led to the development of strategies to try and improve these, and in turn make teaching a more attractive profession. These include induction programmes, access to professional development, enhancing leadership skills in schools, flexible working. In England, the Early Career Framework (ECF) [28], introduced in 2020, is part of the wider Teacher Recruitment and Retention Strategies [29] to support new teachers during the induction years via high-quality mentoring and professional development, and a reduced teaching timetable.

An increasingly important issue is teachers' workload. Correlational studies indicate that teachers' perceptions of workload are strong predictors of their decision to leave teaching [30–32]. In England, a report published by the DfE [27], based on interviews with 101 former teachers, suggested workload as the most important factor influencing teachers' decision to leave the profession. A recent survey of teachers and those who have left the profession found that 'workload' and 'improving work-life' balance were the most cited reasons for exiting teaching [33]. An earlier report based on a survey of over 1000 teachers also identified workload, policy changes and accountability pressure among the top

reasons for teacher attrition [34]. The DfE launched the Workload Strategy in 2014 to understand and address 'unnecessary' tasks that teachers undertake in the course of their duty. There is no evidence available on the extent to which schools have engaged with the recommendations put forward from this consultation, but the Teacher Workload Survey carried out in 2019 showed a reduction in teachers' reported working hours since 2016. However, the majority of teachers still reported workload as a serious problem.

2. Background

2.1. Common Approaches Used to Improve Teacher Recruitment and Retention

2.1.1. Financial Incentives (Including Scholarships, Bursaries, Higher Wages)

Financial incentives are commonly used to try and improve recruitment and retention. These are premised on the assumption that if sufficiently well compensated, people can be encouraged to go into teaching or be persuaded to stay on in the profession. Identifying what a well-designed pay incentive should be is difficult because of the numerous challenges and parameters that need to be considered [35]. Some commentators have suggested that effective incentive plans must offer relatively large awards to induce behavioural changes [36,37]. A number of American studies have pointed to the level of financial incentives needed in different contexts. Goodnough and Kelly [38] suggested that teacher salaries in New York be increased by up to 25% in the lowest-performing schools as the 15% increase that was offered in 39 of those schools appeared to have little impact in terms of attracting qualified teachers. Boyd et al. [39] and Hanushek et al. [40] estimated that considerable pay rises (up to 50%) may be needed to induce more teachers to work in schools with high proportions of ethnic minority or socioeconomically disadvantaged students. However, if working conditions and the relative attractiveness of the schools can be improved, then the size of the pay increase may need not be as large.

Others have argued that a single pay scale does not provide incentives for teachers with skills that are in high demand in non-teaching fields. Successive governments in England have offered more to trainees in certain subjects. However, once qualified, teachers then tend to enter the profession at a similar pay level irrespective of subject area. Recent reports have recommended that teachers in shortage subjects be paid a 'salary supplement' to encourage their retention, particularly in the early years of their careers [41,42]. Increased flexibility of pay was also introduced in England in 2013–2014 meaning that schools no longer had to use the seniority-based national pay scale but instead could determine annual pay awards by 'performance'. Studies have suggested limited impact on teacher mobility or retention in the same schools thus far [43–45]. There is very little evidence to indicate whether performance-related pay works either in improving teachers' performance or retaining them within the profession [46].

2.1.2. Alternative Routes into Teaching

Another approach often used to address the critical shortage of teachers is alternative certification or alternative pathways into teaching. These offer options different to the 'standard' or 'traditional' routes within a particular region or country and often provide ways into the profession for those wishing to train 'on the job' or who are working in other careers or roles (e.g., Troops to Teachers in England, or routes permitting teaching assistants to qualify as teachers). Traditional teacher-preparation programmes tend to emphasise pre-service training on the assumption that the learning and practical experiences that trainees engage with will give them the requisite skills and knowledge needed for success in the classroom. Alternative programmes may try to reduce barriers to entry and/or aim to enable teachers to enter the classroom more quickly (e.g., Teach First, School Centred Initial Teacher Training programme in the UK; Teach for America, the Teacher Residency Programs and Peace Corps Program in the US).

Many studies have been conducted to evaluate the effectiveness of different teacher preparation routes, but most focus on outcomes relating to teacher performance with mixed results [47,48]. Typically, these studies assess the relationship between certain attributes and qualifications of teachers and teacher performance (usually measured using students' performance as a proxy). There has been less research on the effects of teacher preparation for teacher recruitment or retention.

2.1.3. Induction Programmes and Mentoring

In recent years, there has been growing policy interest in induction and mentoring strategies. These are designed to provide additional support and development for teachers, usually in the early years of their career, with a view to retaining them within the profession [49,50]. While there is a large body of research on mentoring/induction programmes that has purportedly examined the "impact" of induction and mentoring on teacher retention, most are limited to single-group causal comparative analysis, correlating teachers' participation in these programmes with their self-reported intention to stay in teaching [51]. A systematic review on the role of mentors on retention of newly qualified teachers could not find conclusive evidence of a positive impact [52]. Only three studies within the review reported positive effects, but all were correlational studies (not based on experimental designs). The report called for closer scrutiny of the relationship between induction and retention and highlighted the need for more robust and reliable research in this area. Moreover, given the often complex or multi-faceted nature of induction/mentoring programmes, it can sometimes be difficult to understand which of the mechanisms or 'ingredients' within them are likely to drive any impact on retention. Further high-quality research focusing on these areas is particularly needed at present in order to inform the development and implementation of new policies on induction and mentoring (see e.g., the Early Career Framework in England [28]

2.1.4. Professional Development

Investing in high-quality professional development is widely believed to be an effective way of improving both teachers' and, in turn, students' performance [53]. More recently, however, professional development is also being considered as a method for improving teachers' satisfaction with their job and potentially reducing their workload. This, it is hoped, may lead to increased retention [54]. There are numerous professional development opportunities available to schools and teachers at present. However, there is considerable variation in the aims and quality of such programmes and teachers' access and engagement with them. Further, there is very little robust evidence which points towards an impact on retention [55,56]. A study by Allen and Sims [57] in England indicated that high-quality subject-specific professional development may be beneficial for retaining teachers. The study examined teachers' engagement with the National STEM (Science, Technology, Engineering and Mathematics) Learning Network development scheme, finding that while participants were no more likely to stay at their current school, they were more likely to stay in the profession for the first and second year after taking the courses. Recent analyses by Worth and van den Brande [58] found an association between teachers' reported autonomy over their professional development and their intention to stay in teaching. While not established as a causal relationship, this finding does suggest that some experimental work on this issue would be helpful for understanding whether increasing teachers' autonomy in relation to professional development might be a potential strategy for improving retention.

2.1.5. Leadership Support

Some studies have highlighted the importance of school environment factors for teacher retention, with school leadership often being viewed as influential in determining the ethos and working conditions within a school. A series of observational studies point to teachers' perceptions of administrative support and leadership as being strong predictors of teachers' intention to leave [59–61]. Johnson, Kraft and Papay [62] argue that while working conditions generally appear to be important to

teachers and their future career plans, it is the social conditions which form part of these—such as the principal's leadership, school culture and relationships with colleagues—which are most influential. Analysis in England, based upon the international TALIS dataset, also highlights the importance of good leadership. Sims [63] found that better school leadership is associated with higher job satisfaction for teachers and a reduction in the odds that they would want to leave their school.

2.1.6. Additional Incentives

In addition to the financial incentives noted above and school working conditions, research has looked at other incentives to encourage teacher recruitment and/or retention, including offering below market rental rates, living allowances (e.g., London living allowances) and discounted housing in certain areas. Examples from Australia and America include housing subsidies or offering rental accommodation at below-market rate for teachers willing to work in rural areas [64]. Unfortunately, the evidence on these kinds of incentives is limited and is often based on small-scale descriptive work or tangential research about wider compensation. Unlike for more direct financial incentives and wage compensation, there have been no rigorous evaluations of housing incentives to determine if they work in improving recruitment and retention especially in hard-to-staff areas, as Anne Podolsky at the Learning Policy Institute acknowledged [64].

2.2. Previous Reviews of the Literature

To the best of our knowledge, there have been no large-scale comprehensive reviews on teacher recruitment and retention policies, with quality appraisal of individual studies included as a key feature. Most have been narrative reviews of available literature [65,66] or focused on particular issues or groups of teachers/schools. Przygocki [67], for example, looked specifically at teacher supply issues in Catholic schools, while Fore et al. [68] and Billingsley [69] were concerned with the shortage of special education teachers. Previous reviews have also tended not to be systematic, instead summarising a collection of studies seeking the consensus view [70,71] or providing a discussion of a small number of key articles [72–74]. Borman and Dowling's [75] comprehensive review focused on the factors that moderate attrition outcomes rather than examining policy initiatives or interventions designed to improve retention.

A review by Guarino et al. [76] examined the individual and school characteristics linked to teacher recruitment and retention, as well as synthesising the evidence for a range of policies and initiatives aiming to reduce the shortage. The authors apply four quality criteria based on sample, measurement procedures, model specification and interpretation to these studies. These quality criteria, however, were used to determine whether studies would be included in the review or not. They were not used to assess the weight that should be allocated to the findings in relation to each intervention. More recently, Hanover Research's review [77] examined both financial and workplace incentives connected to teacher recruitment and retention. However, the study provides little critical analysis nor consideration of the quality of each of the included evaluations. Gunther [78] examined non-financial factors influencing teacher recruitment and retention, including a range of research design and quality criteria used for rating of included studies. However, the study focused on personal, school, community and job characteristics or factors, rather than examining the effectiveness of policy interventions introduced to tackle the teacher shortage.

Where attempts have been made to consider the quality of the included studies the results have been disappointing. Laurence et al.'s [79] review of programmes aimed at attracting and retaining teachers in the US concluded that it was difficult to be certain about the evidence of effectiveness as many of the programmes tended to be small and piecemeal and hence difficult to replicate on a large-scale. Lonsdale and Ingvarson [80] reviewed recruitment strategies employed in Australia, the US, UK, Canada and New Zealand and cautioned that many of the strategies have not been formally evaluated, or where they have, evidence tended to be anecdotal and informal.

Our review exclusively includes studies that can contribute to answering causal questions on the impact of strategies or policy initiatives to improve the recruitment and retention of teachers.

3. Methods

This review summarises the evidence of initiatives in addressing teacher recruitment and retention. Knowing more about what 'works' and what does not will allow policymakers and schools to make informed and targeted decisions on strategies to use or avoid, to attract and retain teachers. Given the huge amount of research in this area, we have to be careful these decisions are supported by the best available evidence. It is therefore imperative that the trustworthiness of each research finding is evaluated. To do this we used a multi-factor method for judging the quality of evidence of each study included in the review.

The research questions are:

1. What are the most promising approaches in attracting teachers into the profession?
2. What are the most promising approaches in retaining teachers into the profession?
3. What are the 'best bets' for schools, regions, and policymakers to improve the recruitment and retention of school teachers?

In the context of this review, teachers refer to classroom teachers who deliver teaching in state-funded schools from early years to post-secondary education. We have intentionally not limited ourselves to teachers of any phase or subjects as attracting and retaining different types of teachers may require different strategies. In any case, the majority of research conducted in this area focused on teachers in state-funded schools, with a small number covering special education teachers. There was also a disproportionate number about shortage subjects, such as maths and science or languages. This is not surprising as recruiting and retaining shortage subject teachers has been a concern in most education systems across the world.

3.1. Search Strategy

To search for relevant studies, a list of search terms was developed as follows:

Teacher supply OR teacher demand OR teacher retention OR teacher shortage OR teacher recruitment AND initiative OR incentive OR policy/scheme AND experiment OR quasi-experiment OR randomised control* trial RCT OR regression discontinuity OR difference in difference OR time series OR longitudinal OR systematic review OR review OR meta-analys* AND impact OR evaluation OR effect.*

These were applied to 13 educational, psychological and sociological electronic databases including:

- Education Resources Information Clearinghouse
- JSTOR
- The Scholarly Journal Archive
- Social Sciences and Education Full Text
- Web of Science
- Sage
- Science Direct
- Proquest Dissertations and Theses
- British Education Index
- ERIC (Educational Resources Information Center)
- IBSS (International Bibliography of the Social Sciences)
- Ingenta Journals (full text of a large number of journals)
- EBSCOhost (which covers the following databases: PsychINFO, BEI, PsycARTICLES, etc, ProQuest, IBSS

- Plus Google and Google Scholar.

These searches were supplemented by studies known to us and snowballing of relevant studies cited in the retrieved studies and from prior reviews of the literature.

The search terms were tested on well-known sociological, educational and psychological databases to ensure that they picked up relevant pieces of literature, and pieces already known on this topic. Following this, a very general and inclusive statement of search terms was generated for each database. These were adjusted to suit the idiosyncrasies of each.

To determine the causal evidence of policies and initiatives on teacher recruitment and retention, we included only studies using experimental (e.g., randomised control studies) or quasi-experimental designs (e.g., regression discontinuity, matched comparison, difference-in-difference, longitudinal time-series analysis and instrumental variables) and large-scale longitudinal studies, or similar.

The scoping review and previous reviews of literature suggested that there were few robust experimental evaluations of policy initiatives or approaches for teacher recruitment and retention. The decision was therefore made to include any empirical studies with at least some type of comparative design, but would have low ratings for trustworthiness in terms of causal claims.

The search was limited to studies published or reported in the English language. We intentionally did not set any date limits, to keep the search open. To avoid publication bias, the search included any material published or unpublished that mentions both substantive and causal terms.

A total of 6731 potentially relevant records were identified from titles alone. An additional 347 were added from following studies in previous reviews, studies known to us from previous work and from references in identified studies. These included 58 research reports from ProQuest Premium which were specifically related to the effects of induction and mentoring on teacher retention. All were exported to EndNote (a reference manager) for screening.

3.2. Screening

In any review, a broad search of the databases will invariably pick up huge numbers of irrelevant materials. This is even more so in our case as we intentionally kept the search as broad and comprehensive as necessary to ensure that we did not miss potentially relevant materials. A large majority of records were not relevant but contained some of the keywords. To remove these, we eyeballed the entries looking at the title and abstracts and removed those that were clearly not relevant to the topic. We then screened for duplicates using the EndNote function. Some studies were presented in different forms, or for different audiences, e.g., as a working paper or a report as well as journal articles. These were treated as one study.

In the next stage of screening the full reports were skim-read by one researcher. Any studies thought not to meet the inclusion criteria were then reviewed by other members of the research team for consensus. Four members of the team independently reviewed 10 randomly selected reports to agree on their inclusion or exclusion. The full texts of the included studies were screened by applying pre-defined inclusion and exclusion criteria as presented below.

3.3. Inclusion Criteria

Inclusion and exclusion criteria were determined prior to completing the searches and were applied after the initial screenings. Studies were included if they were:

1. Empirical research
2. About activities aimed at attracting people into teaching or about retaining teachers in teaching
3. Specifically about recruitment and/or retention of classroom teachers
4. About incentives/initiatives/policies or schemes on teacher recruitment and retention
5. About mainstream teachers in state-funded/government schools
6. Studies that had measurable outcomes (either retention or recruitment)

3.4. Exclusion Criteria

Studies were excluded if they were:

1. Not relevant to the research questions
2. Not primary research
3. Not reported in English
4. Not a report of research
5. Descriptions of programmes or initiatives with no evaluation of strategies or approaches used in teacher recruitment and retention
6. Not about strategies or approaches to improve recruitment or retention of teachers (e.g., observational or correlational studies of factors influencing recruitment and retention)
7. Studies that had no clear evaluation of outcomes
8. Studies with non-tangible or measurable outcomes (e.g., surveys about teachers' attitude or beliefs or perceptions
9. Ethnographic studies, narrative case studies, opinion pieces
10. Outcome is not teacher recruitment or retention
11. Focus only on specific groups of teachers, e.g., special education teachers or ethnic minority teachers
12. Not relevant to the context of English speaking developed countries
13. Recruitment and retention of school leaders, teaching assistants or school administrators
14. Anecdotal accounts from schools about successful strategies
15. Surveys collecting ideas about the best way or most effective ways to attract and retain teachers

The PRISMA diagram (Figure 1) tracks the number of studies included and excluded at each stage of the review process.

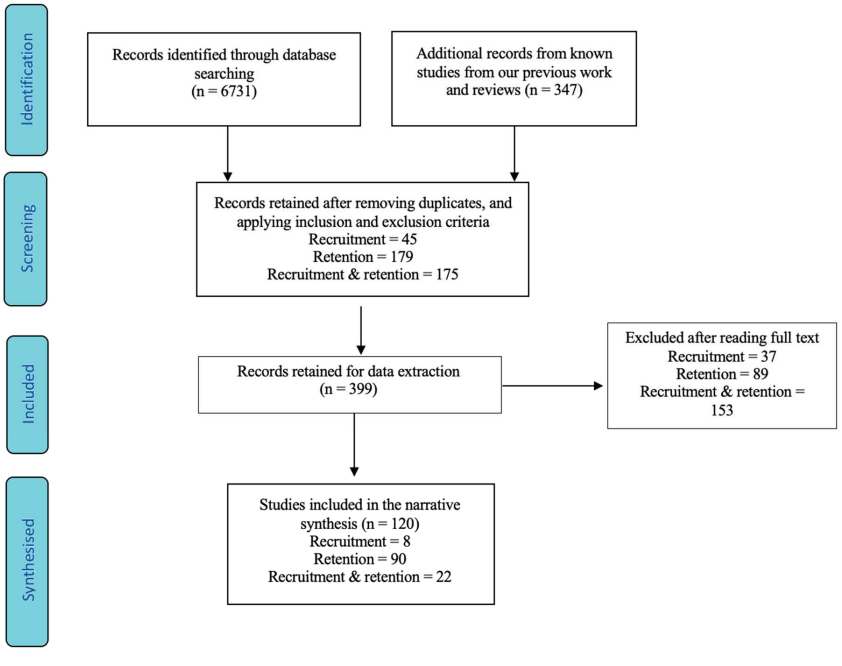


Figure 1. PRISMA Diagram of included/excluded studies

The majority of the studies that were excluded were because they were not relevant to the topic (i.e., not about teacher recruitment and retention), not about impact evaluation, recruitment and retention were not the outcomes, or were simply reports about best practices.

3.5. Data Extraction

The included studies were assembled and considered for synthesis. Key information from each of the included studies was extracted and summarised using a template (Appendix A). Such information included the research design, sample size, group allocation, outcome measures, missing data, methods of analysis and the results.

We then judged the research evidence and reliability of the finding for each piece of included study using the "Gorard Sieve" [81]. This uses five criteria to assess the trustworthiness of the research (see Appendix B): the design (whether it includes a fair counterfactual), scale of study (size of smallest cell, as small studies can be volatile and vulnerable to threats to validity), scale of missing data (missing data can bias the findings), quality of data obtained (data obtained from standardized test instruments or administrative data are considered more robust than data collected from participants' self-report or from instruments aligned with the intervention) and other threats to validity (e.g., conflicts of interest). All such factors are important [82] for ensuring that the evidence that informs policy and practice can be relied upon. Each study was then given a padlock rating ranging from 0 (no weight can be placed on the study) to 4🔒 (the most robust that could be expected in reality). This is an indication of how secure the findings are. We use the term "quality" to refer to the security of the findings and not necessarily the quality of the research. The ratings take no account of whether the intervention was deemed successful or not, or whether the report author claimed the intervention was effective. To ensure inter-rater reliability, four members of the team reviewed and rated a sample of five papers. Team members were in constant consultations with each other throughout the process to ensure consistency.

We did not compute the average effect sizes for each of the approaches as the study designs were so varied that averaging effect sizes across different studies which use different scales (e.g., odds ratios, hazard risks, mean effect sizes and r-coefficients) for measuring different aspects of similar intervention may not reflect the real impact of each individual type of programme. To illustrate, the studies in this review employ a range of methodologies (e.g., instrumental variables, regression discontinuity, time-series analysis, difference-in-difference and randomised control trials) to estimate the effects of a wide spectrum of measures aimed at improving teacher supply. One of these is financial incentives. This is an umbrella term which encompasses differential salary compensation, bonus incentive scheme, pension enhancement, scholarship and bursaries and tuition fee waivers. These strategies are not identical, or even similar in some cases, therefore averaging the effect sizes across the different strategies may not reflect the impact of each particular strategy. It is also the case in this review that there were often only one or two studies for each type of financial incentive that meet our causal criteria. It was therefore not possible or desirable to average the effect size for each type of programme [83].

The key matter is whether the effect is positive or not. The size of the effect can be misleading as studies with small samples, those without control groups, used non-randomised controls and those that are based on teachers'/pupils' self-reports of outcomes invariably show huge effect sizes [83]. To overcome these problems, we examined the substantive and methodological features of each study using the "sieve". Therefore, rather than reporting effect sizes, we present the direction of effect (positive, negative or no change) and the strength of the evidence (i.e., how secure is the finding).

3.6. Synthesising the Evidence

To facilitate the synthesis, we sorted the research reports by outcomes according to whether they were about recruitment, retention or both. A broad classification of incentives/initiatives was created. These include financial incentives (e.g., signing bonuses, wage uplifts, scholarships and loans), and other non-financial incentives (e.g., alternative routes into teaching, staff development, mentoring & induction and workload reduction) or a combination.

Approaches with the most highly rated studies showing positive effects are considered the most promising. Likewise, approaches rated highly (i.e., 2🔒 and above) showing negative or no effects are considered least promising given the existing evidence. All outcomes, whether positive or negative are considered. It is just as important to identify approaches that do not have evidence of effectiveness as it is to identify those that do work. It has to be made clear that approaches with no evidence of impact does not mean that they are not effective, but rather that the existing evidence is such that its effectiveness cannot be determined.

4. Results

The 120 included studies reported 157 individual outcomes relevant to recruitment and retention (Table 1). We discuss the approaches for improving recruitment and retention separately. However, a number of studies report on both recruitment and retention and these are included in both sections where appropriate. Studies receiving a 0🔒 or 1🔒 rating are not discussed in any further detail as their limited design or methodological quality means that they offer little in terms of indicating promising (or otherwise) approaches. Appendix C summarises the weaker studies (rated 0🔒 and 1🔒). These are mainly studies with very weak design. They either had very small samples, non-randomly allocated comparison groups, had no clear comparators, high attrition or based on models that made a number of unrealistic assumptions. All this makes it difficult to attribute the effect to the policy initiative or intervention. Therefore, including them in the discussion will add little to the overall finding.

Table 1. Number of studies with each security rating: all included studies.

Security Rating	Positive Outcome $n = 92$	Unclear/Mixed Outcome $n = 15$	Neutral or Negative Outcome $n = 50$
4🔒	-	-	-
3🔒	6	2	3
2🔒	43	6	17
1🔒	40	5	27
0🔒	3	2	3

4.1. Approaches to Attracting Teachers

Fourteen studies that focus on approaches relating to recruitment are rated 2🔒 and above (Table 2). All but two involve some kind of financial incentives (Table 3). This is perhaps because large-scale administrative/panel data relating to financial incentives are more readily available and accessible, and efficient to examine for researchers.

Table 2. Security ratings of studies on recruitment.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒		Rosen (2012)	
2🔒	Boyd et al. (2012) DeFeo, Hirschberg & Hill (2016) Dolan, Metcalfe & Navarro-Martinez (2012) Falch (2017) Fitzgerald (1986) Glazerman et al. (2013) Hough and Loeb (2013) Steele et al. (2010) Zarkin (1985)	Fulbeck & Richards (2015)	Bueno & Sass (2016) Gorard et al. (2020) Kraft et al. (2020)

Table 3. Security ratings of studies on financial incentives in teacher recruitment.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	1	-
2🔒	8 + 1	1	2

4.1.1. Financial Incentives

The strongest study (3🔒) shows mixed outcomes, but otherwise the results from the 2🔒 studies are predominantly positive (Table 3). This suggests that there is promising, but far from definitive evidence that financial incentives may be an effective strategy in attracting teachers into the profession and to specific regions, subjects or hard-to-staff schools.

The only study rated 3🔒 on recruitment [84] evaluated the impact of financial incentives on recruitment and retention of shortage subject teachers. The results were mixed. The study utilised an instrumental variables model using data from the School and Staffing Survey from 1999/2000 to 2007/2008 which contained data from 106,930 public school teachers in 6540 public school districts. This is perhaps the largest study of its kind and several models were employed within it. One compared teachers in districts that offered incentives with matched teachers in other districts. This does not overcome the problem that districts that did and did not offer such incentives may have other differences that could influence teacher recruitment and retention. There was no clear evidence that the use of incentives improved teacher recruitment or quality. Incentives were most attractive to those who were already interested in becoming teachers.

There were eight studies rated 2🔒 showing positive effects. These were not rated higher because of some limitations in the research design. These studies suggest that financial incentives, such as higher wages, stipends and bonuses can entice teachers to teach in challenging schools. DeFeo et al. [85] estimated that higher salaries are needed to attract more qualified teachers to teach in hard-to-staff schools. They analysed data from twelve Alaskan school communities in three districts to determine the minimum salary needed to attract highly qualified teachers in rural communities in Alaska, and how much more is needed to get teachers to teach in difficult-to-staff schools. Their analysis suggests that to compensate for factors that might make a community or school more or less attractive, salary differential would have to be between 0.85 and 2.01 with hard-to-staff schools having higher differentials. The differentials include costs of living among other working and living conditions that affect teachers staying or leaving communities. So, it might be the case that to attract maths and science graduates (who would command higher salaries elsewhere), the salary differential would have to be big enough to compensate for the difference they would otherwise get. It has to be mentioned that the amount of the bonus would have to be the salary differences on the teacher's actual salary and not the state salary schedule as some districts were already paying teachers more than was stipulated in the state salary schedule. Otherwise even with compensatory bonus, teachers' salaries could be the same or even below what they were already getting.

Fitzgerald [86] evaluated the impact of the High Priority Location Stipend Program in Miami, Florida. The stipends varied between $500 and $2000 annually depending on the position of the staff and the number of years they worked in the high priority areas. Over three years, vacancies in treatment schools fell substantially from the base year while teacher and vacancies in the control schools went up (effect size 1.3). This was a quasi-experimental study comparing 25 schools in the programme with 25 comparable control schools. High priority schools were those with a high proportion of students receiving free/reduced lunches. Control schools were similar in pupil and teacher characteristics but not designated as high priority schools. The treatment schools were selected for the stipend based on their high vacancies and lower retention of teachers suggesting that there were some inherent differences between them. Control schools may be more attractive to teachers by virtue of the fact that they had lower vacancies to start with. The design was unable to take account of changes in

circumstances within the schools (such as pupil intake), which could have affected teacher satisfaction and thus the retention rates.

Glazerman et al. [87] examined the impact of the Talent Transfer Initiative, which offered bonuses to the highest performing teachers for agreeing to move to and stay in low-performing schools. The incentive was $20,000 paid in instalments over a two-year period. Teachers who were already teaching in low-performing schools received a $10,000 retention stipend if they remained in the school over the two-year period. The participants included 85 teacher pairs matched on school characteristics and randomised to intervention or not, across 114 elementary and middle schools. Because the teacher pairs changed their personnel between randomisation and the start of the school year, the two groups were no longer equivalent at the beginning of the study. Of the vacancies assigned to the scheme, 88% were filled, compared to 44% the year before, and 71% in the comparison group.

Hough and Loeb [88] used a difference-in-difference approach, comparing the recruitment and retention of 1611 applicants in the San Francisco Unified School District. The district awards higher salaries/bonuses for teachers teaching shortage subjects, and in schools with a high proportion of poor and ethnic minority students. Teachers were also given a retention bonus if they stayed on after four years and more after eight years. The results showed an increase in the proportion of shortage subject teachers in hard-to-staff areas from 27% to 37%. There was also an increase in the proportion of new hires in the targeted group (those that received the incentives) from 49% to 54%.

Steele et al. [89] evaluated the Governor's Teaching Fellowship (GTF) scheme, involving a $20,000 incentive to attract and retain new teachers to low-performing schools for four years. Teachers had to repay $5000 for each year that they did not meet the commitment. An instrumental variable design was used, based on 718 GTF teachers, excluding those who could not be tracked, were missing data, or not enrolled at recognised institutions. GTF recipients were not randomly selected, and so may have had a predisposition to teach in low-performing schools. Twice as many teachers were enrolled during GTF as in the years before and after, and 28% more taught in low performing schools. It seemed that money was an attractor.

A UK study suggested (indirectly) that monetary incentives may be effective only in attracting those already intending to teach, not those who would not have considered teaching anyway [90]. This was an experiment with 1574 undergraduates (but data for 1496 was analysed) to test whether financial incentives would attract high ability students into teaching. Instead of asking students directly whether they would be motivated by financial incentives, which runs the risk of students giving answers which they think are desirable or acceptable, the authors presented participants with a hypothetical task for which they were rewarded for effort. In addition, they were offered an initial up-front payment or "endowment" conditional on their subject and predicted degree classification. This was to mimic the incentives offered for initial teacher training (ITT) bursaries. In England, the government offered differentiated bursaries for different degree subjects and degree class with high priority subjects attracting higher bursaries. Bursaries were found to be strongly and positively associated with intentions to become a teacher and to do initial teacher training, although the causality appears to be in the opposite direction. Those intending to be teachers were more likely to give greater importance to bursaries, instead of (or as well as) the other way around. The effect was stronger for women who were more likely to want to be primary school teachers than secondary. Those in the third year of study were also less likely to express intention to teach. This study was based on hypotheticals and on participants' expression of intention to become a teacher which weakens its validity.

Using a difference-in-difference approach, Falch [91] compared the recruitment rate of teachers within Norwegian public schools with variable wage premium using data from the 1990s when wages were centralized. Treatment schools had a certain level of teacher shortage and were thus eligible for wage premium. Of 79,135 teachers, 10,868 worked in one of the three counties with treatment schools, and 2034 worked at a treatment school. Because control schools did not have recruitment issues, comparisons were made with schools with persistent teacher shortages outside the three counties, which were not eligible for the wage premium. The results showed that the recruitment rate was

higher in treatment schools than non-treatment schools (effect size 0.13). A 10% increase in wage increased recruitment by about 30%. The wage premium appeared to be more effective in attracting young female teachers into teaching than older male teachers. Although a large study, this was a passive design.

In a longitudinal time-series analysis, Zarkin [92] developed an economic model to test how responsive the "reserve pool" of teachers is to the teacher salary at the time. The reserve pool of teachers in one year was estimated as the average proportion of certified teachers to the total certified over the 20-year period, multiplied by the total number meeting the minimum certification requirements in that year. They estimated that a 20% increase in wages could induce a 14% increase in the supply of secondary school teachers, and that secondary teachers were more responsive than primary teachers to an increase in salaries.

One 2🔒study showed mixed outcome—successful for some schools only. Fulbeck and Richards [93] explored the effects of ProComp, a performance-based financial incentive, on teacher mobility in Denver, CO, USA. Teachers were awarded an additional $24,000 if they taught in top performing schools, high growth schools or hard-to-staff schools. Seven such incentives were given to individual teachers for meeting student performance targets, and three were school-based incentives awarded to teachers who taught at hard-to-staff schools serving low-income population, high performing schools and schools that make the most progress in maths and reading. However, ProComp was eligible only to those who were members of teacher unions and who did not work in Charter schools. The sample included all public school teachers in Denver from 2006–2010 who were eligible for the incentive (regardless of whether they received it) and who made at least one voluntary move within the district (n = 989). Using conditional logit models, the authors predicted which school a teacher would transfer to given their individual characteristics, the characteristics of their current school, and the characteristics of the schools they could be transferring to. The results portrayed the incentive as successful in attracting teachers to high growth and high performing schools, but less successful in getting teachers into schools with a high proportion of low-income pupils or hard-to-staff schools. Financial incentives also did not encourage teachers to move out of the area they were currently in.

Another 2🔒 studies found no impact of financial incentives on teacher recruitment. Bueno and Sass [94] assessed the impact of the Georgia's bonus system (a monetary compensation) on the recruitment and retention of maths and science teachers. The bonus system increased the pay of new maths and science teachers to make it equal to that of a teacher with six years of experience. A difference-in-difference model was used to estimate the impact of the differential pay programme on the likelihood of becoming a teacher by comparing the difference between graduates with majors in maths and science and other education majors in the change before and after the programme period. They found that differential pay did not increase the number of maths or science teachers nor did it encourage people to switch to maths or science.

Gorard et al. [95] compared three groups of 4469 UK undergraduates, classified as never considered teaching, considered teaching but rejected it, and intending to teach. Before being asked about teaching, students were asked about what they were looking for in a career. The never considered teaching group was clearly the most different, and already on a trajectory to a "vocational" outcome like dentistry, medicine, architecture, engineering and so on. Once background factors, especially prior qualifications, had been accounted for, there was no difference between those intending to be teachers and the rest in terms of the extent to which prospective pay was a factor in their decision.

4.1.2. Alternative Routes into Teaching

Only one study that looked at the impact of an alternative teacher preparation programme for teachers (Table 4) was rated 2🔒, and so is discussed here.

Table 4. Number of studies with each security rating: Alternative routes and teacher supply.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	-	-
2🔒	1	-	-

Boyd et al. [96] compared the qualifications and retention of the Maths Immersion Program (MIP), teachers to New York City mathematics teachers who began their careers through other pathways. The study found the Maths Immersion programme was successful in attracting highly qualified teachers to teach in some of the most challenging schools. The number of such teachers increased from 2003 to 2008 at a faster rate than those who were prepared through the traditional college (CR), Teaching Fellowships (TF) and Teach for America (TFA) routes. They also had better academic qualifications than traditionally prepared peers, but weaker qualifications than TFA teachers.

4.1.3. Teacher Accountability

There is only one study that examined teacher accountability on the supply and quality of teachers (Table 5). High stakes teacher evaluation reforms were introduced across the different states in the US at different times. Kraft et al. [97] took advantage of this differential timing to estimate the impact of teacher accountability reforms on the supply and quality of new teachers using a combination of panel datasets from 2002 to 2016 in a difference-in difference approach comparing teacher supply (the number of licenses granted) and teacher quality (measured using the Barron's ranking of the teachers' training college) across different states. They compared the outcomes seven or more years prior (pre-reform) to a reform and three or more years after a reform (post-reform). High-stakes evaluation reforms reduced the number of licenses granted in a state by 2.69 per 10,000 18-to-65-year-olds. The reforms also made it difficult for hard-to-staff schools to fill vacant positions. On the other hand, teacher evaluation reform did raise the quality of teachers, increasing the likelihood of a teacher graduating from a higher ranking college by 8.1 percentage points.

Table 5. Number of studies with each security rating: Teacher accountability and teacher supply.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	-	-
2🔒		-	1

In summary, the most promising approach appears to be financial incentives. While substantial increases in salary may be linked to better recruitment in general, and perhaps in hard-to-staff areas and schools as well, studies that take the background of teachers or potential teachers into account suggest that salaries are not as important [95]. There is also the suggestion that monetary inducements, like bonuses and bursaries, may attract teachers in high demand subjects, but the evidence indicates that such incentives disproportionately attract those already interested in teaching, and are more successful in getting trainees into desirable schools rather than hard-to-staff ones.

The evidence for the different routes into teaching is not strong as there is only one medium-quality piece on this. There are no robust studies at all on most of the other approaches.

4.2. Approaches to Retaining Teachers

The evidence on teacher retention is more mixed than for recruitment (Table 6). There are no 4🔒 studies and the eight studies with a 3🔒 rating, all had unclear, neutral or negative outcomes. The majority of studies in this section either focus on financial incentive interventions or those which

provide professional development and/or mentoring. Several of those relating to financial incentives have already been described above under recruitment, and so are referred to only briefly below.

Table 6. Security rating of studies on retention.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒		Rosen (2012) Shifrer et al. (2017) Springer et al. (2016)	Clotfelter et al. (2007, 2008) Fryer (2013) Glazerman et al. (2010) Helms-Lorenz et al. (2016) Steele et al. (2010)
2🔒	Allen & Sims (2017) Bueno & Sass (2016) Cohen (2005) De Angelis et al. (2013) De Jong & Campoli (2018) Falch (2011) Feng & Sass (2015, 2018) Fitzgerald (1986) Glazerman & Seifullah (2012) Glazerman et al. (2013) Ingersoll & Smith (2004) Koedel & Xiang (2017) Latham & Vogt (2007) Murnane & Olsen (1990) Papay et al. (2012) Ronfeldt & McQueen (2017) Speidel (2005) Springer & Taylor (2016) Springer et al. (2010)	Booker & Glazerman (2009) Choi (2015) Fuchsman et al. (2020) Fulbeck (2011) Fulbeck (2014) Shirrell (2014) Silva et al. (2014/2015) Weisbender (1989)	Anders et al. (2019) Boyd et al. (2012) Dee & Wyckoff (2015) Hendricks (2014) Hough & Loeb (2013) Jones 2013

4.2.1. Financial Incentives

The evidence on financial incentives for retention is less clear than for recruitment. All of the stronger studies (3🔒) do not suggest clear benefits (Table 7).

Table 7. Number of studies with security rating: Financial incentives and retention.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	3	3
2🔒	9	4	4

The large study by Rosen [84], discussed more fully in the recruitment section above, for example, found no clear evidence that districts offering incentives had higher teacher retention, at least after the first year.

Shifrer, Turley and Heard [98], another 3🔓, looked at whether actual receipt and the amount of performance pay award in an urban school district as opposed to eligibility made a difference to teachers' decision to leave or stay. Using the difference between a large and a small award as the cut-off threshold, they conducted a regression discontinuity analysis using census data for 12,000 teachers although they focused only on 3363 teachers. Teachers in the top quartile of value-added scores were rewarded with a large award and teachers with a value-added score in the second quartile a small award. Their analysis showed that likelihood of retention was slightly higher for teachers who received a small award rather than no award. However, this study found that teachers who received a large award were less likely than teachers who received a small award to be retained in the district. Perhaps teachers in receipt of a large award are high performing teachers who can easily find better paid jobs elsewhere.

Springer, Swain and Rodriquez [99] evaluated the US$5000 retention bonus program for effective teachers in Tennessee's Priority Schools (high poverty, high minority schools). The study showed that the bonus incentive increased the retention of teachers in tested subjects and grades, but not the retention of Level 5 (Diploma in Education and Training) teachers. This was a quasi-experimental study using a regression discontinuity design exploiting the sharp cut-off in a teacher's overall evaluation rating that determines eligibility for the retention bonus in participating schools. Nationally representative administrative data supplemented by county-level economic data and data from the TVAAS and Tennessee's online teacher evaluation platform, CODE (contains value-added estimates for teachers) were utilised for the analysis. The sample included all teachers working in Priority Schools in Tennessee during the 2012–2013 school year.

Three moderately high evidence studies (3🔒) found no effect of financial incentives on teacher retention. Steele et al. [89] evaluated the Governor's Teaching Fellowship (GTF) scheme, involving a $20,000 incentive to attract new teachers to low-performing schools. Teachers had to repay $5000 for each of the first four year that they did not meet the commitment. There was no difference in retention rates (75% over four years) between recipient and non-recipients, despite the penalty clause.

Using a difference-in-difference-in-difference approach, Clotfelter et al. [100,101] compared hazard rates before and after the implementation of the bonus programme; eligible and ineligible teachers in the same schools using a hybrid of a randomized experiment and a regression discontinuity design. This is the North Carolina bonus incentive aimed at retaining qualified teachers in targeted subjects in high poverty or academically challenging schools. Under this scheme teachers were eligible for an annual bonus of $1800 if they taught in an eligible school, and will continue to receive the bonus as long as they stayed in the same school and taught the same subjects. Overall, the results suggest that the bonus incentive did not reduce turnover rates. However, it is not clear whether this is because the $1800 bonus was not large enough or is it because there was a flaw in the design and implementation of the program as not all teachers who were eligible actually received the bonus. Survey responses from principals and teachers indicated that the $1800 bonus alone was not enough to retain teachers. They suggested that administrative support, improving school conditions and facilitating professional development might be better options. Comparison was made with teachers across eligible schools and those in schools that narrowly missed out based on the threshold eligibility. The results showed that teachers receiving a bonus were 15% less likely to leave at the end of the school year compared to other teachers in the same school. This increased to 17% after controlling for subject taught. A 10% increase reduces the probability of teachers leaving by 1–4% points. However, this reflects a pattern already in place even before the programme was introduced. Including the school fixed effects in the regression the effect was negative.

In a cluster randomised control trial, Fryer [102] examined a school-wide performance bonus scheme that provided performance bonuses to school staff based on their schools' progress report. Using both an intention-to-treat (ITT) and treatment-of-the-treated analyses, the results showed that the teacher performance bonus program had no effect on teacher retention. Some reasons suggested for the nil effect included incentives not being large enough, incentive scheme was too complex and group-based incentives may not be effective. Participating schools were given a lump sum incentive on $3000 per full union teacher. Schools could decide to award a subset of teachers with the highest value-added or divide among teachers by lottery. The majority of schools opted for group incentives. Data on students and teachers from 396 high-need public elementary, middle, and high schools from 2007–2008 through 2009–2010 were analysed. Schools were selected based on some criteria, e.g., level of poverty. Of these schools, 233 were randomly assigned to the intervention group and 163 to the comparison group. Retention outcomes in schools that were offered participation in the program—even if they ultimately declined to participate—were compared with the outcomes in schools that were not offered the opportunity to participate.

Nine studies rated 2🔒 reported positive outcomes of financial incentives on teacher retention, but the effects were either short-lived or involved some kind of a tie-in. Bueno and Sass [95] found

that salary compensation only had a short-term effect on the retention of teachers. They compared teachers who were eligible with those who were not. The attrition rate for bonus recipients was lower than non-recipients, but only in the first five years when they were receiving the bonus. Working and living conditions, lack of community engagements were reported to be important factors in teachers' decision to stay or leave.

Falch [103] used a natural experiment taking advantage of changes in the wage system in Norway over a nine-year period (1993/94–2002/03) to look at whether giving teachers a higher salary would make them more likely to stay in teaching. Over that time, teachers in schools with high teacher vacancies were eligible to receive a wage premium of between 7.5% and 12%. In total, there were 161 treatment schools. Of these 104 received wage premium for less than four years. The difference-in-difference analysis comparing the turnover rates before and after wage premium was introduced showed that wage premium reduced the probability of quitting a school by 4.8 percentage points. Taking into account school district characteristics, the effect of the wage premium increased to 5.8 percentage points. The wage effect was larger for males and for the married teachers than for females and unmarried. Teachers' age and whether they have children or not also affect the size of the effect. There was no impact on the retention of younger teachers, and female teachers were less responsive to wage increases than older and male teachers.

Another incentive scheme is the Florida Critical Shortage programme is a state-wide programme to increase the supply of teachers in shortage subjects. Feng and Sass [104,105] evaluated the effects of two components of the scheme. The loan forgiveness component of the programme was targeted at beginning teachers teaching in shortage areas where teachers qualified in that subject were given up to $10,000 to pay off their student loan if they continued to teach in shortage subjects for at least 90 days. The other component was a one-off retention bonus for teachers teaching certain subjects and grade levels. Retention bonus was capped at $1200 per teacher. To be eligible teachers would have taught in a shortage area, agreed to continue teaching that subject the following year and have had a favourable performance appraisal. Using difference-in-difference and instrumental variable approaches, the authors compared the probability of attrition and recruitment of eligible and non-eligible teachers for each shortage subject. The effect of loan forgiveness was estimated by comparing changes in retention of eligible teachers when a subject was designated as a shortage subject with those of non-eligible teachers over time. The results showed that loan forgiveness reduced the probability of overall attrition by 12% (10% for maths and 9% for science teachers). The effect disappeared when the funding was reduced. The one-off retention bonus resulted in a reduction of likelihood of shortage subject teachers leaving Florida by 25%., but no effect on retention in the school they were currently teaching.

Fitzgerald [86] also found that the impact of financial incentives is short-lived. The effect of the High Priority Location Stipend Program was observed only in the first year after implementation. No differences were found in the following years. Staff who left indicated that while they were appreciative of the incentives, they did not think the stipend was high enough. Their main concerns were the working conditions, discipline in school, management support and admin/teacher relations. Control teachers also indicated that they would be happy to work in the high priority areas if student discipline, working conditions and admin/teacher relations were improved.

Pension enhancements have also been used in states in the US to encourage teachers to stay until their retirement. Koedel and Xiang [106] examined one such scheme in St Louis, Mississippi using the six-year administrative panel data covering the school years 1994–1995 through 1999–2000. They compared the likelihood of eligible teachers (i.e., those retiring after June 1999) with those not eligible using a difference-in-difference analysis. This resulted in a 60% increase in pension wealth for the eligible teachers, and had the effect of delaying the retirement only of teachers who were a year close to retirement, but not for other groups.

Glazerman et al.'s [87] analysis of the Talent Transfer Initiative, described in the section on recruitment, was also linked to increased retention. But this is because a condition of the incentive

is that teachers agree to move or stay in the low-performing school Retention in the first year was 93% (70% in the comparator group), but dropped to 60% after two years (compared to 51% in the comparator group). This suggests that the effect of the incentive was not long lasting and loses its effect once the payment stopped.

Murnane and Olsen [107] examined the career histories of 13,890 North Carolina teachers to estimate the impact of salaries on teachers' longevity in teaching. Regression models were developed using a number of key explanatory variables to predict the length of stay in teaching. Results of the analysis indicated that a $1000 increase in each step of the salary scale (measured in 1987 US Dollars) is associated with an increase in median duration of two to three years for a teacher starting their career in 1970. The findings suggest that a uniform salary scale may not work in retaining teachers in fields such as chemistry and physics that are in demand in business and industry. This echoes the findings of other studies which found that for financial compensation to be effective it has to be large enough to cover the differential salary that teachers would get if they had not gone into teaching.

Springer et al. [108] evaluated the District Awards for Teaching Excellence (D.A.T.E), a state-funded incentive pay award in Texas. All districts in the state were eligible to receive grants, but participation was voluntary. The average award for teachers ranged from $1361 in districts with district-wide plans to £3344 in districts with select school plans. The study showed that the likelihood of leaving increased for teachers who did not receive the award, but the probability of leaving fell sharply for those who did receive the award. The size of the award also matters. In districts with relatively small maximum awards, turnover increased, but turnover fell as the awards increased until it exceeds $6000 beyond which it makes no difference. However, not all districts and not all schools were eligible for participation, and it is not clear what the eligibility criteria were. Districts and schools that received the award may be systematically different to those not eligible for the award. The factors that exclude them for eligibility may be relevant to teacher turnover.

In a pilot study, Springer and Taylor [109] found mixed effects on a pay-for-performance program (Governor's Educator Excellence Grants/GEEG) in Texas. The Texas GEEG programme, was a three-year programme involving 100 schools (analysis performed with 94 schools) identified as the highest-poverty, high-performing schools in the state. Schools were awarded non-competitive grants ranging from $60,000 to $220,000 each year for three years. The individual award for each full-time teacher was between $3000 and $10,000. Using a combination of data from different sources the author analysed teacher turnover for six academic years. The results showed that turnover was higher among beginning teachers in schools with only individual incentives than in schools with only schoolwide incentives, but only in the first year. No differences were detected in subsequent years. The opposite was true for experienced teachers where turnover was lower in schools with school-level incentives than in schools with a combination of individual and school level incentives in the first year, but the pattern was reversed in the second year. No differences were detected between school and individual level incentives in the third year, suggesting the short-term effects of such incentive award.

A further four 2🔒 studies showed unclear or mixed outcomes. Booker and Glazerman [110] evaluated the Missouri Career Ladder (CL) Program to test the effect of pay increases on teachers at different stages of their career. Based on their performance-level eligible teachers received supplementary pay for taking on certain responsibilities or professional development outside their contracted hours. Teachers were observed and evaluated as they moved up the career ladder in three stages. The amount of bonus was also related to the length of teaching experience. For each stage teachers received more supplementary pay up to £1500 for Stage 1, £3000 for Stage 2 and £5000 for Stage 3. The authors compared the retention rates of teachers in districts offering the Career Ladder incentive with similar teachers in non-Career Ladder districts. There was no difference in retention rates between CL and non-CL districts after controlling for observable differences such as wealth, size and population density in regression models. Using instrumental variables controlling for district selection into CL participation, teachers in CL districts were less likely to move to a different district. The model predicted that after 10 years teachers in CL districts were less likely to move compared

to similar teachers in non-CL districts (81% remain vs 77%). The oldest teachers (after 11 years and receiving the biggest bonuses) were half as likely to move compared to their non-CL peers. It was more effective in retaining younger teachers in the profession but not necessarily in the district. The authors estimated that incentive payments need to exceed 25% of teacher salary to neutralise the effects of turnover in hard-to-staff urban schools. One complication is that this programme also had an element of enhancing teacher autonomy. Therefore, it is not clear how much of the effect was due to the incentive and how much was the result of teachers' enhanced autonomy.

In another study Fulbeck [111] used interrupted time-series and difference-in-difference regression models to analyse the impact of Denver's Professional Compensation for Teachers Program (ProComp), a teacher incentive programme that awards salary increases and/or annual incentives to teachers who meet a range of requirements, such as having advanced qualifications, complete professional development, teach in a hard-to-staff school or shortage subject and work at a high-achieving school. ProComp was championed by Barack Obama as a model for teacher compensation reform. The ProComp hard-to-serve incentive initiative (HTS) is one of 10 financial incentives aimed at retaining teachers in schools with a high proportion of poor students. The number of teachers under the scheme was between 3900 and 4200 each year. Panel data, teacher interview data, and data on school characteristics were taken from Denver Public School and ProComp school-level information. The study compared the retention rates of teachers before and after ProComp. It reported that participation in ProComp increased retention rates by 2.1 percentage points. Regression analysis showed that ProComp accounted for 2.5% of the variation in changes in retention rates. ProComp is reportedly more effective in challenging schools at or above average participation (ES = 0.30), but less meaningful for non HTS schools (ES = 0.05). The findings, however, are really difficult to interpret as the graphs seem to contradict the findings reported. Also the incentive came in at the time of the economic recession, which may have affected individual's propensity to move.

Using multinomial hierarchical regression modelling of data taken over a year, Fulbeck [112] estimated the risk of teachers moving within district and moving out of the district by comparing the hazard rates of teachers who received ProComp with those who did not, and also between teachers who taught in high poverty schools with those who did not. The results of the analysis showed that receipt of ProComp reduced the odds of teachers leaving the district, but not out of schools within the district. This relates only to those who volunteered to participate in ProComp and received the $5000+ incentive. There was no effect on those who volunteered but did not receive the incentive. These are likely to be teachers who did not meet the eligibility criteria in terms of performance and knowledge/skills. However, ProComp was not effective in high poverty schools. In other words, ProComp did not compensate for poor working conditions, school leadership and climate.

Choi [113] reported positive effects of the Quality Compensation program (Q Comp) on teacher retention but only in schools that have implemented the scheme for five years—6.3 percentage points higher compared to schools with less than five years of implementation. There was no benefit for charter schools (retention rates 10.5 percentage points lower than other schools). Q Comp is an alternative teacher compensation program (ACPs) under which teachers' pay was based on their performance, measured in terms of student achievement, leadership, professional knowledge and skills, and instructional behaviour. The study used a difference-in-difference-in-difference approach analysing data for 12,708 teachers and 1734 schools over 8 years. Teacher retention was calculated by comparing the list of teachers in two subsequent years.

Four other 2🔒 indicated that financial incentives did not improve retention of teachers. A study in England looked at whether pay reforms in England where schools are given the freedom to set pay based on performance rather than seniority have impacted on teacher retention. Anders et al. [43] compared three groups of schools—the positive adopters where pay progression on average was faster than pre-reform seniority-based salary schedule; negative adopters where pay progression was slower than expected under pre-reform; and mean-zero adopters where pay progression was as expected under pre-reform pay schedule based on seniority. Using a difference-in-difference framework the

authors estimated the effect of pay reforms on teacher retention, using adopters as treatment groups. The effect of the reform increased teachers' pay at positive adopter schools by 4% while pay of teachers in negative adopter schools fell by 3%. However, there were no effects on retention.

Dee and Wyckoff [114] reported a performance incentive programme (IMPACT) aimed at retaining effective teachers in the District of Columbia. IMPACT had been successful in removing low performing teachers and retaining high-performing teachers. Teachers were evaluated on a multifaceted measure of teacher performance. Based on these evaluations low-performing teachers may be dismissed and high performing teachers receive large financial incentives. The financial incentives included one-time bonuses of up to $25,000 and permanent increases to base pay of up to $27,000 per year. Employing a regression discontinuity design, they compared the retention and performance outcomes of 4000 low-performing teachers whose ratings placed them near the threshold at risk of strong dismissal threat. The study also compared outcomes among 2000 teachers who had IMPACT scores just above and just below the threshold between Effective and Highly Effective. The high stakes incentive programme was successful in removing teachers at the threshold of being labelled minimally effective, but did not improve the retention of high-performing teachers.

Hendricks [115] compared the attrition of teachers in districts which award teachers via pay for year of experiences, with districts that do not. The study found no relationship between teacher pay and turnover. Districts differ in terms of labour and market outcomes so those districts that award pay increases by years of experience may already be experiencing high attrition of more senior teachers.

Hough and Loeb [88], described under recruitment, found no difference in the retention rates of targeted and non-targeted teachers for higher salaries/bonuses. Over 90% of teachers stayed on in the district and over 85% stayed in their school, in both groups. The comparison is made difficult because of the economic downturn in 2008 when unemployment was high.

4.2.2. Teacher Development and Support

Previous studies have suggested that teacher development, which includes mentoring for inexperienced teachers and induction for early career teachers can help support and retain teachers in the profession. Our review found mixed results with the strongest studies showing no obvious benefit of teacher induction, while the weaker studies are largely positive about mentoring and induction (Table 8).

Table 8. Number of studies with security rating: Teacher support and retention.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	-	2
2🔒	10	1	-

Although 10 studies of moderate quality 2🔒 suggested positive effects, the two strongest studies rated 3🔒, using randomized control designs, showed that mentoring and induction did not make a difference to teacher retention.

For example, Glazerman et al.'s [56] evaluation of a comprehensive teacher induction programme in Princeton, New Jersey (US) found no impact on retention of teachers who received either one or two years of comprehensive induction within school, district or teaching profession over the first four years of the teachers' careers. This was one of the strongest studies using a randomised control design involving 1009 teachers in 418 schools. The mentoring programme consists of a year-long curriculum for beginning teachers that focuses on effective teaching. Mentees also had the opportunity to observe experienced teachers. In the second year, monthly Teaching and Learning Communities were held where mentors and mentees met for peer support and to discuss aspects of classroom instruction. In the second year, beginning teachers also received between 35 and 42 h of professional development.

Another randomised control evaluation of an induction programme for beginning teachers in the Netherlands also showed no clear effect on teacher retention [116]. It involved 71 schools with 338 beginning secondary education teachers who were randomly allocated to receive the induction arrangements or a business-as-usual control group. Because schools routinely provide beginning teachers extra support, control teachers also received some induction albeit only for a maximum of one year. Experimental teachers, on the other hand, followed the programme for three years under controlled conditions arranged by the schools, which included workload reduction and professional development. Both groups were similar in background characteristics. The results showed that three years later, 14% of the control group and 12% of the experimental group had left. (ES = +0.076). Importantly, the study found that it was the lack of certification and the low teaching skills that most explained teachers leaving the profession.

The 2🔒 studies mostly reported positive outcomes. Allen and Sims [57] evaluated STEM Learning Network professional development courses intended to improve teachers' subject, pedagogical and career knowledge, confidence and motivation. They used retention data of teachers from England's Department for Education (DfE) School Workforce Census. This was matched with the National STEM Learning Network to identify teachers who participated in the CPD courses. The authors used propensity score matching, matching participants with non-participants by known characteristics. To control for unobserved differences, comparisons were made between those who participated in 2010 with those who participated later. The authors argued that these individuals were therefore more likely to be similar in terms of motivation and career plans. Further analyses were also made comparing science departments in schools before and after the treatment. The study suggests that taking part in National STEM Learning Network professional development is associated with an increase in retention in the profession as a whole. The odds that a participant stays in the profession one year after completing these courses was around 160% higher than for similar non-participants, and the positive association is sustained two years later for recently qualified teachers. Using the more rigorous double-difference and triple-difference models that takes into account factors that are not included in the demographic and background measures, the positive association is maintained. However, there is no evidence that completing CPD courses improves retention within the schools that teachers were working in at the time of participation.

Cohen [117] used administrative data for 51,811 US public school beginning teachers comparing whether they had received a formal induction programme or not, and their perceptions of workload and classroom support. They correlated these variables with whether teachers stayed on the following year. Analysis on teacher induction was based on 3172 new public school teachers. This indicated that teachers who left reported less mentoring than stayers (effect size 0.12) and less supportive communication (effect size −0.04) and less common planning (effect size 0.11). Higher workload reduction levels did not relate to turnover.

De Angelis, Wall and Che [118] found that having more comprehensive mentoring and induction support significantly decreased the odds of new teachers changing districts and leaving the profession after one year. Quality of teacher support was based on teachers' self-report of their perceptions. It is therefore possible that teachers who were more likely to leave or had no intention to stay in teaching were more likely to report less favourable perceptions of programme quality.

De Jong and Campoli [119] analysed the observational data from the 2007–2008 Schools and Staffing Survey (SASS) to see if the use of curricular coaches is associated with teacher retention. Curricular coaching provides new teachers with the techniques to incorporate evidence-based instructional methods in their local context. Using multinomial logistic regression analysis, they compared the likelihood of teachers leaving profession, staying or moving school of those who had a curricular coach and those who did not. They found that early career teachers in a school with a curricular coach was less likely to leave the profession (relative risk ratio = −0.52). The effect was stronger for first year teachers, but much less so for second and third year teachers. However, having a curricular coach did

not have an influence on early career teachers' decision to move school. It is possible that this was the period of economic recession when there is less incentive to change profession.

Glazerman and Seifullah [23] evaluated the Chicago Teacher Advancement Program (TAP), a teacher development and compensation programme. The implementation of the programme was staggered across all schools with schools randomly assigned to implement sooner or later, creating comparison group for analysis. Teacher retention was measured by comparing the retention of a matched sample of over 2600 teachers in Chicago TAP and conventional public schools. In this programme teachers and mentors met weekly in their "cluster groups". Teachers were also given performance incentives and had the opportunity to assume leadership roles. The results showed positive effects on school retention only for the first cohort but the effect was not consistent across cohorts. More teachers from the first cohort returned to their same school three years later compared to teachers in non-TAP schools, an impact of nearly 12 percentage points. In other words, teachers in Chicago TAP schools were about 20% more likely than teachers in comparison schools to be in those same schools three years later. For teachers in schools that started the Chicago TAP in later years, the impact was not obvious. There was some evidence of impacts on retention for subgroups of teachers, such as those with less experience, but there was no consistent pattern.

A correlational study using a nationally representative sample showed a positive correlation between participation in induction/mentoring programmes and the likelihood of teachers leaving or moving school. However, it is not just having mentors, but having same-subject mentors that mattered [120]. Having mentors from different subject areas had no influence on beginning teachers' decision to leave. The study analysed data from the School Staffing Survey (SASS) and the Teacher Follow-up Survey (TFS) which included a sample of 3235 beginning teachers in their first year of teaching. The survey asked teachers about their participation in any form of induction programme including mentoring, CPD, collaboration with other teachers and support. The multiple kinds of support included in these induction programmes meant that it was not possible to isolate which of these were most effective. Although the authors controlled for school and teacher effects, they were unable to control for unobserved differences between teachers and schools. Because those who received mentoring and those who did not were not randomly allocated, there may be inherent differences between these two groups. It could be that schools or districts that offer mentoring support are generally more supportive of their teachers, or have better working environment.

Latham and Vogt [121] compared the retention propensity of 506 elementary education graduates in Illinois who had opted to undertake teacher preparation in a professional development school (PDS) with another group of 559 traditionally prepared graduates matched on demographic characteristics. The authors claimed that those trained in PDSs (defined as having elements of field placement, onsite coursework and professional development) were more likely to stay in teaching for longer (about 0.25 of SD more than those who did not). It is important to note that the PDS group were self-selected and hence were likely to be different to those that were in the non-PDS group.

Papay et al. [122] found that graduates of the Boston Teacher Residency Programme were less likely to leave teaching in the first year (12%) than other new Boston public school teachers (27%). By the fifth year, retention rates among BTR teachers were still higher than other public school teachers in Boston (49% vs 25%). However, it has to be mentioned that BTR teachers were committed to teach in Boston for three years after their residency year or pay a penalty equivalent to the programme tuition fees of up to $10,000. They were more likely to stay until their fifth year, and did not leave suddenly after their third year when their commitment had been fulfilled.

Ronfeldt and McQueen [123] drew on the SASS, TFS and BTLS data to investigate whether different kinds of induction supports predict teacher turnover among first-year teachers. To mitigate against unobserved factors, the authors used propensity score matching of demographic characteristics to link 1600 teachers receiving extensive induction (i.e., 4 to 6 induction supports) with 1130 teachers not receiving extensive induction (i.e., 0 to 3 types of support). Unlike previous studies that focused on only one cohort, this study looked at three recent cohorts of teachers. In total, there were 13,000

across the three waves, but only 2340 were first year teachers that could be linked to both teacher and school characteristics. The authors correlated the level of induction support with teacher outcomes (leaving school and leaving profession). Multilevel regression analyses showed a negative correlation between the number of combined induction supports and teachers' likelihood of leaving school or teaching in their second year and across five years. Receiving extensive induction supports reduced migration by 5% compared with not receiving extensive induction supports. Of all the induction supports, supportive communication with school leadership had the biggest impact, reducing the odds by 55% to 67%. Every additional induction support was associated with an average decrease in the odds of leaving teaching by between 18% and 22%. One major limitation of this study is that the measure of induction is based on teacher self-report and this is prone to reporting biases.

Speidel [124] evaluated a teacher development programme, known as the Skills, Tips, and Routines for Teacher Success (STARTS), in the Volusia County Schools (Florida) designed for teachers of students with special needs. The study utilized data on the employment histories of 771 new special needs teachers for school years 1998/99 to 2003/2004. The findings suggest that the programme makes a positive difference in the retention rate of teachers who took part in STARTS. However, there were no controls for differences between the two groups of teachers. There were other variables that might have been in play with respect to teacher retention that were not accounted for.

Further, one 2🔒 showed mixed outcome. Weisbender [125] evaluated the California Mentor Teacher Program which was developed to retain experienced teachers and to assist new teachers in the transition into teaching. Under this scheme, highly talented classroom teachers (mentors) were given the incentive to continue teaching and to use their instructional expertise to mentor their peers and new teachers (mentees). The study included 336 mentors and 638 of their mentees in 240 schools and 46 retirees in the Priority Staffing Program serving 46 schools. Personnel records and questionnaires over a 5-year period were collected to assess the length of time each cohort stayed in the district. Comparisons were made between mentors and a matched group of non-mentors. Results varied from cohort to cohort. There was no effect on retention for the first cohort, with non-mentees being more likely to stay within the school district compared to mentees. With the subsequent cohorts, mentees were more likely to stay compared to non-mentees. On the other hand, mentors were also more likely to leave over the 5-year period than non-mentors. Although comparison mentors were matched, the selection of highly effective teachers suggest that the two groups may not be equal. As Shifrer et al. [98] noted, it may be the case the high performing teachers can find jobs more easily and are therefore more mobile.

4.2.3. Alternative Routes to Teaching

There is no clear evidence that offering alternative routes into teaching is beneficial in retaining teachers. Two studies rated 2🔒 that examined alternative routes into teaching showed no clear advantages of any alternative pathways in retaining teachers (Table 9).

Table 9. Number of studies with security rating: Alternative routes and retention.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	-	-
2🔒	-	1	1

Silva et al. [126] evaluated a teacher residency programme (known as the Teaching Quality Partnership Grants Program), which works in partnership with local school districts and universities where prospective teachers complete a coursework with supervised fieldwork experience teaching in a school for at least a year. The data shows that there is no difference in the retention rates of TRP and non-TRP teachers within district (89% and 87% respectively) and within schools (77% for TRP and 79%

for non-TRP). Teachers who moved schools were more likely to move to higher performing schools with a smaller proportion of ethnic minority children.

The second study by Boyd et al. [96] compared the Maths Immersion Programme with traditional certification and Teach for America (TFA). Compared to their traditionally prepared peers, immersion teachers were more likely to leave teaching in NYC (ES = −0.14) although less so than TFA teachers (ES = −0.3). They were also more likely than traditionally prepared teachers to transfer or leave their school (ES = −0.2). TFA teachers were more likely to leave teaching after four years but less likely to leave their schools. This is a large study using administrative data.

4.2.4. Teacher Accountability

One commonly cited reasons for teachers leaving the profession is high stakes tests and accountability pressures [127–129]. However, our review found that removing or reducing teacher accountability does not seem to have a clear benefit on retention, although the evidence base is particularly weak here (Table 10).

Table 10. Number of studies with security rating: Accountability and retention.

Security Rating	Positive Outcome	Unclear/Mixed Outcome	Neutral or Negative Outcome
3🔒	-	-	-
2🔒	-	2	1

High stakes tests which increase teacher accountability are a reported source of stress. Fuchsman, Sass and Zamarro [129] took advantage of a policy change in Georgia, US in 2011 when testing was removed for Grades one and two and from 2017 onwards when testing for science and social science were removed for Grades 6 and 7. The study compared the attrition rates of teachers in grades one to eight, before and after testing and with teachers in other grades where testing had not been removed using a difference-in-difference approach. The study found no impact on teachers' likelihood of leaving teaching, changing schools within a district, or moving between districts. However, there is a reduction in the probability of teachers with 0–4 years of experience leaving the profession when testing requirements were relaxed from 14 to 13 percentage points for teachers in grades 1 and 2 and from 14 to 11 percentage points in grades 6 and 7. Although comparisons were made before and after testing, the comparisons were not between similar groups.

Shirrell [130] estimated the impact of accountability under the No Child Left Behind Act (NCLB) where schools were held accountable for the performance of minority ethnic subgroups only if the number of students in those subgroups exceeded a minimum subgroup size. Using the minimum subgroup size threshold of 40 for a regression discontinuity analysis, Shirrell compared schools on either side of the threshold before and after NCLB. A difference-in-difference analysis was also used to compare teachers of different ethnic sub-groups. The study found that Black teachers in schools that were held accountable for the performance of Black student subgroup were less likely to leave than Black teachers in schools not accountable for the Black subgroup's performance. There was no difference in attrition for the White subgroup. One reason suggested could be that Black teachers were more likely to be paired with minority ethnic pupils and it is possible that these teachers were motivated to stay on in the school seeing that the schools were taking action to address the achievement gap between Black and White students. Shirrell also surveyed student teachers before they began teaching and after. The results showed that challenging working conditions generally do not predict changes in student teachers' career plans, although poor working conditions in training schools are associated with decreases in the lengths of time they plan to teach during their careers. Overall, there was no evidence that working conditions and accountability had any effect on attrition of ethnic minority primary school teachers.

Jones [131] used an instrumental variable approach to estimate teacher turnover under performance pay incentives for maths and English teachers (an accountability system), employing nationally representative datasets. Teachers in performance pay districts earned a salary that was $2825 less than their counterparts in non-performance pay districts and the performance pay may be used to compensate for the difference. Data from Teacher Follow-up Survey showed that performance pay was not considered the most important reason for teachers' decision to leave. Since the performance pay incentives were rewarded at the school level, this finding may also suggest that other teachers were free-riding on the efforts of Math and English teachers. Because the sample consisted of only 64 teachers caution is urged in interpreting this result. Also, implementation of performance pay incentive vary between districts. For example, performance pay was more effective in reducing turnover when it was implemented on a school level than on an individual level, and male teachers also responded more positively than female teachers to performance pay. In summary, the evidence on retention is not clear.

5. Discussion

5.1. The Evidence on Recruitment and Retention to the Teaching Profession

In summary, financial incentives seem promising for attracting new teachers into teaching, and in increasing the number of teachers in challenging schools with a high proportion of poor or disadvantaged children. However, the effect is stronger for high performing schools and schools with lower proportions of disadvantaged children. The evidence suggests that for financial incentives to work, they have to be large enough to compensate for the challenges of working in less desirable schools and areas, or to compensate for the salary that teachers would receive if they had been in comparable profession. This is especially so for shortage subject teachers like maths and science where graduates from these subjects tend to command a higher salary in the labour market. In England, bursaries are offered to trainees in secondary shortage subjects. The lower proportion of bursary holders in state-funded schools compared to non-bursary holders, suggests that the bursaries are not attracting shortage subject teachers to state-funded schools. This may be because the bursaries are not large enough. It is also possible that individuals who were awarded bursaries do not eventually enter teaching because, unlike in many states in the US, there are no bonds or tie-ins to commit bursary recipients to teaching. The effect of financial incentives is also not consistent across genders and age groups. Wage premiums, for example, are potentially more effective in attracting young female teachers than older male teachers, but more effective in retaining older male teachers.

As for retention, financial incentives do not seem as be as effective. Although many studies do show positive results, the more robust studies which control for context suggest that teachers only stay while the incentive is available. Such short-term results are not useful in solving the chronic shortage of teachers. In fact, the evidence suggests that the use of discriminatory incentives may even worsen overall retention. Eligibility for an incentive, or a small incentive, seems to make little difference. Where incentives are used, they need to be substantial.

In many cases, monetary incentives work only because teachers are required to commit to teach for a specified period or certain subjects in specified schools or areas as part of the contract agreement. These incentives often entail a penalty for breaking the contract, raising questions about the value of such an approach and the potential for a kind of enforced retention where teachers feel 'tied-in' to a role that they no longer wish to do.

In recent months, in England, the government have announced pay increases for teachers across the board [132,133]. These are not specific incentives nor attached to particular individuals, subjects or regions as we have examined in the section above. However, the plans do indicate that teachers new to the profession will receive a higher increase than those who are more experienced. The link between teacher pay and recruitment/retention is still fairly unclear but it will be important to examine whether these reforms do appear to have any impact on the number or type of graduates entering teaching, or the number of teachers choosing to continuing working in schools.

5.2. Beyond Financial Incentives—Implications for Policy, Practice and Research

Financial incentives have been used for decades to try and encourage increased numbers of graduates to enter the teaching profession. Despite the evidence that they can have some positive impacts, there is nothing to suggest that they are likely to 'solve' the recruitment challenges that countries such as England and the US currently face on a longer term basis. Economists have argued that the use of monetary incentives needs to consider how they interact with intrinsic and social motivations and what happens when they are withdrawn [134].

Thus, it is important that policymakers and school leaders look to alternative or additional approaches too. Relying primarily upon financial inducements is unlikely to be an effective and sustainable strategy for recruitment and retention, and arguably it is not desirable either. It is not clear, for example, whether the extrinsic motivation offered by financial incentives, leads to the best or most-suited graduates entering teaching.

While monetary incentives may be potentially useful as a way to increase the supply of shortage subject teachers and to attract teachers to challenging schools or areas, we question, for example, whether they are necessarily the best approach to improving recruitment as it is quite clear that the attraction is not lasting. Where prospective or qualified teachers are asked to report the factors that influence their likelihood for entering or staying in teaching, or that might encourage those in the profession to leave, money rarely features highly [27,32,33,94]. Instead there are other factors which may offer more promise for recruitment and retention if addressed.

Support in schools for teachers in the early stages of their careers and continuing professional development for established teachers appear to offer potential benefits for retention but are also arguably important in their own right. The evidence for mentoring and professional development is uniformly positive for mentees but the studies that we have to make these judgements are unfortunately not of the strongest quality. The stronger studies do not show consistent positive effects [23,119]. There is also little evidence on the effectiveness of specific induction programmes for retaining new teachers. The few studies that have looked at this area are either methodologically fairly weak and/or report mixed or unclear findings. The stronger studies find little or no impact. A consistent issue though is the multi-faceted nature of these interventions which makes it difficult to identify, accurately measure and understand the elements of the induction programmes. It is not always clear whether it is the induction alone or a combination of other factors that makes a difference. Some of these studies also use 'intention' to stay in the profession as an outcome rather than actual attrition figures. These are issues which need to be factored into future design and evaluation of induction programmes.

In England, the government has recently introduced the Early Career Framework (ECF) with a view to providing teachers with a strong induction programme, including early professional support, mentoring and a reduced teaching timetable [28]. The potentially promising findings on some of these areas in this review are therefore welcome news, but given the relative weakness of the studies, it is not clear to what extent the ECF is an evidence-informed initiative, or how confident we can be in its outcomes. Robust evaluations of the ECF in its early years, however, would provide some much-needed evidence in this area and will be vital for informing ongoing iterations of the policy or those like it.

Our review also tentatively points to the importance of improving school cultures and ethos for recruitment and retention. While we found very few rigorous studies that evaluated interventions related to areas such as accountability, teacher stress, working conditions, behaviour, workload or levels of support from teachers/leaders, some of the correlational and survey-based studies indicate that these could be valuable areas to explore further. As we note above, there are likely myriad other reasons for improving some of these wider factors too, including pupil/teacher performance and wellbeing. These could well be enough to justify the trialling of interventions that seek to achieve such aims, with a view to also understanding their impact on recruitment and retention. While we acknowledge that measuring and evaluating some of these school-level approaches at scale is arguably

more complex than the evaluation of government-administered financial incentives, we would contend that such evaluations are vital and should form an integral part of any new policy initiative.

Many of the interventions also seemingly address the symptoms rather than the cause of teacher shortages. As See and Gorard [14] have shown, government policies that aim to improve the quality of teachers has led to a reduction in the number accepted into teacher training. Manipulating the number of teachers that can be trained in higher education institutions and reducing school funding all have ramifications on the number of teachers in schools. A more coherent and long-term approach to policies is therefore needed.

5.3. Strengths and Limitations of the Review

This review systematically and rigorously synthesised approaches and interventions used for improving the recruitment and retention of teachers. It is the only comprehensive single-study review that we know of which includes robust appraisal of the research design and methods used within each study. This quality-appraisal is key to the claims that we are able to make in terms of the most effective approaches and the strength or amount of evidence available to support them. While our review focuses on the inclusion of studies using experimental or quasi-experimental designs, we have also sought to acknowledge where evidence from other types of study might be helpful for understanding particular issues or highlighting where there are potential evidence gaps. Finally, although our review had broad parameters, included over 7000 studies at the outset and a process of careful and rigorous screening, the criteria that we applied do mean that it is of course possible that relevant and potentially useful studies have been missed or excluded.

Author Contributions: Design and methodology of the review, S.G., B.H.S. and R.M.; Formal analysis, B.H.S. and R.M.; Data extraction and evidence rating, B.H.S., R.M., S.G., S.A.; Original draft preparation and editing, D.K.; Writing—review and editing, B.H.S., R.M., S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Economic Social and Research Council, grant number ES/R007349/1.

Acknowledgments: We would like to thank Nada El Soufi for assisting with the database searches and screening of the data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Template used for extracting data from each included study

Overview

Brief Description of the Intervention

- Aim and type of intervention: e.g., financial incentives (performance-related pay, scholarships, bursaries, housing benefits, pension scheme)
- Phase: Primary/secondary/general
- Country:
- How the intervention works: There must be enough information to enable identification of key features of a successful intervention, if it works.

Method

Research Design

- Does it have a control and comparison group?
- Does it have pre- and post- event comparison?

- How is randomisation or other allocation to groups carried out?
- Was there an intervention?

Sample

- Size of sample
- How were samples identified?
- School characteristics, e.g., primary, secondary, rural, urban, challenging schools
- How many cases were lost at each stage?

Outcome Measures

What are the outcomes and how were they collected?

- Is there a pre-defined primary outcome, or is there an element of 'dredging' for success?

Analysis (if Relevant)

- What kind of analysis was carried out?
- Are there pre- and post-test comparisons?
- Are effect sizes cited or calculable?
- How was the performance of treatment and comparison groups compared?

Findings

- Reviewers' analysis of the results (re-calculate effect sizes if not estimated or if in doubt).

Commentary

Aspects of the study that might threaten or enhance its validity. This could include fidelity to treatment, quality of counterfactual, extraneous/confounding variables, other programmes going on that may have affected the results, and conflicts of interest.

Appendix B

A 'sieve for judging the trustworthiness of causal research studies (Gorard, 2017)

Design	Scale	Dropout	Outcomes	Other Threats	Rating
Fair design for comparison (e.g., RCT)	Large number of cases per comparison group	Minimal attrition, no evidence of impact on findings	Standardised pre-specified independent outcome	No evidence of diffusion or other threat	4🔒
Balanced comparison (e.g., RDD, Difference-in-Difference)	Medium number of cases per comparison group	Some initial imbalance or attrition	Pre-specified outcome, not standardised or not independent	Indication of diffusion or other threat, unintended variation in delivery	3🔒
Matched comparison (e.g., Propensity score matching)	Small number of cases per comparison group	Initial imbalance or moderate attrition	Not pre-specified but valid outcome	Evidence of experimenter effect, diffusion or variation in delivery	2🔒
Comparison with poor or no equivalence (e.g., volunteers)	Very small number of cases per comparison group	Substantial imbalance and/or high attrition	Outcome with issues of validity or appropriateness	Strong indication of diffusion or poorly specified approach	1🔒
No report of comparator	A trivial scale of study, or N unclear	Attrition not reported or too high for any comparison	Too many outcomes, weak measures, or poor reliability	No consideration of threats to validity	0

Appendix C

The table summarises pieces that were reviewed and rated below 2. These tended to be small-scale, have considerable attrition, no clear comparator, and/or rely on reports of intention to stay or leave.

Study	Strategy	Impact	Evidence
Adnot et al. 2017	Performance incentive (financial incentives)	Positive effect in keeping high-performing teachers in high-poverty schools but not in low-poverty schools	The analysis did not compare teacher retention rates before and after IMPACT nor did it evaluate whether IMPACT improve retention of teachers in general. The study was unable to identify high-performing teachers who leave DCPS because of IMPACT, the estimates indicated that replacing high-performing teachers who exit with teachers who perform similarly is difficult. Also leavers include both voluntary and involuntary leavers.
Afolabi 2013	Professional development (Cross Career Learning Communities)	Positive effect Fewer treatment teachers left teaching or moved from their school than control teachers	QED Groups were matched on individual and school characteristics Teachers participating in CCLC were already in schools with a culture of professional development (groups are not equivalent) The study period also coincided with economic recession which may explain the high retention and lower mobility
Barnett and Hudgens 2014	TAP (Teacher and Student Advancement Programme)	Small positive effect (ES = 0.05)	TAP schools are self-selected. These schools are likely to be different to the national average. Schools that stopped TAP were not included in the analysis. These maybe schools where the programme had not worked. In other words, only successful schools were considered in the analysis.
Beattie 2013	Mentoring	No difference between groups but teachers receiving support from full-release mentors reported more positive experience	Small sample (87) Some teachers were selected to receive full-release mentors and some to school-based mentors Evidence based on teachers' report of intention rather than actual attrition
Bemis 1999	Mentoring	There is no clear impact of mentoring on retention despite the author's claim that mentoring programs were found to be most influential on new teacher retention for elementary level teachers.	Small sample Retention based on teachers' self-report High attrition, therefore, those who did not respond may be different to those who did. The results are therefore not reliable. Districts with mentoring may be different to disctricts with no mentoring. Different attrition rate may be a reflection of differences in the districts.
Bobronnikov et al., 2013	Incentive grant	+ Increase in number going into teaching, 80% teaching in high need areas (but no comparator). Not enough data to calculate ES Unclear retention Majority indicated they'd stay on. But of the 6 states, 2 states showed negative impact (no comparison groups)	The study design was unable to test whether recipients of the Noyce programme would have gained teacher certification in STEM subjects and go on to teach in high needs areas in the absence of the programme

Study	Strategy	Impact	Evidence
Bond (2001)	Salary	+ States where salary was markedly lower than similarly-education professionals, there was higher teacher turnover and reverse is true (after controlling for family background)	It is a correlational analysis and the states being compared are not the same, there are confounding factors that are not accounted for.
Bowman 2007	Mentoring	Negative impact on retention Experimental teachers were more aware of the career commitment which negatively affect their withdrawal intention.	Small sample (n = 30) Comparison groups were not equivalent. Control teachers had more teaching experience than experimental teachers. No actual data on retention was collected
Brown &Wynn 2009	Role of principal	Positive effect of principal awareness of issues affecting teachers on retention	Not an impact evaluation
Cheng and Brown 1992	Peer support/mentoring	Mixed results • Positive effect in the first year (ES = +0.12) but no effect in the second year (ES = +0.03	Evidence was based on teachers' self-report. The sample was small and imbalance. The 2 groups were not equivalent. Comparison teachers were those that were not eligible for the programme. In the second year, comparison teachers were randomly selected to be in the experimental group. Experimental teachers were also designed to include those that did not have prior experience.
Chou 2015	Mentoring (full-time release for mentors with financial rewards)	Negative result of full-time release mentoring	The 2 school districts being compared are different and the sample size of only 23 is too small to make any sensible judgements on effectiveness.
Clamp 2011	Mentoring	No effect	Comparison groups were self-selected, coupled with the high attrition rates and the self-report survey, the evidence is weak.
Clewell and Villegas 2001	Alternative certification	Impact on recruitment unclear (more pathways graduates completed (75% vs 60%) and ended up teaching in HTSS (84% no comparison) than traditionally certified teachers + on retention ES = 0.1	Comparisons were made with national average and traditionally certified teachers. The 2 groups of people are therefore likely to be different. Paraprofessionals and emergency-certified teachers, for example, were already working in the schools. It is therefore, hardly surprising that they were more likely to stay in the school or district where they were trained. There was also no comparison of before and after data.
Colson and Satterfield 2018	Financial incentive (The Innovation Acceleration Fund grant, a compensation scheme)	+ impact on retention 80% of teachers on the scheme were retained compared to 70% not on the scheme (ES = 0.07)	The very small non-random sample, and exclusion of teachers who did not have TVAAS results meant that the sample might be biased. Comparisons were made with volunteers and non-volunteers
Counts 2012	Induction	Positive effect Administrative support and workload were the strongest predictor of teachers' commitment to stay in the school (R^2 = 0.19 for both).	Calculation of means was used for categorical variables (e.g., strongly agree to strongly disagree). Only 22% of teachers responded to the survey. The views of the majority 78% of new teachers were not captured.

Study	Strategy	Impact	Evidence
Cowman 2004	Alternative certification	Unclear results But looks like mentoring did not influence retention All programs had relatively high rates of retention; ACP had the highest retention (96.81%), followed by ECP (90%) and then CPDT (89.9%). CPDT teachers reported receiving the most support as they were paired with experienced teachers during the internship, they have the highest attrition. This suggests that factors other than mentoring and support could determine teachers' decision to leave. ACP had the highest retention rates likely because of their selective process.	Record of attrition may not be accurate. Teachers who are still teaching but have left the state of Texas are treated as teachers who have left the profession because their employment histories are no longer trackable. Those who left temporarily (e.g., maternity) sare treated as having left teaching.
Croffut 2015	Mentoring and Induction	No effect Turnover rate of beginning teachers in the district decreased by 1 percentage point between 2012–2014 and 2014–2015. Comparing teachers' self-report intention to stay or not, showed no difference between expected and actual response rate. In fact, actual response rate was 88% compared to the expected rate of 90%.	High level of missing data (only 29% responded to survey). Therefore responses could be from self-selected individuals. Evidence of bias in reporting Despite the data showing no effect, the author concluded "While there is no statistically significant difference, the data reveal the district is maintaining the beginning teacher turnover rate which would indicate the district's beginning teacher program is positively impacting the teacher retention rate"
Dwinal 2012	Alternative certification (Teach For America)	No effect	Based on interviews with superintendents and principals with low response rates (under 20%). Poor reporting. Based on vacancies not placements.
Eberhard, Reinhardt-Mondragon and Stottlemyer 2000	Mentoring and Alternative Certification	+ effect of mentoring (compared to no mentoring) + effect of alternative certification (compared to standard certification) Negative effect of emergency certification compared to fully certified teachers	The groups were no randomly selected and as the authors reported, this may be reflective of the kind of pre-service students who would sign up for the more intensive one-year programme. No actual retention data presented.
Elmore 2003	Mentoring	No difference in retention rates although retention of teachers using MTC continued to increase over 2 years while those using peer mentors continued to decrease	No pure control Comparison was with Peer Mentors and Mentor Teacher Consultants Schools were selected for MTC based on high turnover rates and low performance. Schools are therefore different
Fleener 1998	Alternative certification	Positive effect for field-based training (2.1% attrition) compared to university-based training (6.7%)	The 2 groups are self-selected so may be different in terms of motivation and commitment. Also a large number who did not end up in state-funded teaching were excluded. This may have already excluded those who would be likely to leave teaching anyway

Study	Strategy	Impact	Evidence
Fowler 2003	Massachussets Signing Bonus	- No effect on recruiting to high need districts (no comparator, so cannot calculate ES)	There was no comparison group. It was simply an analysis of the data on bonus recipients and their outcomes.
Fuller (2003)	Mentoring	+ effect on retention Although differences in retention rates of participants and non-participants are "significant" effect sizes calculated by reviewer are small (around 0.05 for all the 3 years)	Participants were self-selected or "qualified" for inclusion. Therefore groups being compared were different. The programme had a lot of components, so it was difficult to isolate the effects of mentoring In some all beginning teachers had a mentor, in others there were few or no mentoring for new teachers
Gaikhorst et al., 2015	Professional development for beginning teachers	No effect on retention	Evidence based on teachers' report of their intention to stay. Experimental teachers were those who volunteered to take part. These were compared with those who did not take part
Gold 1987	Mentoring (New York City retired teachers-as-mentors programme)	Lowers attrition rates among mentored teachers compared to non-mentored, but tiny numbers	This was a small-scale RCT. Although principals were asked to assign mentors at random, it was not clear how this was done. In some cases teachers rejected the offer of a mentor. Assignment was therefore no longer random
Goldhaber, Destler and Player 2010	Financial incentives	+ effect Additional $5790 needed for a 50% increase in number of teachers teaching in schools with high proportion of minority children, but only $706 extra for a 50% increase in number of teachers teaching in high poverty schools	Not focused on recruitment and retention specifically
Gordon and Vegas 2004	FUNDEF (Financial incentives)	Increase in number of teachers in poorer regions but no effect on proportion of secondary teachers with higher degrees	Not relevant to English context (funding reform in Brazil). The analyses are correlational and did not take into account other confounding factors
Hancock 2008	External support, mentoring and induction and financial incentives	Mentoring and induction did not predict likelihood of attrition Parent and administrative support reduced the risk of attrition Salary is also significant. For every I unit increase in salary bracket (c. $10,000), there is a 38% reduction in risk (OR = 0.62).	The evidence is based on a large sample of participants based on administrative data. But because the evidence is based on self-report of intention to stay or leave, the evidence is not strong

Study	Strategy	Impact	Evidence
Hansen et al., 2016	Alternative certification (Teach for America)	Effects are mixed. Clustering has a positive effect on retention of teachers in schools in the district. The higher the density of TFA corps members in a school increases, they are less likely to move schools within district However, it has a negative effect on retention of teachers within district. A 1 percentage point increase in TFA density in the school is associated with a 1.5% greater likelihood of exiting the district	This study can only establish correlation but not causality. It also cannot determine the direction of causation. It is possible that schools with high out-of-district exits are more likely to rely on TFA staffing.
Hardie 2008 [full paper not available)	Alternative preparation	No effect on retention	The two groups of teachers were not randomly allocated and no controls were made of teacher background characteristics
Harrell and Harris 2006	Alternative certification (Online post-baccalaureate teacher certification programme)	+ effect on recruiting males (ES = 0.2) and minority candidates (ES = 0.19) + effect on recruiting maths and science teachers (ES = 0.2) + effect on recruiting career changers (no comparison for ES calculation)	Because of self-selection into programmes candidates who signed up for traditional programmes are likely to be different to those who signed up for the online programme. The groups are therefore not balanced. Also comparison is made for only one year, it is not possible to rule out other exogenous factors (e.g., economic performance) which may have affected a larger number of people who change career Data was taken from one faculty in one institution and for one academic year only. Sample may not be generalised to other years and institutions. Hence the 1 🔒 rating.
Harris-McIntyre 2014	Induction	No clear effect No evidence that alternative (on-the-job training as in Teach First in England) has been effective in retaining teachers in the district. However, non lateral teachers were over twice more likely to stay in teaching in the first and second year, but no difference in the 3rd year	The teachers were neither randomised nor matched by background characteristics. There are likely to be unobservable differences which have not been controlled for in the analysis.

Study	Strategy	Impact	Evidence
Henke, Chen and Geis 2000	Induction	+ effect on retention (15% left compared to 26% not on induction programme, ES = 0.27)	Used data from the Baccalaureate and Beyond Longitudinal Survey (n = 7294) It is not clear how many missing cases there were that had not been accounted for. Also the two groups may be different as teachers participating in induction programmes may be in more supportive schools with better working conditions etc. So it is not possible to attribute the lower attrition rate simply to induction alone. • The analysis is based on bivariate correlations between two factors. It could not account for unobserved factors.
Henry, Bastian and Adrienne 2012	Financial Merit-based scholarships	+ recruitment of high quality graduates (SAT scores of high school scholars 113 points higher than traditionally prepared teachers and GPA scores are 0.6 points higher non-teaching fellows; ranked among the top 10% of graduates) + retention (scholarship recipients more than 1.1 times more likely to stay on for 5 years than other in-state prepared teachers)	Comparisons were not made with similar teachers Scholarship recipients were high-flying graduates who applied and were therefore self-selected. Unobserved confounders such as scholars' motivations and intentions could not be controlled for.
Hopkins 1997	Induction	No effect on retention (Effect size = 0.03)	Groups not equivalent Missing cases and non-response meant that the groups were no longer balanced Retention based on reported intention
Humphrey et al., 2018	Behaviour management as CPD	No impact on teacher retention (ES = −0.01)	A lot of missing data Low compliance No actual retention data (based on teachers' expression of intention)
Ingersoll, Merrill and May 2014	Teacher preparation	Positive effect Those that have more pedagogy in their training were less likely to leave Training in teaching strategies and methods made no difference	The study could not control for unobserved differences. Those who chose the traditional teacher preparation route may view teaching as a career to which they are committed. Those with an education degree may be more committed to teaching because they have fewer alternative career options than those with a maths or science degree.
Jacobson 1988	Salary differentials	+ recruitment (positive correlation between entry-level salary ranking and recruitment of highly qualified teachers + retention (positive correlation between salary ranking of mid-career teachers and retention of mid-career teachers)	It is correlational in design, it is not able to control for other confounding factors such as the economic and political differences in the districts

Study	Strategy	Impact	Evidence
Jones 2004	Mentoring	No effect No difference between the in-house and full-time mentoring in terms of teachers' reported intention to stay (Cramer's V effect size = 0.0067) No differences between the two groups in terms of reasons for leaving Lack of collaboration with colleagues and administrative and mentor support as top reasons for leaving	1 🔒 Schools offering Full-Time mentoring programme were selected based on certain criteria, not randomised. Measure of retention was based on participants' self-report.
Kelley 2004	Induction and mentoring	Positive effect on retention	Compare 10 cohorts of new teachers with national average. These teachers were self-selected based on their qualifications and also they received higher salaries after completion than most novice teachers. The number involved in each year is small (under 50)
Kelly and Northrop	Teacher preparation	Teachers from less selective training colleges are less likely to leave their school (including moving school and leaving profession	Those from highly selective colleges may have greater job opportunities. Large amount of missing data. Very small sample from selective colleges.
Lawrason 2008	Teacher induction	Some positive responses but weak links	Results collected from surveys of participants' reported intention (compared with other induction programmes) Small sample of 54
Lyons 2007	Induction programme (known as left X programme)	+ effect • Beginning teachers who were exposed to all programme types (i.e., better prepared) were less likely to leave classroom teaching or education than those who were not.	This study was based on a comparison of observed and predicted rates of retention using logistic regression analysis to control for observable characteristics.
McBride 2012	Induction and mentoring	Positive effect Association between induction and mentoring variables, and likelihood of teacher remaining in teaching for the following year	Uses 3 admin datasets looking at the outcomes of those involved in induction and mentoring.
McGlamery and Edick 2004	Teacher induction The CADRE project	Positive effect Compared with national sample (40% attrition rate), retention of CADRE participants was 89% over 5 years	153 1st and 2nd year CADRE teachers Risk of selection bias
Mordan 2012	Mentoring of beginning Career and Technical Education teachers	Positive effect on retention. Beginning CTE teachers assigned a mentor were 6.64 times more likely to remain in teaching	Uses 3 admin datasets (SASS, TFS and BTLS) Weak comparisons Small target group (N = 110) Focus of study was on teachers' experience rather than retention outcomes

Study	Strategy	Impact	Evidence
Morrell and Salomon (2017)	Scholarship scheme	Inconclusive	Claims that it was successful in assisting undergraduates with a STEM background into teaching, but not supported by the data
Murphy 2004	Grow Your Own (A collaborative partnership with local education agencies, community colleges, private and public schools)	Positive effect Large percentage of participants who have received Consortium services have remained in continuous employment in North Carolina's schools	Weak causal evidence Focus on participants in the Consortium programmes No comparison with non participants
Odell and Ferraro 1992	Mentoring	+ effect on retention	There was no control group and the groups were not matched nor was there an attempt to find similar, or matched districts to serve as the comparison. This is important since the districts in question might have already been higher-retaining districts (or at least higher than the state average.
Ogunyemi 2013	Mentoring	Some claims about perceived impact of mentoring on retention	Self-report, no comparison group and high attrition
Oliver 2016	Mentoring	Suggests that the use of social media platform increases retention of induction year maths teachers	Ethnographic accounts based on participant observations and field notes—not a study which aims to find causal/correlational outcomes linked to retention
Parker, Ndoye and Imig 2009	Mentoring	Positive effect of same subject and grade level mentors on retention	Sample included 8838 beginning teachers being mentored for 2 years. Outcome was teachers' intention to stay not actual retention
Partridge 2008	Mentoring	No effect of mentoring on participants' intention to stay	Survey based on 71 teachers (only 12 were assigned a mentor). The data was delimited to information provided by a portion of elementary teachers in one public school district so might not reflect the opinion of all members of the included population. Responses were subject to the validity of self-perceptions regarding mentoring.
Perry 2008	Induction	Minority teachers	Small sample (n = 22). No clear data presented to make judgements about the validity of the findings
Protik et al., 2015	Cash transfer incentive	No effect—uptake was low	0 No comparison so not possible to say what the uptake would be in the absence of the incentive
Quartz 2003	Induction and ongoing professional development in left X	Positive effect Over 5 years 70% of left X graduates remain in classroom compared to 61% nationally based on SASS (ES = 0.69)	Comparison with national figures Participants were self-selected (bias selection) The focus of the study is on the reason why teachers stay or leave
Randall 2009	Mentoring	The teachers reported that the mentors had no effect on their decision to remain in the classroom.	Not impact evaluation.

Study	Strategy	Impact	Evidence
Reynolds and Wang 2005	Professional development	Positive effect PDS graduates less likely to leave teaching (20%) than non-PDS graduates (17%) ES = 0.26	Compared PDS with non-PDS graduates High attrition/nonresponse
Reynolds, Ross and Rakow 2002	Professional development	No effect No retention differences between PDS and non-PDS route	Small sample (N = 191) Attrition 58% No data on retention presented
Ridgely 2016	Induction	Compare two models of induction. Suggests that dual-role induction was more effective in keeping teachers thana site-based induction.	Comparison was between 2 types of induction programme. No counterfactual. So cannot rule out other differences between the 2 districts who could have explained the different retention rates. There was also a huge disparity in numbers between the two districts being compared.
Robertson-Kraft 2014/2018	Teacher performance management	Quicker turnover rates in INVEST pilot schools Paperwork relating to INVEST contributed to wanting to leave	Schools are not randomly allocated High non-response No report of actual retention data (based on teacher's self-report)
Robertson-Phillips 2010	Teacher induction Beginning Teacher Support and Assessment Program	No effect on retention Retention of BTSA teachers similar to the intern programme	Compared RIMS/BTSA teachers with intern teachers Groups not randomly assigned Data based on perceptions of participants
Rothstein (2015)	Types of contract (permanent vs temporary	No impact on supply. Bonus contract is less effective than the tenure contract in increasing the number of high ability teachers (ES +0.004 and +0.033 respectively). Retention policies are effective only if there is substantial increase in salary. If budget is fixed, may need to increase class sizes to offset the higher salary of teachers	The models are based on a number of caveats which are not possible in reality. It assumes that teacher performance assessment is unbiased and that new teachers are recruited from the same population as current teachers ignoring the fact that there are potentially high ability teachers who would not consider teaching at all.
Rogers 2015	Induction	Found no link between induction programme and retention	Online survey, very low response (34%), no clear comparator. Evidence based on school leaders' and administrators' report. No actual retention data
Scott et al. (2006)	Scholarship, tuition fee remission and mentoring	+ effect on recruitment (an increase of over 100% from in 37 1st year to 80 in the 3rd year). In the 4th year 100 enrolled 80% indicated that they would stay on. (no comparison group). Retention is based on participants' self-report of intention to stay on the course, not teaching in general.	There is no comparison group, so it is not possible to attribute the increase in the number of students enrolled on the teacher certification course solely to the MASS programme. The retention rate is the retention on the programme and is based on students' report of their intention rather than actual staying on

Study	Strategy	Impact	Evidence
Shen, J. 1997	Alternative route to teaching	Successful in recruiting minority and shortage subject teachers and increasing supply of teachers in urban areas However, AC teachers tend to have lower qualifications AC less successful in attracting experience personnel from other occupations Most new college graduates opted for the AC to avoid the traditional teacher education programme AC teachers less likely to treat teaching as a lifelong career No impact on retention (retention not measured but based on participants' report of intention to stay)	Given that AC and TC teachers were not randomised there are important differences between them. Those who chose the AC route may have different motivations from those who chose the TC route. It's also possible that those who entered via the AC route were not eligible for the TC programme because of their lower academic qualifications.
Shepherd 2009		Claimed that the Induction program had a positive effect, but given the data presented, it is not possible to know if this can be attributed to the program.	Data gathered from stakeholders through surveys, focus group discussions and interviews. No causal/correlational evidence clearly presented. Poor reporting of samples.
Sims (2017)	Salary compensation	+ effect on recruitment and retention Increase in the total supply of teachers (recruitment deficit ES = 1.3 for science and 1.4 for maths	The model made a number of assumptions, e.g., Teachers missing in the School Workforce are taken to have left teaching, the reduction in probability of leaving the profession is evenly spread across each year of the policy, Increased pay does not incentivize more people to train in each cohort
Spuhler and Zetler 1993–1995	Mentoring	Positive effect on retention. In the second year 92% of mentored teachers compared to 73% of non-mentored teachers were still teaching. Effect size is 0.12. In the 3rd year all the mentored teachers continued teaching but only 70% of non-mentored teachers remained in teaching (ES = 0.12)	The small sample size meant that the results could not be generalised. The comparison teachers were not matched in any way.
Stinebrickner 1998	Wages	+ impact on retention Teachers paid higher salary 9% more likely to stay on in teaching for more than 5 years than teachers paid the mean wage Attrition was 70%, hence the 1🔒	The data is poor with only 30% of teachers being tracked. We are therefore not sure how different the results would be if data for all the teachers were available. Those that did not respond are likely to be different to those who did. Also the survey asked teachers to recall their teaching experience. This can be subjective depending on their experience at the time of the survey and may not accurately reflect what actually happened.
Tai, Liu and Fan (2006)	Alternative certification of maths and science teachers	No difference between alternative and traditionally certified teachers	Used admin data (SASSand TFS) Missing data Lapse time between SASS and TFS is only one year. Longer evaluation needed to test sustained effect

Study	Strategy	Impact	Evidence
Toterdell, Heilbronn, Bubb and Jones 2002	Induction	Focused on the positive experience of NQTs	Not impact evaluation. Limited focus on retention or attempts to measure this in a coherent way. Looks at perceptions of new programme and some implementation but little in the way of actual outcomes.
Troutt 2014	Professional Learning Communities (PLCs)	Claims PLCs improve retention	No pre- post comparison. Made conclusions based on comparison of a high retention and low retention school. The schools may be systematically different in terms of pupil intake, location etc, which could have influenced retention. Therefore, not possible to attribute success to the programme. Used school-level rather than individual teacher retention Poor reporting.
Uttley 2006	Mentoring	Suggests positive effect	Evidence based on survey of teachers' perceptions about the effectiveness of the programme, collected at one time point. Non response was 45%.
Van Overschelde, Saunders and Ash 2017	Professional development programme Texas State University teacher preparation programme	Positive effect 85% of Texas State University's graduates teaching after 5 years compared to 71% for average state retention rate (ES =0.9) Retention also higher.	Comparison institutions not randomly allocated. Did not control for teacher and institutional characteristics.
Wells 2011	Financial incentives Team performance pay	No effect in the 1st and 2nd year	Difference-in-difference approach comparing retention before, during implementation and a year later Teachers' report of retention and the district data not consistent
Wilkinson 2009	Induction for alternative certification programme students		Comparisons were made with 7 different cohorts of students, who were lumped together as one despite possible differences in contexts/backgrounds. Evidence based on survey collecting respondents' report of satisfaction with the programme and correlation analysis of their responses with their intention to stay
Zavala 2002	Alternative certification vs field-based training	CPDT (field-based training) appears to impact retention positively	Two types of teacher preparation not randomly assigned. So not sure how field-base training is compared to traditional teacher preparation.
Zhang and Zeller 2016	Alternative routes into teaching	Long-term retention rates are greater for traditional certification programme than ACP	Small sample (58 teachers were tracked over 7 years. 22 regular, 20 lateral entry and 18 NC teachers. Groups self-selected not randomly assigned.
Zumwalt et al., 2017	Alternative route to teaching	• Positive results for recruitment but weak evidence as not comparison group data available. • Negative results for retention of maths teachers	The evidence is weak as these measures were largely based on correlation and pre-post comparisons without any control. e.g., the increase in the proportion of qualified primary teachers coincided with the legislation that teachers should be qualified.

References

1. Sibieta, L. *Teacher Shortages in England: Analysis and Pay Options*; Education Policy Institute: London, UK, 2020.
2. DfE. *School Workforce Census: Teachers Analysis Compendium 4*; DfE: London, UK, 2018.
3. Gerritsen, S.; Plug, E.; Webbink, D. Teacher quality and student achievement: Evidence from a sample Dutch twins: Teacher quality and student achievement. *J. Appl. Econom.* **2016**, *32*, 643–660. [CrossRef]
4. Goe, L. *The Link between Teacher Quality and Student Outcomes: A Research Synthesis*; National Comprehensive Center for Teacher Quality: Washington, DC, USA, 2007.
5. Sanders, W.L.; Horn, S. Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *J. Pers. Eval. Educ.* **1998**, *12*, 247–256. [CrossRef]
6. Sorensen, L.; Ladd, H. *The Hidden Costs of Teacher Turnover*; Working Paper No. 203-0918-1; National Center for Analysis of Longitudinal Data in Education Research (CALDER): Washington, DC, USA, 2018.
7. Sutton Trust. *Improving the Impact of Teachers on Pupil Achievement in the UK—Interim Findings*; Sutton Trust: London, UK, 2011.
8. European Commission/EACEA/Eurydice. *Teaching Careers in Europe: Access, Progression and Support. Eurydice Report*; Publications Office of the European Union: Luxembourg, 2018.
9. Dee, T.; Goldhaber, D. *Understanding and Addressing Teacher Shortages in the United States*; Brookings: Washington, DC, USA, 2017.
10. Aldeman, C. Teacher Shortage? *Blame the Economy. TeacherPensionsorg*. 2015. Available online: http://educationnext.org/teacher-shortage-blame-the-economy/ (accessed on 12 December 2019).
11. Ingersoll, R. Do we produce enough maths and science teachers? *Phi Delta Kappan.* **2011**, *92*, 37–41. [CrossRef]
12. Hutchings, M. *What Impact does the Wider Economic Situation Have on Teachers' Career Decisions? A Literature Review*; DfE Research Report DfE-RR136; DfE: London, UK, 2011.
13. Dolton, P.; Tremayne, A.; Chung, T. *The Economic and Teacher Supply. A Paper Commissioned by the Education and Policy Division, OECD for the Activity Attracting, Developing and Retaining Effective Teachers*; OECD: Paris, France, 2003.
14. See, B.H.; Gorard, S. Why don't we have enough teachers? A reconsideration of the available evidence. *Res. Pap. Educ.* **2020**, *35*, 416–442. [CrossRef]
15. Lortie, D.C. *Schoolteacher: A Sociological Study*; University of Chicago: Chicago, IL, USA, 1975.
16. Ingersoll, R.; Smith, T. The wrong solution to the teacher shortage. *Educ. Leadersh.* **2003**, *60*, 30–33.
17. Ward, H. One in Three Teachers Leaves within Five Years. *Times Educational Supplement*. 2019. Available online: https://www.tes.com/news/one-three-teachers-leaves-within-five-years (accessed on 15 July 2020).
18. Ingersoll, R. *Is There Really a Teacher Shortage?* University of Pennsylvania, Consortium for Policy Research in Education: Philadelphia, PA, USA, 2003.
19. Ingersoll, R.; Perda, D. Is the supply of mathematics and science teachers sufficient? *Am. Educ. Res. J.* **2010**, *47*, 563–595. [CrossRef]
20. Sutcher, L.; Darling-Hammond, L.; Carver-Thomas, D. *A Coming Crisis in Teaching? Teacher Supply, Demand and Shortages in the US*; Learning Policy Institute: Palo Alto, CA, USA, 2016.
21. Worth, J.; De Lazzari, G. *Teacher Retention and Turnover Research. Research Update 1: Teacher Retention by Subject*; NFER: Slough, UK, 2017.
22. Bowsher, A. Recruiting the 'Best and Brightest': Factors that Influence Academically-Talented Undergraduates' Teaching-Related Career Decisions. Dissertation Abstracts International Section A: Humanities and Social Sciences. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2016.
23. Glazerman, S.; Seifullah, A. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years. Final Report*; Mathematica Policy Research, Inc.: Princeton, NJ, USA, 2012.
24. Fryer, R.G., Jr.; Levitt, S.D.; List, J.; Sadoff, S. *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment (No. w18237)*; National Bureau of Economic Research: Cambridge, MA, USA, 2012.
25. Springer, M.G.; Ballou, D.; Hamilton, L.; Le, V.; Lockwood, J.; McCaffrey, D.; Pepper, M.; Stecher, B. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching. NCPI Policy Evaluation Report*; National Center on Performance Incentives: Nashville, TN, USA, 2011.
26. Ingersoll, R.; May, H. *Recruitment, Retention, and the Minority Teacher Shortage*; University of Pennsylvania, Consortium for Policy Research in Education: Philadelphia, PA, USA, 2011.

27. Cooper-Gibson Research. *Factors Affecting Teacher Retention: Qualitative Investigation. Research Report*; DfE: London, UK, 2018.
28. DfE. *Early Career Framework*; DfE: London, UK, 2019.
29. DfE. *Teacher Recruitment and Retention Strategy*; DfE: London, UK, 2019.
30. Torres, C. Is This Work Sustainable? Teacher Turnover and Perceptions of Workload in Charter Management Organizations. *Urban Educ.* **2014**, *51*, 891–914. [CrossRef]
31. Higton, J.; Leonardi, S.; Richards, N.; Choudhoury, A.; Sofroniou, N.; Owen, D. *Teacher Workload Survey 2016*; DfE: London, UK, 2017.
32. Lynch, S.; Worth, J.; Bamford, S.; Wespieser, K. *Engaging Teachers: NFER Analysis of Teacher Retention*; NFER: Slough, UK, 2016.
33. Perryman, J.; Calvert, G. What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *Br. J. Educ. Stud.* **2020**, *68*, 3–23. [CrossRef]
34. DfE. *Analysis of School and Teacher Level Factors Relating to Teacher Supply*; DfE: London, UK, 2017. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/682023/SFR86_2017_Main_Text.pdf (accessed on 10 December 2019).
35. Figlio, D.; Kenny, L.W. Individual teacher incentives and student performance. *J. Public Econ.* **2007**, *91*, 901–914. [CrossRef]
36. Milanowski, A.T.; Longwell-Grice, H.; Saffold, F.; Jones, J.; Schomisch, K.; Odden, A. Recruiting New Teachers to Urban School Districts: What Incentives Will Work? *Int. J. Educ. Policy Leadersh.* **2009**, *4*, 1–13. [CrossRef]
37. Prince, C.D. *Higher Pay in Hard-To-Staff Schools: The Case for Financial Incentives*; Scarecrow Press: Lanham, MD, USA, 2003.
38. Goodnough, A.; Kelley, T. Newly certified teachers, looking for a job, find a paradox. *New York Times*, 1 September 2000.
39. Boyd, D.; Lankford, H.; Loeb, S.; Wyckoff, J. *Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets*; Working Paper 9878; National Bureau of Economic Research: Cambridge, MA, USA, 2003.
40. Hanushek, E.; Kain, J.; Rivkin, S. Why public schools lose teachers. *J. Human Resour.* **2004**, *39*, 326–354. [CrossRef]
41. Sibieta, L. *The Teacher Labour Market in England: Shortages, Subject Expertise and Incentives*; Education Policy Institute: London, UK, 2018.
42. Sims, S. *What Happens When You Pay Shortage-Subject Teachers More Money? Simulating the Effect of Early-Career Salary Supplements on Teacher Supply in England*; The Gatsby Charitable Foundation: London, UK, 2017.
43. Anders, J.; Bryson, A.; Horváth, H.; Nasim, B. The Effects of Pay Decentralisation on Teachers' Pay and Teacher Retention. Available online: https://scholar.google.co.uk/scholar?hl=en&as_sdt=0%2C5&q=effect+of+pay+decentralisation+on+teachers%27+pay+and+teacher+retention&btnG= (accessed on 10 December 2019).
44. Burgess, S.; Greaves, E.; Murphy, R. *Evaluation of Teachers' Pay Reform*; DfE: London, UK, 2017.
45. Karbownik, K. *The Determinants of Teacher Mobility in Sweden (No. 2014: 13)*; Working Paper; Institute for Evaluation of Labour Market and Education Policy (IFAU): Uppsala, Sweden, 2014.
46. Duffrin, E. What's the value in value-added? *Dist. Adm.* **2011**, *47*, 48–50.
47. Greenwald, R.; Hedges, L.; Laine, R. The effect of school resources on student achievement. *Rev. Educ. Res.* **1996**, *66*, 361–396. [CrossRef]
48. Rivkin, S. *Teacher Characteristics, Market Forces, and Distribution of Teacher Quality among Schools and Districts [Commissioned Paper]*; National Center for Education Statistics: Washington, DC, USA, 2007.
49. Kearney, S. Understanding beginning teacher induction: A contextualized examination of best practice. *Cogent Educ.* **2014**, *1*, 967477. [CrossRef]
50. Martin, K.L.; Buelow, S.M.; Hoffman, J.T. New teacher induction: Support that impacts beginning middle-level educators. *Middle Sch. J.* **2016**, *47*, 4–12. [CrossRef]
51. Jenkins, C. The Impact of a Teacher Induction Program on Teachers' Perceived Persistence. Ph.D. Thesis, Texas Southern University, Houston, TX, USA, 2012.
52. Totterdell, M.; Woodroffe, L.; Bubb, S.; Daly, C.; Smart, T.; Arrowsmith, J. *What Are the Effects of the Role of Mentors or Inductors Using Induction Programmes for Newly Qualified Teachers (NQTs) on Their Professional Practice, with Special Reference to Teacher Performance, Professional Learning and Retention Rates? Research Evidence*

in Education Library; EPPI-Centre, Social Science Research Unit, Institute of Education, University of London: London, UK, 2008.
53. Darling-Hammond, L.; Hyler, M.E.; Gardner, M. Effective Teacher Professional Development. 2017. Available online: https://www.yu.edu/sites/default/files/inline-files/Effective_Teacher_Professional_Development_REPORT.pdf (accessed on 15 July 2019).
54. ColdwelL, M. Exploring the influence of professional development on teacher careers: Developing a path model approach. *Teach. Teach. Educ.* **2017**, *61*, 189–198. [CrossRef]
55. Humphrey, N.; Ra, H.; Ashworth, E.; Frearson, K.; Black, L.; Petersen, K. *Good Behaviour Game. Evaluation Report and Executive Summary*; Education Endowment Foundation: London, UK, 2018.
56. Glazerman, S.; Isenberg, E.; Dolfin, S.; Bleeker, M.; Johnson, A.; Grider, M.; Jacobus, M. *Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study*; NCEE 2010–4027; National Center for Education Evaluation and Regional Assistance: Washington, DC, USA, 2010.
57. Allen, R.; Sims, S. *Improving Science Teacher Retention: Do National STEM Learning Network Professional Development Courses Keep Science Teachers in the Classroom*; Wellcome Trust: London, UK, 2017.
58. Worth, J.; Van den Brande, J. Teacher Autonomy: How Does It Relate to Job Satisfaction and Retention? Available online: https://www.nfer.ac.uk/teacher-autonomy-how-does-it-relate-to-job-satisfaction-and-retention/ (accessed on 20 September 2020).
59. Allensworth, E.; Ponisciak, S.; Mazzeo, C. *The Schools Teachers Leave: Teacher Mobility in Chicago Public Schools Consortium on Chicago School Research at the University of Chicago Urban Education Institute*; University of Chicago: Chicago, IL, USA, 2009.
60. Boyd, D.; Grossman, P.; Ing, M.; Lankford, H.; Loeb, S.; Wyckoff, J. The influence of school administrators on teacher retention decisions. *Am. Educ. Res. J.* **2011**, *48*, 303–333. [CrossRef]
61. Marinell, W.H.; Coca, V.M. *Who Stays and Who Leaves? Findings from a Three Part Study of Teacher Turnover in NYC Middle Schools*; The Research Alliance for New York City Schools: New York, NY, USA, 2013.
62. Johnson, S.M.; Kraft, M.A.; Papay, J.P. How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teach. Coll. Rec.* **2012**, *114*, 1–39.
63. Sims, S. *TALIS 2013: Working Conditions, Teacher Job Satisfaction and Retention*; DfE: London, UK, 2017.
64. Loewus, L. Does It Make Sense to Offer Housing Perks for Teachers? *Education Week*. 23 January 2018. Available online: https://www.edweek.org/ew/articles/2018/01/24/does-it-make-sense-to-offer-housing.html (accessed on 20 September 2020).
65. Kelley, C.; Finnigan, K. Teacher compensation and teacher workforce development. *Yearb. Natl. Soc. Study Educ.* **2004**, *103*, 253–273. [CrossRef]
66. Lynch, M. Recruiting, Retaining, and Fairly Compensating Our Teachers. *Int. J. Progress. Educ.* **2012**, *8*, 121–135.
67. Przygocki, W.F. Teacher retention in Catholic schools. *Cathol. Educ. J. Inq. Pract.* **2004**, *7*, 523–547. [CrossRef]
68. Fore, C.; Martin, C.; Bender, W.N. Teacher burnout in special education: The causes and the recommended solutions. *High Sch. J.* **2002**, *86*, 36–44. [CrossRef]
69. Billingsley, B.S. Special education teacher retention and attrition: A critical analysis of the research literature. *J. Spec. Educ.* **2004**, *38*, 39–55. [CrossRef]
70. SACE. *A Review of Teacher Demand and Supply*; South African Council for Educators: Centurion, South Africa, 2010.
71. Achinstein, B.; Ogawa, R.; Sexton, D.; Freitas, C. Retaining teachers of color: A pressing problem and a potential strategy for "hard-to-staff" schools. *Rev. Educ. Res.* **2010**, *80*, 71–107. [CrossRef]
72. Dauksas, L.; White, J. Should I stay or should I go? How teacher leadership can improve teacher retention. *AASA J. Scholarsh. Pract.* **2010**, *7*, 27–32.
73. Newton, R.M.; Witherspoon, N. Recruiting teachers, principals, and superintendents: A job choice theory perspective. *AASA J. Scholarsh. Pract.* **2007**, *3*, 37–43.
74. Oke, A.O.; Ajagbe, M.A.; Ogbari, M.E.; Adeyeye, J.O. Teacher retention and attrition: A review of the literature. *Mediterr. J. Soc. Sci.* **2016**, *7*, 371. [CrossRef]
75. Borman, G.; Dowling, N. Teacher attrition and retention: A metaanalytic and narrative review of the research. *Rev. Educ. Res.* **2008**, *78*, 367–409. [CrossRef]

76. Guarino, C.M.; Santibañez, L.; Daley, G.A.; Brewer, D.J. *A Review of the Research Literature on Teacher Recruitment and Retention*; RAND: Santa Monica, CA, USA, 2004.
77. Hanover Research. *Review of Teacher Incentive Programs*; Hanover Research: Washington, DC, USA, 2014.
78. Gunther, J. Evaluating the Non-Monetary Factors in Teachers' Employment Decisions. Ph.D. Thesis, Utah State University, Logan, UT, USA, 2018.
79. Laurence, W.; Hass, B.; Burr, E.; Fuller, B.; Gardner, M.; Hayward, G.; Kuboyama, E. *Incentives for Attracting and Retaining K-12 Teachers: Lessons for Early Education. Policy Brief*; Policy Analysis for California Education: Stanford, CA, USA, 2002.
80. Lonsdale, M.; Ingvarson, L. Initiatives to Address Teacher Shortage. *ACER Policy Brief.* **2003**, *5*, 1–47.
81. Gorard, S.; See, B.H.; Siddiqui, N. *The Trials of Evidence-Based Education: The Promises, Opportunities and Problems of Trials in Education*; Routledge: London, UK, 2017.
82. Slavin, R.; Smith, D. Effects of sample size on effect size in systematic reviews in education. In Proceedings of the Annual Meetings of the Society for Research on Effective Education, Crystal City, VA, USA, 3–4 March 2008.
83. Slavin, R. Cherry Picking or Making Better Trees? Available online: https://robertslavinsblog.wordpress.com/2020/04/16/cherry-picking-or-making-better-trees/ (accessed on 16 April 2020).
84. Rosen, R. Shortage-Field Incentives: Impacts in Teacher Retention and Recruitment. Ph.D. Thesis, Columbia University, New York, NY, USA, 2012.
85. DeFeo, D.; Hirshberg, D.; Hill, A. It's More Than Just Dollars: Problematizing Salary As the Sole Mechanism for Recruiting and Retaining Teachers in Rural Alaska. In *Wellness and Healing: Indigenous Innovation and Alaska Native Research, Proceedings of the Alaska Native Studies Conference, Juneau, AK, USA, 13–15 April 2018*; University of Alaska Anchorage: Anchorage, AK, USA, 2018.
86. Fitzgerald, C.T. *Report on the High Priority Location Stipend Program*; Dade County Public Schools: Miami, FL, USA, 1986.
87. Glazerman, S.; Protik, A.; Teh, B.R.; Bruch, J.; Max, J. *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment*; NCEE 2014–4004; National Center for Education Evaluation and Regional Assistance: Washington, DC, USA, 2013.
88. Hough, H.J.; Loeb, S. *Can a District-Level Teacher Salary Incentive Policy Improve Teacher Recruitment and Retention?* Policy Brief 13-4; Policy Analysis for California Education, PACE: Stanford, CA, USA, 2013.
89. Steele, J.L.; Murnane, R.J.; Willett, J.B. Do Financial Incentives Help Low-Performing Schools Attract and Keep Academically Talented Teachers? *Evidence from California. J. Policy Anal. Manag.* **2010**, *29*, 451–478. [CrossRef]
90. Dolan, P.; Metcalfe, R.; Navarro-Martinez, D. *Financial Incentives and Working in the Education Sector*; Research Report DFE–RR251; Department for Education: London, UK, 2012.
91. Falch, T. Wages and recruitment: Evidence from external wage changes. *ILR Rev.* **2017**, *70*, 483–518. [CrossRef]
92. Zarkin, G.A. *The Importance of Economic Incentives in the Recruitment of Teachers. Final Report*; Duke University: Durham, NC, USA, 1985.
93. Fulbeck, E.S.; Richards, M.P. The Impact of School-Based Financial Incentives on Teachers' Strategic Moves: A Descriptive Analysis. *Teach. Coll. Rec.* **2015**, *117*, 9.
94. Bueno, C.; Sass, T. The Effects of Differential Pay on Teacher Recruitment, Retention and Quality. Calder Research Conference: 3 February 2017. Available online: https://edre.uark.edu/lecture-series/effectsofdifferentialpaysass.pdf (accessed on 20 September 2020).
95. Gorard, S.; Ventista, O.; Morris, R.; See, B.H. *Who Wants to Be a Teacher? Findings from a Survey of Undergraduates in England, Working Paper*; Durham University Evidence Centre for Education: Durham, UK, 2020.
96. Boyd, D.; Grossman, P.; Hammerness, K.; Lankford, H.; Loeb, S.; Ronfeldt, M.; Wyckoff, J. Recruiting effective math teachers: Evidence from New York City. *Am. Educ. Res. J.* **2012**, *49*, 1008–1047. [CrossRef]
97. Kraft, M.A.; Brunner, E.J.; Dougherty, S.M.; Schwegman, D.J. Teacher accountability reforms and the supply and quality of new teachers. *J. Public Econ.* **2020**, *188*, 104212. [CrossRef]
98. Shifrer, D.; Turley, R.L.; Heard, H. Do teacher financial awards improve teacher retention and student achievement in an urban disadvantaged school district? *Am. Educ. Res. J.* **2017**, *54*, 1117–1153. [CrossRef]
99. Springer, M.G.; Swain, W.A.; Rodriguez, L.A. Effective teacher retention bonuses: Evidence from Tennessee. *Educ. Eval. Policy Anal.* **2016**, *38*, 199–221. [CrossRef]

100. Clotfelter, C.; Glennie, E.; Ladd, H.; Vigdor, J. Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *J. Public Econ.* **2007**, *92*, 1352–1370. [CrossRef]
101. Clotfelter, C.T.; Glennie, E.J.; Ladd, H.F.; Vigdor, J.L. Teacher bonuses and teacher retention in low-performing schools: Evidence from the North Carolina $1800 teacher bonus program. *Public Financ. Rev.* **2008**, *36*, 63–87. [CrossRef]
102. Fryer, R.G. Teacher incentives and student achievement: Evidence from New York City public schools. *J. Labor Econ.* **2013**, *31*, 373–407. [CrossRef]
103. Falch, T. Teacher mobility responses to wage changes: Evidence from a quasi-natural experiment. *Am. Econ. Rev.* **2011**, *101*, 460–465. [CrossRef]
104. Feng, L.; Sass, T.R. *The Impact of Incentives to Recruit and Retain Teachers in "Hard-to-Staff" Subjects: An Analysis of the Florida Critical Teacher Shortage Program*; Working Paper 141; National Center for Analysis of Longitudinal Data in Education Research (CALDER): Washington, DC, USA, 2015.
105. Feng, L.; Sass, T.R. The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects. *J. Policy Anal. Manag.* **2018**, *37*, 112–135. [CrossRef]
106. Koedel, C.; Xiang, P.B. Pension enhancements and the retention of public employees. *ILR Rev.* **2017**, *70*, 519–551. [CrossRef]
107. Murnane, R.J.; Olsen, R.J. The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *J. Human Resour.* **1990**, *25*, 106–124. [CrossRef]
108. Springer, M.G.; Lewis, J.; Ehlert, M.; Podgursky, M.; Crader, G.; Taylor, L.; Gronberg, T.; Jansen, D.; Lopez, O.; Stuit, D. *District Awards for Teacher Excellence (D.A.T.E.) Program: Final Evaluation Report*; NCPI Policy Evaluation Report; National Center on Performance Incentives: Nashville, TN, USA, 2010.
109. Springer, M.G.; Taylor, L.L. Designing incentives for public school teachers: Evidence from a Texas incentive pay program. *J. Educ. Financ.* **2016**, *41*, 344–381. [CrossRef]
110. Booker, K.; Glazerman, S. *Effects of the Missouri Career Ladder Program on Teacher Mobility*; Mathematica Policy Research: Princeton, NJ, USA, 2009.
111. Fulbeck, E.S. Teacher Retention: Estimating and Understanding the Effects of Financial Incentives in Denver. Ph.D. Thesis, University of Colarado, Boulder, CO, USA, 2011.
112. Fulbeck, E.S. Teacher mobility and financial incentives: A descriptive analysis of Denver's ProComp. *Educ. Eval. Policy Anal.* **2014**, *36*, 67–82. [CrossRef]
113. Choi, W.S. The Effect of Alternative Compensation Programs on Teacher Retention and Student Achievement: The Case of Q Comp in Minnesota. Ph.D. Thesis, University of Minnesota, Minneapolis, MN, USA, 2015.
114. Dee, T.; Wyckoff, J. Incentives, selection, and teacher performance: Evidence from IMPACT. *J. Policy Anal. Manag.* **2015**, *34*, 267–297. [CrossRef]
115. Hendricks, M.D. Does it pay to pay teachers more? Evidence from Texas. *J. Public Econ.* **2014**, *109*, 50–63. [CrossRef]
116. Helms-Lorenz, M.; van de Grift, W.; Maulana, R. Longitudinal effects of induction on teaching skills and attrition rates of beginning teachers. *Sch. Eff. Sch. Improv.* **2016**, *27*, 178–204. [CrossRef]
117. Cohen, B.A. Enhancing the 'Learning Profession': Improving New Teacher Retention with Teacher Induction. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2005.
118. DeAngelis, K.J.; Wall, A.F.; Che, J. The impact of preservice preparation and early career support on novice teachers' career intentions and decisions. *J. Teach. Educ.* **2013**, *64*, 338–355. [CrossRef]
119. De Jong, D.; Campoli, A. Curricular coaches' impact on retention for early-career elementary teachers in the USA. *Int. J. Mentor. Coach. Educ.* **2018**, *7*, 191–200. [CrossRef]
120. Ingersoll, R.; Smith, T. Do teacher induction and mentoring matter? *NASSP Bull.* **2004**, *88*, 28–40. [CrossRef]
121. Latham, N.I.; Vogt, W.P. Do professional development schools reduce teacher attrition? Evidence from a longitudinal study of 1000 graduates. *J. Teach. Educ.* **2007**, *58*, 153–167. [CrossRef]
122. Papay, J.P.; West, M.R.; Fullerton, J.B.; Kane, T.J. Does an urban teacher residency increase student achievement? Early evidence from Boston. *Educ. Eval. Policy Anal.* **2012**, *34*, 413–434. [CrossRef]
123. Ronfeldt, M.; McQueen, K. Does New Teacher Induction Really Improve Retention? *J. Teach. Educ.* **2017**, *68*, 394–410. [CrossRef]
124. Speidel, M. Teacher Attrition and Retention in Exceptional Student Education: An Evaluation of the Skills, Tips, and Routines for Teacher Success (STARTS) Initiative of Volusia County, Florida Schools. Ph.D. Thesis, University of Central Florida, Orlando, FL, USA, 2005.

125. Weisbender, L. Preventing Teacher Dropout: Volume 1—Mentors Helping New LAUSD Teachers, 1984–1988; Volume 2—Retention Rates for LAUSD Mentors, Mentees, and Other Teachers, l984 to l988. 1989. Available online: https://eric.ed.gov/?id=ED332975 (accessed on 20 September 2020).
126. Silva, T.; McKie, A.; Knechtel, V.; Gleason, P.; Makowsky, L. *Teaching Residency Programs: A Multisite Look at a New Model to Prepare Teachers for High-Need Schools*; NCEE 2015-4002; National Center for Education Evaluation and Regional Assistance: Washington, DC, USA, 2015.
127. Ryan, S.V.; von der Embse, N.P.; Pendergast, L.L.; Saeki, E.; Segool, N.; Schwing, S. Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teach. Teach. Educ.* **2017**, *66*, 1–11. [CrossRef]
128. Carver-Thomas, D.; Darling-Hammond, L. *Teacher Turnover: Why It Matters and What We Can Do about It*; Learning Policy Institute: Palo Alto, CA, USA, 2017.
129. Fuchsman, D.; Sass, T.R.; Zamarro, G. *Testing, Teacher Turnover and the Distribution of Teachers across Grades and Schools*; EdWorking Paper No. 20-2020; Annenberg, Brown University: Providence, RI, USA, 2020.
130. Shirrell, M.A. School Working Conditions and Teacher Attrition: The Roles of Policy, Teacher Preparation, and School Principals. Ph.D. Thesis, Northwestern University, Evanston, IL, USA, 2014.
131. Jones, M.D. Teacher behavior under performance pay incentives. *Econ. Educ. Rev.* **2013**, *37*, 148–164. [CrossRef]
132. DfE. Teachers Set for Biggest Pay Rise in Fifteen Years. 2020. Available online: https://www.gov.uk/government/news/teachers-set-for-biggest-pay-rise-in-fifteen-years (accessed on 23 July 2020).
133. Weale, S. Teachers to Get Pay Rise in Attempt to Boost Recruitment. 2020. Available online: https://www.theguardian.com/education/2020/jan/21/teachers-to-get-pay-rise-in-attempt-to-boost-recruitment (accessed on 22 January 2020).
134. Gneezy, U.; Meier, S.; Rey-Biel, P. When and why incentives (don't) work to modify behaviour. *J. Econ. Perspect.* **2011**, *25*, 191–210. [CrossRef]

education sciences

MDPI

Article

Teacher Mobility in Punjab, Pakistan: Stayers and Movers within the Public and Private Schools

Nadia Siddiqui [1,*] and Sadia Shaukat [2]

[1] Durham Evidence Centre for Education, School of Education, Durham University, Durham DH1 1TA, UK
[2] Faculty of Education, Bank Road Campus, University of Education, Lahore 54770, Punjab, Pakistan; sadia.shaukat@ue.edu.pk
* Correspondence: nadia.siddiqui@durham.ac.uk

Abstract: Understanding the determinants of teacher mobility is important in order to implement effective policies for the recruitment, retention, and fair allocation of teachers. The teacher transfer policy implemented in Punjab, Pakistan, is intended to address corruption and a fair allocation of teachers in public schools. However, the policy has implications for teacher mobility. This paper presents survey findings on teacher mobility patterns in public schools in Punjab, Pakistan, examining the extent and determinants of mobility in comparison with teachers in private schools. In the survey, 1002 in-service teachers participated, and 46% reported changing school at least once during their teaching career. The findings show that teachers who changed schools in their early career, with an average of two years of teaching experience, gained higher salary benefits by changing schools compared to experienced teachers, with an average of 14 years of teaching experience, who never changed schools. In comparison with early career teachers, experienced teachers who never changed school had lower salaries but higher satisfaction with life in general and with the school as their workplace. The most common reasons for changing school were lack of teaching resources, difficulty in commuting to school, unmanageable student–teacher ratio, and no chance of promotion in their teaching career. Teacher mobility was slightly higher in public schools compared to private schools, despite implementation of a merit-based transfer policy. A binary logistic regression model was constructed with the outcome of teacher mobility (or not), with a base figure of 54%. By adding gender, marital status, school type, length of teaching experience, and teachers' satisfaction, the model increased correct predictions to 62%. Teacher salary and teaching workload did not explain any variation in the model. These findings have implications for teacher transfer policy in public schools and lessons for private schools to retain teachers by offering longer contracts and reliable pension schemes. Policymakers must consider facilitating teachers' satisfaction with their workplace, particularly by making public schools in rural areas attractive places for the retention of teachers in early career phases.

Keywords: transfer policy; rationalisation; mobility; retention

Citation: Siddiqui, N.; Shaukat, S. Teacher Mobility in Punjab, Pakistan: Stayers and Movers within the Public and Private Schools. *Educ. Sci.* **2021**, *11*, 358. https://doi.org/10.3390/educsci11070358

Academic Editors: Beng Huat See, Rebecca Morris and Eila Jeronen

Received: 30 April 2021
Accepted: 13 July 2021
Published: 16 July 2021

1. Introduction

Teachers play an essential role in educational change and development [1]. In order to achieve children's universal access to primary and secondary education, all public schools are expected to have fair deployment and allocation of teachers. Governments cannot achieve the target of education for all without having teachers for all. The role of the teacher as a mediator of social and economic transformation has never been more evident than in the current times [2]. However, implications of meaningful teaching are increasing quickly in the contemporary world, and the criteria for education and learning are higher than ever before [3]. A standard of education is crucial for any country, and teacher education plays a significant role in regulating it. Since instruction is provided mostly by teachers, the size, alignment, and distribution of the teaching work force are significant factors for

the effectiveness of public schools. Hence, public education is mainly the government's responsibility and is considered essential to the country's social and economic well-being and as well as its global competitiveness. Therefore, the appropriateness of the supply of teachers in relation to the demand and retention has been the substance of ongoing concern among teacher educators and policy makers accountable for certifying the efficiency of public schools [4].

This is a cross-sectional survey based on teacher mobility to understand patterns of teachers staying in or changing schools during their teaching career. All in-service teachers working in public and private schools in Punjab, Pakistan, were invited to complete a survey on job experience, their teaching career, and job mobility during their careers. In Punjab, public schools have a provincially centralised and merit-based transfer policy, while private and public–private partnership schools do not have a regular teacher recruitment system or transfer policy. The determinants of school job mobility might be different across the two sectors. Salary, nature of job contracts, transfer policies, student–teacher ratio, workload, and satisfaction with work environment can contribute to teachers' decisions of changing schools which in turn has implications on teacher retention, teacher shortage in disadvantaged regions, and compositional disparities in teaching quality.

This survey study shows teacher mobility is slightly higher in public schools compared with private schools, despite after a successful implementation of rationalization and merit-based transfer policies. The rationalization policy introduced in Punjab was intended to make teacher deployment and transfer free from corruption, improve teacher retention, and equalise teacher composition in urban and rural schools. The policy has tackled corruption to some extent, but the implementation has not successfully met the targets of fair allocation of teachers. Teachers have shown dissatisfaction with their job placements, and this is associated with teacher mobility within the public schools. Schools in socioeconomically disadvantaged areas particularly located in remote and rural locations have vacant positions, which is a sign of teacher shortage. However, schools in central urban locations have surplus teachers.

A social media campaign by a large number of public school teachers protested against the impact of transfers on their lives. Female teachers protested against transfer orders and deployment to hard-to-reach areas, which had practical implications on travelling to schools in distant locations, as it is financially unaffordable, unsafe for travelling alone, and time consuming. It is perhaps the implementation of rationalisation and transfer policy that has made teacher mobility more volatile in public schools compared with private schools. The findings of this survey study show that higher salaries and time allocated for workload are not related to the teachers' decision to change schools; instead, an important factor that retains teachers is their satisfaction with work environment. The findings of this survey have to be treated with caution due to the voluntary and unrepresentative sample, but they provide an insight into the possible association between teacher retention and working environment.

2. Background

The school teacher workforce in Pakistan is divided into public and private sectors. According to Pakistan Education Statistics (2018–2019), public schools have more student proportion (29 million) as compared to the private schools (21 million) [5]. Teacher workforce is slightly higher in private schools (51%) when compared with teachers in public schools (49%). Although public schools have a larger student population when compared with private schools, the public sector has a shortage of teachers. Teacher training or licenses for working in a school are not compulsory requirements in public or private schools. Therefore, anyone can apply for a school teaching job. There could be several reasons for the imbalance in teacher workforce distribution in the systems, which run parallel but independent of each other [6]. Fixed-term and permanent contracts, salary, pensions, and privileges are very different across the two systems [7]. Each of the four provincial governments have independent but centralised merit-based systems for teacher

recruitment, deployment, promotion, and job transfer. Private schools exist in all four of the provinces and are independent in recruiting teachers and implementing policies to meet their needs of teaching workforce [8].

In this paper, we are presenting the case of Punjab in Pakistan, where in the last ten years several public school education policies and initiatives have been implemented to address corruption in education and fairly manage teacher workforce.

In 2016, the School Education Department (SED) in Punjab introduced standardised test-based performance measures for selecting public school teachers on permanent positions. In addition to setting standards of selection in the teacher recruitment and deployment process, the provincial government implemented new policies for a fair distribution of teachers in public schools. There were also concerns regarding teacher shortages in rural and geographically remote areas, exasperated by the corruption in the process of teacher transfers in public schools. Teacher transfers were managed through manual applications. Bribery, nepotism, and contacts in the public offices made teacher deployment and transfer less transparent, more unfair, and a longer process [9,10].

In 2013, the education minister in Punjab implemented teacher transfer policy on a point-based system [5]. The criteria currently in practice include: Distance (more than 8 km from home are eligible) (20 points), years of service (30 points), seniority (20 points) (teachers in first 3 years of service cannot apply), wedlock basis (15 points), and compassionate grounds (15 points).

Rationalisation measures were introduced in public school of Punjab in 2010 where surplus teachers in schools were re-allocated when there was a teacher shortage. The student–teacher ratio was introduced where 1 teacher for 40 students was an agreed ratio at primary school level. This is a high student–teacher ratio when compared to countries with similar education contexts such as Morocco, Tunisia, and Turkey [10]. In the policy documents, there is a lack of clarity in how rationalisation of teachers in public schools is conceptualised in a centrally managed system of teacher recruitment and deployment. Theoretically, there is no need to have a transfer policy if the system is centrally managed because demand for teachers is fulfilled by the teacher recruitment system when there is a need. However, the school systems are dynamic due to factors such as changes in student population, teachers completing service period, changes in circumstances of teachers' life, schools merging or closing, and most importantly teachers as permanent public employees have protected right to work, salary, and pensions [10]. Closure of any public school means that the teachers during their service period need to be reallocated if they become redundant or a surplus. However, there is no promising evidence on existing policies which has fairly and equitably managed teacher deployment or transfer.

Teachers in the public schools can now apply for transfers to another school of their choice within a district. All transfer applications are managed through E-transfer process which is part of the digitalisation initiative introduced in 2018. An internet-based application system has made the process accessible and transparent. Depending on the points gained on the given set of criteria, the transfer application is accepted or rejected. Nearly 100,000 teachers applied for the transfer, and 50,000 were accepted [7]. Applications for public school transfers are accepted if their choice of school meets the criteria of rationalization. There are less chances of corruption involved in the process. However, this has not addressed teacher shortages in disadvantaged areas.

Public school teachers can apply for promotions when there is a vacancy available that offers an upscale in teaching position. This is a transfer by promotion to other school in the same district. These transfers are based on seniority level in the period of service, which means if more than one teacher applies for a vacancy, then the senior applicant would have more chances of acceptance. Transfer by promotion favours experienced teachers. SED, Punjab, Pakistan, has maintained administrative data of teacher transfer, transfer decisions, and promotions can show patterns, the extent, and direction of mobility in rural and urban regions. The existing international evidence based on administrative datasets shows that teacher mobility by seniority level adversely affects the quality of teaching in schools and

mainly impact disadvantaged students' learning experience who are concentrated in the regions of high socioeconomic poverty [8].

In a centrally managed system, public school teachers have a condition-based choice to change schools, whereas private school teachers can leave and join any private school where there is a vacant position subject to acceptance of the application by the headteacher or school management system. These two systems attract a large number of teacher workforce, but differences in policies have implications on who join and retain in teaching career, leading to subsequent outcomes such as teachers' satisfaction with their work environment with their salary [9,10] and passion for teaching profession [11]. Teachers make choice to leave, stay or move within public and private system. There is no evidence of teacher exchange or crossover in the two sectors, and perhaps, this is related with differences in policies for teacher recruitment, deployment, salary structure, school environment, and pension schemes [12].

In the public sector, higher salaries and permanent job positions show higher rates of teacher's attendance in comparison with teachers on lower salaries and shorter job contracts [13]. It is well known that teachers are less likely to remain on a long-term basis in private schools. This phenomenon of hiring teachers on short-term contracts is well described by the analogy of 'schools as sausage machines' where teachers work on short-term contracts and are hired and fired on demand basis [14]. In context of Punjab, Pakistan, administrative large-scale data are required to understand differences in the teacher workforce across the public and private schools. The determinants of teacher mobility within the systems are unclear, and there is no sufficient evidence if the government policies such as rationalisation and teacher transfer are appropriately targeting the underlying issues of teacher job mobility in Pakistan. This study answers the following questions:

- Are teachers who stay in school different from teachers who change schools?
- What factors determine teachers' job mobility?

SED, Punjab has made several improvements in digitalisation of school and student data introducing new regulations for public school teachers so that they can apply for transfer under the new merit-based policy. Teachers can also challenge their transfer orders using the same system. The SED has large amounts of data on teacher supply, deployment, transfer, and promotions, but access to this valuable information is not available for independent research use. Unfortunately, we were not successful in our request to the department in achieving these data, which could be the most appropriate for answering research questions on teacher school mobility, at least in the public school sector. We launched an online survey to collect self-reported information on teachers' job satisfaction, workload, and school mobility during their teaching career. The online teacher survey included in-service teachers from private and public school sectors.

We investigated mobility of teachers within public and private school system in schools of Punjab. All in-service teachers were invited to complete an online survey. Punjab public schools have a centralised and merit-based transfer policy, while private and public–private partnership schools do not have a regular teacher recruitment and teacher transfer policy. Considering the differences, the hypothesis in this research was that the teacher job mobility would be less prevalent in public schools when compared with teachers in private schools. Salary, student–teacher ratio, workload, and satisfaction with work environment contribute to teachers' decisions to change schools. The existing evidence based on non-random selection of teacher's interviews has shown that teachers on short-term contracts in private schools are more mobile than teachers on permanent contracts in public schools [15].

This study presents survey findings on teacher workforce differences between two different school systems and analyse the factors that determine teacher school mobility (or lack of it) in their teaching career.

3. Existing Evidence on Teacher Mobility

Teacher supply and mobility have important links to the effective delivery of education. There is international research evidence on effective school mechanisms which shows teachers' stability, consistent professional development, and satisfaction with school play important role in efficient working of school [16]. There is no clear evidence regarding teacher supply mechanism and determinants of teacher mobility in context of Pakistan. This is a complex school education sector in which teacher supply/demand and mobility could have multiple determinants and can lead to various school level outcomes. The knowledge of the factors that drive teachers' recruitment, supply to schools, and mobility during different phases of teaching career can inform effective policy measures in which interest of various stakeholders can be protected.

There is a concern over the influence of teacher job change on both the inclusive level of teacher quality and the supply of teacher quality across schools [17]. Teachers leaving the profession or switching schools located in economically developed regions lead to compositional disparities in terms of teacher quality [18,19]. This has implications on designing feasible policies that can support sustainable school systems, where teacher retention in their prime years of teaching career and experience can benefit student accomplishment [20]. There is an indication that within-profession teacher mobility is contributing some variation in terms of teacher's credentials [21].

Teachers working in schools with a higher proportion of disadvantaged students are more likely to transfer to a new school district [22], and teachers in municipal inner-city schools are more likely to transfer away from their schools than teachers in other areas [23]. Another factor is the efficiency of a teacher relative to their peer teachers may impact teacher job choice, though the direction of the effect is uncertain. Teachers tend to seek out positions where their efficiency matches with the efficiency of their peers. In addition, productive teachers tend to choose those schools where their work compatibility and intellectual interest align with their peer teachers to enhance their intellectual growth and professional development [24]. On the other side, there is some evidence that ambitious and qualified teachers might choose to transfer to schools where the work environment does not demand work efficiency and high-performance goals to achieve [25].

Through the existing period of educational reform beginning in the early 1980s, the teaching force has been extensively considered, on the whole, as insufficient to encounter current national supplies [26,27]. Meaningful teaching delivered effectively by trained and highly competent teachers have a high demand in all education systems [28]. Therefore, the equitable supply of teachers in relation to the demand and retention has been an important concern for policymakers who are accountable for the efficiency of public schools [29].

Many recent reports by blue ribbon commissions have determined that education as a whole, and teachers in actuality, are not producing government school students with the level of knowledge, skill, and discipline required in today's multifaceted and competitive economic environment [30]. Teachers who live farther from school tend to remain more absent; particularly, female teachers are more absent and are less likely to have access to safe transport facility. Teachers are expected to perform tasks that have no direct relevance with children's learning. The education department assign orders to teachers for completing nonteaching duties such as helping with antipolio and antidengue drives, election duties, and administering government exams and conducting surveys. A study conducted by Alif Ailaan and SAHE assessed that government teachers spend a quarter of the academic year on nonteaching activities that can contribute to weakening the performance of school students [31].

The government has introduced initiatives to monitor presence and work performance of the teachers in the classrooms, in response to a long-term problem of teacher absenteeism [32]. There have been initiatives implemented for the least number of improvements in public school quality such as addressing teacher absenteeism. All public school teachers now mark biometric attendance [33], and there are also uninformed inspection visits. These changes implemented in Punjab have made teachers to be compliant and present in

school during expected job hours. However, teacher unions have raised their voice about the time for teaching or preparation for teaching have been increasingly replaced with teachers' involvement in administrative and managerial tasks. Teaching workload and administrative duties such as participation children's enrolment drives, campaigns for the eradication of polio, dengue, and Covid-19 has led to burnout [34]. Teachers in private schools have no involvement in government-led campaigns; therefore, their allocated time is entirely dedicated for teaching-related activities.

Another reason of teacher mobility is the geographical context and school characteristics. Geographical preferences are overall relatively personal, being related to ignored individual preferences; teachers have to take up whatever the location is presented to them and quite often commence their careers distant from their preferred physical location. This obviously affects the efficacy of teachers and of the education system as a whole. Discontented teachers may be poor performers, not only because of overall motivational factors [35] but also since they are merely waiting to move on to a diverse location, exerting low effort into their existing work duties and ignoring any extended term plans for their students [36]. Schools' characteristics such as size of classroom, infrastructure, and being under-resourced contribute to teacher mobility [37].

Teaching experience can be challenging in schools where students and parents depend on additional and constant support from school and teachers. Teachers try to avoid schools situated in a setting with a high poverty rate and a high percentage of people working in agriculture. Moreover, in schools with a higher portion of disabled students, drop-out students, and students failing to gain admission to the next grade make such schools less attractive to teachers [37].

Considering the aforementioned determinants of teacher mobility, Punjab SED initiated E-transfer policy to apply school transfer online and to avoid visiting the School Education Department physically. Government statistics reflecting main shortages in the quantity of teachers in public schools, it is generally perceived that the education system is failing as a result of deficiencies in both the quantity and quality of teachers [38].

Teachers in public schools are always at the good will of school education department and school administration. Transfer policy for rationalisation has been the reason of distress among teachers in public schools. A social media campaign against these new regulations led to a protest in Punjab [39,40].

There has been some evidence that a regular transfer of teachers generates not only conflicts for the teachers, their families, and children but also affect the quality of teaching and learning in the schools [41]. There are no accommodation facilities or additional transport allowance for teachers placed in remote areas. Lack of capacity building opportunities has also shown association with teachers' dissatisfaction with the teaching profession [42,43].

Teachers' decisions to change school or leave the teaching profession can be determined by workload stress and nature of tasks in the teaching duties. Existing studies have shown that workload-related stress and number of hours spent on teaching tasks, such as marking, are associated with teachers' declining quality of life [44].

4. Research Design and Methods for Analysis

This study is a survey-based cross-sectional research design in which respondents volunteered to take part and self-reported their information. An online survey was developed for in-service teachers' participation. The survey language was English because all in service can read, understand, and respond in this language. The online survey link was widely distributed in schools and social media forums for teachers' high response rate. The survey message invited teachers for volunteering participation and sharing information for the purpose of research. There was no sensitive information collected that could potentially harm or breach the confidentiality of the respondents. No rule was set on questions; therefore, respondents could skip giving any information which they did not like to share. Only one question required a compulsory response, which asked all the

respondents to report in 'yes' or 'no', if they had ever changed school during their teaching career. The survey completion time for each respondent was less than 10 min.

The survey included three main sections on general life and job satisfaction, teaching workload, and reasons for changing school. General life and job satisfaction and teaching workload sections were for all respondents. The third section on reasons for changing schools was only available for those respondents who reported that they changed school at least once in their teaching career. Items in all sections were statements to be rated on scale of 0–10 where 0 means 'not at all' and 10 means 'a lot'. The analysis is based on averages and standard deviations of the rated items, which are then used for calculating the effect size.

At the end of each section, there was an open dialogue box for respondents to give any additional information or explanation of responses they wanted to contribute. The item response-scale was consistent in all sections except for the background information. Teachers' self-reported information on age, marital status, work experience, type of school they worked, and take-home salary, which were open-response items or multiple answer choice.

This survey invited volunteering participants, and in order to retain a large sample with minimum dropout, we did not set a compulsory rule in question items except for one. This resulted in missing data, which we coded as missing. We did not replace missing responses with weighted values or used any methods for statistical imputations.

Simple descriptive and bivariate analysis demonstrated the difference between two groups of teachers who changed school and those who never changed school. For each item mean score and standard deviation were used to calculate the effect sizes, which measure the magnitude of the difference in the responses between the two groups.

To identify the determining factors or predictors in explaining the differences the two groups, a binary logistic regression analysis was performed. The binary outcome variable was 'changed' school or not, and other variables were introduced as predictors using a stepwise forward analysis. This means predictor variables, which were not contributing any explanation in the variation (percentage correctness) were removed.

The obtained sample from online survey naturally included 46% of teachers who changed school and 54% of those who never changed school. This is an appropriate balance in the overall distribution of the sample, thereby meeting the condition for binary logistics analysis [45]. This distribution means before adding any explanatory variable in the regression model, there is nearly 54% chance that any case would change a school. By adding predictor variables such as gender, age, work experience, satisfaction with job, workload, take-home salary, school type, etc. We anticipated an increase in the percentage correctness or likelihood of cases to be accurately predicted. The results of regression analysis showed major determinants of teacher mobility and those, which need more focus in policy for teacher recruitment, deployment, and retention.

5. Teachers Who Participated in the Study

This survey is opt-in participation of in-service teachers. This means teachers volunteered to take part in the survey in response to our invitation sent to nearly 680 schools in Punjab. We also used our social networks and personal links in approaching school education forums and groups inviting teachers to take part in the study. It is hard to estimate a response rate because the information was widely shared, but a large number of teachers volunteered to take part in the online survey. However, the online survey is restricted to only those who have access to internet technology to receive the information and complete the survey. The nonrandom sample of 1002 teachers does not represent teacher population of the time in Punjab, Pakistan. However, it is large enough in each category of characteristics to conduct a meaningful analysis. The sample is reasonably large for meaningful calculations of effect sizes and regression modelling (see more details in [45]). Table 1 shows characteristics of the achieved sample.

Table 1. Percentage of teachers.

	Percentage
School type	
Public	54
Private	29
Public-Private Partnership	16
Madrassah	1
Gender	
Male	35
Female	65
Geographical region	
Urban	51
Rural	47
Missing	2
Marital status	
Married	54
Single	43
Missing	3
Highest Qualification	
Completed diploma	3
Completed higher education (Graduate level and above)	88
Completed intermediate studies	6
Completed secondary	3
Missing	1
Take home salary self-reported	95
Teaching experience self-reported	97
Age in years self-reported	97
Total	1002

According to Table 1 the sample has a higher participation of female teachers from public schools in urban areas. The majority of teachers have completed higher-education studies. These are very similar with official figures on teacher workforce in Punjab.

Cases with missing information are retained as much as possible instead of excluding them from analysis. Teachers self-reported their background characteristics, salary, age, and work experience in years. These are important indicators to understand the patterns of school job mobility in teaching career.

Table 2 shows the gender pay gap in teaching workforce. On average, public school teachers earn higher salary than teachers in private and public–private partnership schools. This means public school teachers are advantaged over other teachers by having permanent and pensionable jobs with higher salaries. This is consistent with the evidence on public and private salary differences presented in study called Learning and Educational Achievement in Punjab Schools [43]. Teaching experience in years is also highest for teachers in public school, which could be due to working on permanent job contracts. Private schools offer higher salaries when they recruit teachers, but not many offer pension or benefits, therefore being less likely to retain teachers on a long-term basis.

Table 2. Average teacher salary and work experience.

	Take Home Salary (Average)	Work Experience in Years (Average)
Male	Rs 39,455	9
Female	Rs 27,472	8
Urban	Rs 34,129	8
Rural	Rs 29,191	8
Public	Rs 40,119	11
Private	Rs 22,034	5
Public–Private Partnership	Rs 19,050	6
Madrassah	Rs 15,000	5

6. Who Changed Schools in Their Teaching Career?

In this online survey, 1002 in-service teachers participated and self-reported their responses and background information. The descriptive analysis shows differences in teachers' background characteristics for those who changed schools and those who never changed school (See Table 3).

Table 3. Percentage who changed/never changed school.

	Changed School %	Never Changed School %
School type		
Public	50	50
Private	46	54
Public–private partnership	34	66
Madrassah	33	67
Gender		
Male	45	55
Female	47	53
Geographical region		
Urban	48	52
Rural	45	55
Marital status		
Married	54	47
Single	39	61
Highest Qualification		
Completed diploma	29	71
Completed higher education	47	53
Completed intermediate studies	39	61
Completed secondary	47	53
Take home salary (average)	Rs.31,197	Rs.24,988
Teaching experience in years (average)	2 years	14 years
Age in years (average)	34	55
Total	540	462

Teacher mobility in public schools is slightly higher as compared with private schools. Half of the public school teachers have changed school at least once in their teaching career despite permanent teaching positions, and job transfer is a competitive point-based merit system. There are some other slight differences such as teachers in urban areas changing schools more than teachers in rural areas. Urban areas have more schools and therefore more job opportunities as compared to rural areas. The direction of teacher transfer could be more likely from rural to urban location. Vacant teaching positions in rural schools might offer promotion opportunities, but teachers working in urban school locations would prefer not to transfer to the vacant positions in rural areas due to general lack of resources. More female teachers change schools than male teachers, which could be associated with change in marital status where female teachers are more likely to migrate.

Average salary and years of teaching experience show bigger gaps. Experienced teachers have served in a school for a long time and never changed schools, while teachers in early careers changed schools within two years. On average, teachers' retention in a school on long-term basis does not show association with salary because less experienced teachers who changed schools within average of two years in teaching careers are earning extra Rs 6208. A possible explanation of this difference could be that teachers stay in a school on long-term basis if they are satisfied with their job and school environment, while salary is not a compelling factor to change school. Teachers change school if they are not satisfied with their job and school environment, and joining a new school would mean negotiating and accepting only a higher salary than what they were receiving previously. Public and private salary structures could also be related with these gaps. Regression analysis shows if difference in the mobility patterns can be explained by teachers' satisfaction, workload, salary, and school type or not.

Table 4 shows differences in life and job satisfaction of teachers in public, private, and private–public partnership schools. The differences are reported as effect sizes interpreted as standardised gap between teachers who changed schools at least once in their teaching career and those who never changed school. The effect sizes are separately reported for the school types to show differences in the patterns of teacher mobility more clearly. Highlighted items show noticeable differences.

The overall effect sizes in Table 4 show that teachers who changed school are less satisfied with their life and work experience at school as the effect sizes are consistently negative. Work environment, work–life balance, and assistance in teaching activities have large negative effect sizes. These effect sizes are about their current job experiences in schools where they moved to and were currently working when they completed the survey. It is surprising to see that changing schools does not make any positive difference to their life and job satisfaction. Those who never changed schools are more satisfied with life and work environment in their current school. This is an indication that teacher retention is higher when the work environment has a sufficient amount of support available for teachers and they have a good work–life balance. Public schools have the option of promotion by transferring to other schools where there are vacant positions. However, it might be that public school teachers do not want promotion by transferring to schools where they have to compromise on work environment. These are likely to be schools in rural areas where limited resources can make teachers feel isolated from developed regions and losing development opportunities available in urban regions.

The effect sizes were calculated for the public and private schools, and there is no big difference in these two school types except large negative gaps in two categories of satisfaction: life in general and availability of time in personal life. Public school teachers who changed schools were less satisfied in life in general and have less availability of time in personal life than their counterpart in private schools. This is perhaps associated with rationalisation of teachers when surplus teachers in a school are transferred to schools against their choice. Teacher shortage in rural areas can also increase the workload for teachers in rural regions.

Table 4. Effect size/Life satisfaction/Changed school/Never changed schools.

Life Satisfaction	All	Public	Private and Public–Private Partnership
Generally, I am satisfied with my life	−0.03	−0.15	−0.01
I feel happy most of the time	−0.07	−0.05	−0.01
I have plenty of time to do the things I want in personal life	−0.18	−0.12	0.00
I get excited about teaching	−0.03	−0.10	0.02
I enjoy working with students	0.06	−0.08	0.02
I like the work environment of my school	−0.17	0.03	−0.03
It is convenient to commute to school regularly	−0.15	−0.08	−0.01
I am satisfied with the student−teacher ratio in my school	−0.13	−0.01	0.00
I enjoy working with my colleagues	−0.08	−0.07	−0.01
I am satisfied with my current salary package	−0.05	−0.06	−0.03
I achieve a good balance between my work life and my private life	−0.21	−0.05	−0.06
I can complete my assigned teaching workload during contracted working hours	−0.15	−0.09	−0.01
I work on my teaching duties during weekends and holidays	−0.14	−0.12	0.05
The school has provided me teaching assistance (e.g., helper staff, computer/laptop, etc.) to support my teaching activities	−0.21	−0.08	−0.02
I can quit teaching if I get some other source of income	−0.05	−0.10	0.00
I want to work in another school	0.10	−0.07	0.04
Total	1002	544	449

Time allocation for work-related activities can contribute to overall work experience and satisfaction with teaching job. Table 5 shows the effect sizes of mobile teacher time allocation in workload for activities that can support their teaching. The negative effect sizes show that teacher who changed schools have shown less amount of time allocated to their teaching activities. The items are highlighted where effect sizes are large for consideration. In these highlighted items, there is an underlying need for support and professional growth, which is neglected in the workload model. Time for mentoring, peer feedback, and training for development are not available in workload, which can make overall teaching experience less supported with appropriate measures.

Table 5 on workload time allocation, there are slight but consistent negative effect sizes for those who changed schools within public and private system. Although teachers in public schools have reported having less time to do things in their personal lives, the effect sizes in Table 5 show they have more time allocated in workload for teaching activities and teaching support when compared with teachers in public school. Mobile teachers in private school work slightly more than their allocated time, but these effect sizes are very small to make any definitive conclusions.

Table 5. Effect size/Time allocated for the workload/Changed school/Never changed schools.

Workload	Effect Size All	Public	Private and Public–Private Partnership
Marking students' work	−0.16	0.07	−0.11
Individual student time for counselling	−0.15	0.00	−0.08
Preparation for lessons	−0.13	0.03	−0.08
Mentoring other teachers/colleagues	−0.10	0.05	−0.06
Breaks for lunch/ short period of rest	−0.14	0.03	−0.05
Observe other teachers for feedback and/or learning practice	−0.19	0.05	−0.09
Being mentored by teachers/colleagues	−0.18	0.03	−0.05
Parent teacher meeting	−0.13	−0.02	−0.05
Professional development training	−0.18	0.07	−0.09
Participation in the civic campaigns such as information on Dengue, Corona, etc,	−0.07	0.05	−0.05
To be involved in the National Day, Kashmir Day, Election officer and different rallies	−0.02	0.03	−0.08
To be involved in board exam duty as invigilator or exam marker	−0.04	0.03	−0.09
Total	1002	544	449

7. Determinants of Teacher Mobility

Binary logistic regression model was constructed with an outcome variable of changed school or not. The obtained sample from online survey naturally included 46% of teachers who changed school and 54% of those who never changed school. This is an appropriate balance in the overall distribution of the sample and therefore meets the condition for binary logistics analysis [44]. This distribution means before adding any explanatory variable in the regression model, there is nearly 54% chance that any case would change a school. By adding predictor variables such as gender, age, work experience, satisfaction with job, workload, take-home salary, school type, etc., we anticipated an increase in the percentage correctness or likelihood of cases to be accurately predicted.

Table 6 shows that after introducing teacher background details, the percentage correctness of the model increased by 8% points. This means that there is still a high percentage of 38% which remains unexplained. It could be due to limitation in the sample and data quality. However, the known predictors have shown some interesting patterns. The constructed model individually excluded the variables to observe changes in the percentage correctness. Upon exclusion of age, take-home salary, and time for workload did not change the percentage correctness. This could mean that teacher mobility is not determined by these variables.

Table 6. Binary logistic model changed (Changed school/Never changed schools).

			Predicted		
			Have you changed school(s) during your teaching career?		Percentage Correct
			No	Yes	
Step 0	Have you changed school(s) during your teaching career?	No	540	0	100.0
		Yes	462	0	0.0
	Overall Percentage				53.9
Step 1	School type, sex, region, marital status, * age, work experience, * Take-home salary, satisfaction with life and work, * time for workload				62

* Do not change percentage correctness if excluded from the model.

Table 7 shows coefficient in the odds ratio of the predictors added in the model. This represents the likelihood of teacher mobility by controlling the predictors.

Table 7. Binary logistic model predicting outcome Changed school/Never changed schools.

Predictors	Odds Ratio
Are you? Female vs. Male	1.30
Geographical region Rural vs. Urban	0.97
What is your marital status? Married vs. Single	1.44
Work experience in years	1.04
Take-home salary	1.00
School Types vs. Public	
School Type (Madrassah)	1.15
School Type (Private)	0.80
School Type (Public–Private Partnership)	1.43
Satisfaction average	0.98
Time for workload	1.07

The coefficients are in the odds ratio, meaning that, all other things taken into account so far, a female respondent in the category of sex is 1.30 times more likely to change the school than male counterpart. Odds of 1 would mean likely to change schools. The negative odds are ≤ 0 and are interpreted in percentage. Teachers in private schools are less likely to change schools. The odds ratio of 0.80 means that private school teachers would likely to be in 20% chance of changing school in comparison with public school teachers. Teachers in urban areas are more mobile than teachers in rural areas.

Age and teaching experience made no difference in the percentage correctness and were excluded from the mode; therefore, the coefficients for only work experience are included. Teachers in early career stages (experienced less than 2 years) are more mobile than older teachers, but mobility seems less explained by experience. Teacher satisfaction and time for workload are average scores rather than categorical values. Teachers less satisfied with work environment are more likely to change schools than those who reported higher satisfaction. Teachers who have less time allocation in workload are more likely to change schools.

8. Why Change Schools?

Teachers who reported changing jobs were asked to rate separate set of items on the scale of 0–10 indicating the reasons of changing schools. This section included 34% of the teacher sample.

Table 8 show effect size differences for public school teachers in comparison with private school teachers.

Table 8. Reasons for job mobility on scale of 0–10. Public vs. Private and Public–Private partnership.

	Public	Private and Public–Private Partnership	Effect Size
I never wanted to change my teaching job to another school	4.58	4.50	0.02
I was transferred by the education authorities	4.88	3.41	0.35
I was offered a better salary in another school	2.30	5.83	−0.86
In the previous school my workload was more than my monthly salary	3.13	6.32	−0.78
In the previous schools, working hour policy was not feasible	2.56	5.34	−0.71
My job contract was not renewed by the previous school	2.00	4.06	−0.53
I was offered a permanent/longer contract by another school	2.51	4.40	−0.46
It was difficult to commute to previous school regularly	3.52	4.54	−0.25
The new school offered a transport facility	2.09	3.42	−0.34
The new school offered reduced fees for my child/children	1.60	3.41	−0.50
The new school had a manageable student–teacher ratio	4.20	5.77	−0.39
In the previous school I was not teaching at the right level of school phase according to my qualification and/or experience	3.53	4.57	−0.26
My spouse/family moved to another address or city	3.58	3.42	0.04
The previous school offered no chances of job promotion	2.89	4.99	−0.50
I fell out with the school leadership and management bodies	2.44	3.99	−0.41
The previous school was in an area of high poverty	3.67	3.36	0.08
The previous school provided very limited teaching resources	3.53	4.72	−0.31
The previous school was closed by the management authorities	1.52	3.02	−0.43
The previous school environment was not intellectually stimulating	2.69	4.18	−0.40
The previous school did not nominate me for a professional development opportunity	2.88	4.53	−0.42

In Table 8, highlighted average scores and effect sizes show noteworthy differences among public and private school teachers in their reasons to change school. Teachers in private schools changed schools due to workload and dissatisfaction with salary. However, teachers in public schools reported a higher average in reasons of changing school due to difficulties in commuting to school and location of school in the regions of high poverty.

Change in family status and transfer by the orders of the education department are the main reasons for public school teachers to have changed school. Intellectually stimulating work environment and lack of professional development opportunities are important reasons for changing schools for private school teachers. Private school teachers offered longer contracts of teaching service, which can be an attractive incentive for teacher retention.

9. Conclusions

The sample achieved in this study suggests that teacher mobility is slightly higher in public schools than in private schools in Pakistan. This follows successful implementation of a rationalisation policy to improve teacher retention and fair teacher composition in urban and rural schools. It is perhaps the implementation of rationalisation and transfer policy which has made teacher mobility volatile in public schools. Teachers who changed school in their early career, with an average of two years teaching experience, gained higher salary benefits by changing schools compared to teachers with an average of 14 years of teaching experience, who never changed schools. It is noteworthy that the proportion of private school teachers in the sample (46%) is lower than public school teachers (51%). In other words, more teachers from the public sector responded to the survey than from the private sector. In comparison with early career teachers, experienced teachers who never changed school had lower salaries but higher satisfaction with life in general and with school as their workplace. Teachers currently in urban areas tend to be more mobile than teachers in rural areas which might suggest that the direction of mobility is toward schools in urban locations. If so, transfer policy needs to introduce measures where rural to urban teacher mobility can be fairly addressed.

Overall, teacher salary does not seem to be the determining factor for teacher retention. Satisfaction with work environment seems to play an important role in teacher retention. Satisfaction with work environment and school conditions can make a teaching job more attractive and is likely to increase teacher retention in school and in profession. These are areas of investment for strengthening the infrastructure of public schools to make them more attractive for teachers to join the profession and retain on long-term basis. Working conditions in schools need more development initiatives where public school intake of children is a mix of pupils from different socioeconomic backgrounds. Public schools in the regions of high poverty and rural areas need additional resources and support provision to overcome the needs of disadvantaged students in schools. Another area of policy and provisions could be to tackle teachers feeling isolated from developing regions and opportunities. Perhaps, schools connected through digital technologies can make a difference in this regard. Better transport facilities, road links, and safe travel routes have been repeatedly suggested for improvement in other areas of development such as increasing school enrolment and attendance, especially for girls. These improvement for development in infrastructure can also improve teacher mobility patterns toward harder areas making schools in rural areas attractive job places.

Implications of these findings are also on public education policies of rationalization and teacher transfer where the intention to overcome corruption and compositional disparities in teacher workforce are likely to impact teachers' lack of satisfaction with their job. The lesson for private schools is that longer and sustainable teaching contracts can be an attractive incentive for teacher retention and long-term commitment to a school. If private schools can offer some reliable pension schemes, then teachers' retention in a school can possibly be for a longer time.

The logistic regression model could only explain 62% of the differences between those who changed school and those who did not. This suggests that there are other factors that are not accounted for that could explain these differences. A bigger and better database is needed for more conclusive evidence.

It is useful for researchers that the Punjab government has successfully digitised data on public school teachers. This information is important for the purpose of fair decisions on teacher recruitment, deployment, and transfer. We hope that the data systems can include

information on private school teachers for observing disparities and finding potential for workforce cooperation in the two parallel systems. Administrative data have great value for research informed and evidence-based policy. Therefore, we would encourage the School Education Department to develop a reliable infrastructure for utilising administrative data and implement a data-sharing policy which can benefit the education department and also contribute to academic research for a high-quality evidence.

Author Contributions: Conceptualization, N.S. and S.S.; Methodology, N.S. and S.S.; Software, Bristol Online Survey, SPSS version 21.; Formal analysis, N.S. and S.S.; Investigation, S.S. and N.S.; Writing—original draft preparation, N.S. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: No funding acquisition.

Institutional Review Board Statement: This study was conducted according to the guidelines of British Education Research Association. The project was formally approved by the University of Education, Lahore, Ethics Review Board.

Informed Consent Statement: All participants volunteered to take part in the anonymous survey and consented to use the information for the research purpose.

Data Availability Statement: The anonymised data files can be accessed by contacting the authors.

Acknowledgments: We thank all the teachers who volunteered to take part in the anonymous survey and consented to use their reported information for the research propose.

Conflicts of Interest: No conflict of interest.

References

1. School Education Department (SED). *Simple, Convenient and Efficient! E-Transfer Public School Teachers 2019*; Punjab Information Technology Board: Punjab, Pakistan, 2019. Available online: Punjab.gov.pk (accessed on 24 February 2021).
2. Siddiqui, N. Socio-economic segregation of disadvantaged children between schools in Pakistan: Comparing the state and private sector. *Educ. Stud.* **2017**, *43*, 391–409. [CrossRef]
3. Alderman, H.; Orazem, P.F.; Paterno, E.M. School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan. *J. Hum. Resour.* **2001**, *36*, 304. [CrossRef]
4. Andrabi, T.; Das, J. *The Rise of Private Schooling in Pakistan: Catering to the Urban Elite or Education the Rural Poor?* The Study of World Bank: Harvard, MA, USA, 2002.
5. Bari, F.; Raza, R.; Aslam, M.; Khan, B.; Maqsood, N. *An Investigation into Teacher Recruitment and Retention in Punjab*; IDEAS—Institute of Development and Economic Alternatives: Punjab, Pakistan, 2013.
6. Siddiqui, N.; Gorard, S. Comparing government and private schools in Pakistan: The way forward for universal education. *Int. J. Educ. Res.* **2017**, *82*, 159–169. [CrossRef]
7. SABER Teachers of Punjab Pakistan. *Systems Approach for Better Education Results (SABER): World Bank*; SABER Teachers Pakistan Punjab Country Report 2018; SABER Teachers of Punjab Pakistan: Punjab, Pakistan, 2018. Available online: Worldbank.org (accessed on 28 January 2021).
8. Hair, N.L.; Hanson, J.L.; Wolfe, B.L.; Pollak, S.D. Association of child poverty, brain development, and academic achievement. *JAMA Pediatrics* **2015**, *169*, 822–829. [CrossRef] [PubMed]
9. Shaukat, S.; Vishnumolakala, V.R.; Al Bustami, G. The impact of teachers' characteristics on their self-efficacy and job satisfaction: A perspective from teachers engaging students with disabilities. *J. Res. Spec. Educ. Needs* **2019**, *19*, 68–76. [CrossRef]
10. Ingersoll, R.M. Teacher Turnover and Teacher Shortages: An Organizational Analysis. *Am. Educ. Res. J.* **2001**, *38*, 499–534. [CrossRef]
11. Toropova, A.; Myrberg, E.; Johansson, S. Teacher job satisfaction: The importance of school working conditions and teacher characteristics. *Educ. Rev.* **2021**, *73*, 71–97. [CrossRef]
12. Falch, T. Teacher Mobility Responses to Wage Changes: Evidence from a Quasi-Natural Experiment. *Am. Econ. Rev.* **2011**, *101*, 460–465. [CrossRef]
13. Asadullah, M.N. Returns to private and public education in Bangladesh and Pakistan: A comparative analysis. *J. Asian Econ.* **2009**, *20*, 77–86. [CrossRef]
14. Allen, R.; Sims, S. *The Teacher Gaps*; Routledge: London, UK, 2018.
15. Khan, T. Teacher Job Satisfaction and Incentive. A Case Study of Pakistan, DFID, Microsoft Word—TMIPa-kistanELDIS.doc. 2005. Available online: Publishing.service.gov.uk (accessed on 5 January 2021).
16. Hanushek, E.A.; Kain, J.F.; Rivkin, S.G. Why Public Schools Lose Teachers. *J. Hum. Resour.* **2004**, *39*, 326. [CrossRef]
17. Feng, L. Teacher Placement, Mobility, and Occupational Choices after Teaching. *Educ. Econ.* **2014**, *22*, 1–24. [CrossRef]

18. Prost, C. Teacher Mobility: Can Financial Incentives Help Disadvantaged Schools to Retain Their Teachers? *Ann. Econ. Stat.* **2013**, *171*. [CrossRef]
19. Neilson, C.; Bobba, M.; Ederer, T.; Leon-Ciliotta, G.; Nieddu, M. Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Peru. 2021. Available online: http://arks.princeton.edu/ark:/88435/dsp01nv935592w (accessed on 24 April 2021).
20. Clotfelter, C.; Glennie, E.; Ladd, H.; Vigdor, J. Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *J. Public Econ.* **2008**, *92*, 1352–1370. [CrossRef]
21. Weiqi, C. The structure of secondary school teacher job satisfaction and its relationship with attrition and work enthusiasm. *Chin. Educ. Soc.* **2007**, *40*, 17–31. [CrossRef]
22. Imazeki, J. Teacher salaries and teacher attrition. *Econ. Educ. Rev.* **2005**, *24*, 431–449. [CrossRef]
23. Jackson, C.K. Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *J. Labor Econ.* **2009**, *27*, 213–256. [CrossRef]
24. Sharma, R.D.; Jyoti, J. Job satisfaction among school teachers. *IIMB Manag. Rev.* **2006**, *18*, 349–363.
25. Kirabo, J.C.; Bruegmann, E. Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers. *Am. Econ. J. Appl. Econ.* **2009**, *1*, 85–108.
26. Rehmani, A. Teacher education in Pakistan with particular reference to teachers' conceptions of teaching. *Qual. Educ. Teach. Leadersh. Challeng. Times* **2006**, *2*, 495–524. Available online: http://ecommons.aku.edu/pakistan_ied_pdck/134 (accessed on 13 July 2021).
27. Delors, J. *Learning: The Treasure Within*; UNESCO: Paris, France, 1996.
28. Darling-Hammond, L. *Powerful Teacher Education: Lessons from Exemplary Programs*; John Wiley and Sons, Inc.: San Francisco, CA, USA, 2006.
29. Bradburb, M.B.; Gilford, D.M. (Eds.) *A Framework for International Comparative Studies in Education*; National Academy Press: Washington, DC, USA, 1990.
30. Halai, A. Equality or Equity: Gender Awareness Issues in Secondary Schools in Pakistan. Equality or Equity: Gender Awareness Issues in Secondary Schools in Pa by Anjum Halai. 2011. Available online: https://doi.org/10.1016/j.ijedudev.2010.06.012 (accessed on 14 July 2021).
31. Naviwala, N. Pakistan Education Crisis: A Real Story. Woodrow Wilson International Center for Scholars. One Woodrow Wilson Plaza. 2016. Available online: https://www.wilsoncenter.org/sites/default/files/pakistanseducationcrisistherealstory2.pdf (accessed on 13 January 2020).
32. Iqbal, H.M. Teachers' content knowledge and pedagogical competence: Issues for teacher status in Pakistan. In *Teacher Status: A Symposium*; Halai, A., Ed.; 2007; pp. 42–65. Available online: http://ecommons.aku.edu/cgi/viewcontent.cgi?article=1006&context=books (accessed on 12 January 2020).
33. Global Partnership for Education (GPE). Technology Boost and Education Reforms in Remote Areas: Pakistan. GPE Transforming Education. 2020. Available online: https://www.globalpartnership.org/results/stories-of-change/pakistan-technology-boosts-education-reform-remote-areas (accessed on 13 January 2020).
34. Chughati, F.D.; Perveen, U. A study of teachers workload and job satisfaction in public And private schools at secondary level in Lahore city Pakistan. *Asian J. Soc. Sci. Humanit.* **2013**, *2*, 202–214.
35. Rivkin, S.G.; Hanushek, E.A.; Kain, J.F. Teachers, Schools, and Academic Achievement. *Econometrica* **2005**, *73*, 417–458. [CrossRef]
36. Barbieri, G.; Rossetti, C.; Sestito, P. *The Determinants of Teacher Mobility. Evidence from a Panel of Italian Teachers*; Ministry of Education: Wellington, DC, USA, 2010. Available online: https://www.bancaditalia.it/pubblicazioni/temi-discussione/2010/2010-0761/en_tema_761.pdf (accessed on 13 January 2020).
37. Bradley, S.; Green, C.; Leeves, G. *The Role of Pecuniary and Non-Pecuniary Factors in Teacher Turnover and Mobility Decisions*; Lancaster University Management School, Economics Department: Lancaster, UK, 2006.
38. Transfer Policy. Government of the Punjab. School Education Department. 2013. Available online: https://schools.punjab.gov.pk/system/files/TransferPolicy-2013_0.pdf (accessed on 17 April 2021).
39. Dawn. Female Teachers Protest against FDE's Repatriation Orders. 14 January 2021. Available online: https://www.dawn.com/news/1601394 (accessed on 17 April 2021).
40. The News Police Make Use of Water Cannons to Thwart Teachers' Protest. 8 September 2020. Available online: https://www.thenews.com.pk/print/711639-female-teachers-protest-transfer-to-remote-areas (accessed on 20 March 2021).
41. Farooq, R.A. *Education System of Pakistan: Issues and Problems*; Asia Society for Promotion of Innovation and Reforms in Education: Islamabad, Pakistan, 1993; pp. 10–20.
42. Bregman, J.; Muhammad, N. *Primary and Secondary Education Structural Issues*; Jan Publications: Islamabad, Pakistan, 1998; pp. 23–31.
43. Andrabi, T.; Das, J.; Khwaja, A.; Vishwanath, T.; Zajonc, T. *Pakistan, Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to Inform the Education Policy Debate*; World Bank: Washington, DC, USA, 2007.
44. Jerrim, J.; Sims, S. When is high workload bad for teacher wellbeing? Accounting for the non-linear contribution of specific teaching tasks. *Teach. Teach. Educ.* **2021**, *105*, 103395. [CrossRef]
45. Gorard, S. *How to Make Sense of Statistics*; Sage: London, UK, 2021.

education sciences

MDPI

Article

Mid-Career Teachers: A Mixed Methods Scoping Study of Professional Development, Career Progression and Retention

Josephine Booth [1,*], Mike Coldwell [1], Lisa-Maria Müller [2], Emily Perry [1,*] and James Zuccollo [3]

[1] Sheffield Institute of Education, Sheffield Hallam University, Sheffield S1 1WB, UK; m.r.coldwell@shu.ac.uk
[2] Chartered College of Teaching, Pears Pavilion, Coram Campus, 41 Brunswick Square, London WC1N 1AZ, UK; lmuller@chartered.college
[3] Education Policy Institute, London SW1W 9TR, UK; james.zuccollo@epi.org.uk
* Correspondence: josephine.booth@shu.ac.uk (J.B.); e.perry@shu.ac.uk (E.P.)

Abstract: Globally, there are ongoing problems with teacher retention, leading to a loss of experience and expertise. In policy and research, the emphasis is often on the professional development and retention of early career teachers, whereas teachers in later stages of their career are relatively underrepresented. This article addresses this imbalance, reporting on a mixed methods scoping study that explores definitions of mid-career teachers in England and their retention and development, via a literature review, primary data collection and secondary analysis of data from the OECD's TALIS 2018 survey. We found that there is no agreed definition of mid-career teacher, relating to time in teaching, role and wider life circumstances and self-definition. Whatever definition is used, mid-career teachers are a heterogenous group, with varying needs, career plans and commitment to the profession. Whilst typically confident in their practice, their learning needs vary and are often experienced as unmet, especially for those looking for progression routes outside leadership and those with family commitments. This indicates that their potential for career development to benefit the profession may not be reached. The article concludes with suggestions for further study, policy and practice to improve understanding of this under-researched group.

Keywords: mid-career teacher; teacher development; teacher retention; teacher careers; scoping study

Citation: Booth, J.; Coldwell, M.; Müller, L.-M.; Perry, E.; Zuccollo, J. Mid-Career Teachers: A Mixed Methods Scoping Study of Professional Development, Career Progression and Retention. *Educ. Sci.* **2021**, *11*, 299. https://doi.org/10.3390/educsci11060299

Academic Editors: Beng Huat See and Rebecca Morris

Received: 30 April 2021
Accepted: 11 June 2021
Published: 16 June 2021

1. Introduction

There is global concern with teacher retention [1,2]. Many countries, including the United States [3–5], Australia [6,7] and some European nations [8], report longstanding and increasing problems with teachers leaving the profession. The situation is complex: teacher retention appears to be variable across contexts, with more acute problems in some subject areas, in more remote geographical and socio-economically deprived locations and for teachers from minoritised groups [9], with further differences deriving from the support offered to teachers, especially in the early stages of their careers [10]. Research has tended to focus on the reasons for teachers leaving the profession, which include high workloads, lack of autonomy and low job satisfaction [11–13]. However, the drivers for teachers to stay in the profession and what can be done to support them, especially beyond the early years of their career, are less well-established. Therefore, there is a need for greater understanding of how policies, initiatives and support mechanisms might lead to more teachers staying in the profession, particularly beyond the first five years of their careers [6,14–16].

In England, recent figures suggest teacher recruitment and retention are critical issues [17], with missed recruitment targets across school phases, and in several subject areas, most notably physics, modern foreign languages and mathematics [18], although there are signs that the COVID-19 pandemic has driven increased recruitment to initial teacher education, which should percolate through to teacher recruitment in 2022 and beyond [19].

The five-year retention rate for teachers continues to drop; for those who qualified in 2014 this is just 67% [18]. In 2019, the Department for Education (DfE) [20] reported that "even small increases in the rate of teachers leaving the profession creates significant, additional pressures on recruitment." Areas of high social and financial deprivation tend to have higher rates of teacher turnover [21,22]. More experienced teachers tend to manage their roles and promote student learning more effectively than those in the early stages of their careers [23] and so teacher attrition removes expertise from the system [22,24], with a negative impact on pupil attainment [23,25,26].

Much research and policy in the area of teacher retention and attrition focuses on teachers in the early years of their careers [27,28]. Initiatives designed to support early career teachers often include professional development aiming to improve teachers' professional knowledge, efficacy, confidence and resilience (for example, [29–31]). For instance, at the time of writing, in England, the DfE's new Early Career Framework [32] is about to be rolled out nationally. This has at its heart a professional development programme, focussed on improving the retention of teachers in their first years in the profession. In relation to teachers at later stages of their careers, in England, the DfE has recently developed a suite of specialist national professional qualifications (NPQs) as professional and career development opportunities for teachers and school leaders who want to "develop their knowledge and skills in school leadership and specialist areas of teaching practice" [33]. However, it remains largely the case that the retention of teachers at later stages of their career is under-researched, and we lack an understanding of appropriate models of support, progression and development for these teachers [34,35].

In this article, we report on a mixed methods scoping study that aimed to explore what is known about mid-career teachers' retention and development. The study used three complementary approaches to address a set of related scoping questions (described below): a literature review, secondary analysis of data from the OECD's 2018 Teaching and Learning International Survey (TALIS) and primary data collection with teachers. As a result of the COVID-19 pandemic, particular attention has been paid recently to teachers' development, retention and recruitment worldwide (for example, [36]). This study largely took place before restrictions and changes to working practices were implemented in England as a response to the pandemic, and it does not attempt or intend to explore the impact of the pandemic on teachers' retention or development.

In the next section, we describe our methods. We move on to addressing the scoping questions in three findings sections, focussing on definitions of mid-career teachers, professional development, and retention and career progression. We then discuss these findings, considering where they complement and contrast each other, our confidence in what we have learned and the gaps and limitations in our findings. In conclusion, we outline opportunities for further research.

2. Methods

2.1. Approach

The questions we aimed to address through this study were:

- in what ways can we define and characterise mid-career teachers?
- how might we categorise mid-career teachers' professional development needs and how do they relate to the professional development that is typically available to teachers?
- can we identify common factors relating to the retention and career development of mid-career teachers, and, if so, what are they?

We used an initial working definition of mid-career teachers as those with five to fifteen years' experience, testing this through the study, as described below. We drew on the OECD's TALIS survey (see below) to define professional development as "activities that aim to develop an individual's skills, knowledge, expertise and other characteristics as a teacher" [37] (p. 49), and considered retention through a variety of measures, again

described below, including teachers' likelihood to stay in or leave the profession, using job satisfaction as a proxy for this where appropriate.

Typically, scoping studies rapidly investigate the current state of the field in terms of published research through a literature review [38]. Their purpose is to map key concepts and the sources and types of evidence available and to identify gaps in the knowledge base [39–42]. In this study, a scoping literature review was supplemented with secondary analysis of publicly available data and primary data collection using quantitative and qualitative methods, as described below. Mixed methods scoping studies appear to be uncommon in education. Our extension of a literature-based scoping approach aligns with other scoping approaches, such as those found in medical literature (see, for example, [43,44]), in which mapping of literature, policy and concepts is supplemented by stakeholder consultation [45].

Our intention in choosing this method was to acknowledge the complex, interacting influences on teacher retention [2] by drawing on multiple sources to compare the existing literature base with current and recent national and international data to quickly test initial hypotheses about gaps in the literature (via secondary and primary data) and to gain stakeholders' responses to emerging issues (via primary data). Through this, we aimed to identify what further research could benefit school leaders, policy makers and mid-career teachers themselves to support their career development and retention in the profession.

In outline, the methods we used (Table 1) were:

- a rapid review of the literature on mid-career teachers;
- secondary analysis of data from the OECD's Teaching and Learning International Survey 2018 (TALIS) study [23];
- primary data collection through surveys and interviews with teachers in England.

Table 1. Methods used, their benefits and limitations.

Method	Summary of Method	Benefits and Limitations
Literature review	• Key search terms • Peer-reviewed literature from the last five years from a range of national contexts • Supplemented by knowledge within the research team of relevant authors and studies • Follow-up relevant literature from references, particularly around definitions of mid-career teachers	• Identifies key issues in existing research evidence, including from systems outside English education • Suggests areas of disagreement and gaps in knowledge • Rapid approach and focus on anglophone studies means key literature may be missed
Secondary data analysis	• Descriptive statistics generated from OECD TALIS 2018 teacher questionnaire on job satisfaction and professional development • Associational regression models relating professional development to job satisfaction	• Tests findings from literature review in current national context • Rigorous data, but limited analysis
Primary data collection	• Survey and focus groups of self-defined mid-career teachers from across school phases	• Rapid, current understanding of issues • Tests findings from other methods • Self-selecting sample of participants

Further details of these methods are given below.

The scoping approach allowed us to be flexible in the sequencing of our methods: each method supplemented and informed the others in an iterative process of feedback, comparison and further analysis. This triangulation of data from three distinct methods provided rich evidence of findings across the knowledge base, enabling identification of

areas of apparent agreement and disagreement, strengthening our emerging understanding from each data source. By using this mixed methods approach, the study builds on the methodological benefits of each approach while mitigating their inherent shortcomings. As is appropriate for a scoping study, the approach also supported the identification of gaps in the knowledge base and areas for further research.

2.2. Literature Review

Initial searches of the past five years of peer reviewed evidence, published in English, were carried out using pre-defined search terms and inclusion criteria (e.g., education systems similar to that of England, such as the other United Kingdom countries, anglophone jurisdictions and systems with high PISA rankings) in relevant databases. Initial search terms included: "mid-career teachers + CPD," "mid-career teachers + professional development" and "mid-career teachers + CPD + retention." These were applied in Google Scholar and educational databases. To supplement these searches, suggestions from colleagues were sought along with studies with which we were already familiar, and references were "snowballed" [46] as appropriate. Grey literature, such as reports produced by government and research organisations, was also included.

Our focus was on studies focussed on mid-career teachers and factors linked to their career progression, retention and professional development in England and other jurisdictions as above. As this was not intended to be a systematic review, we did not assess the strengths of studies, instead using them together as indicators of the nature of research activity in the area, thereby mapping key concepts, summarising emerging themes and identifying gaps [41]. The review of literature identified key areas of focus for secondary data analysis and primary data collection, and in turn, further analysis of the literature was informed by findings from these other methods.

2.3. Secondary Data Analysis

The OECD's Teaching and Learning International Survey (TALIS) is a large-scale international comparative study of teachers and school leaders that focuses on teachers' practices, working conditions and the learning environment in schools. Findings from TALIS are widely used for national and international analysis and comparison, justifications of research and initiatives, and in policy development [47–51].

The TALIS 2018 survey took the form of a 45–60 min online questionnaire for school leaders and teachers of pupils of "lower secondary" age (in England, 11–14 years old), carried out between March and May 2018. Our secondary data analysis used the 2316 teachers' responses from England. To identify mid-career teachers, drawing on initial findings from the literature review, we used a working assumption that they are those with 5–15 years' experience in the profession. There were 995 of these among the responses from teachers in England. The data generated descriptive statistics and correlations between mid-career teachers' job satisfaction and their professional development.

The sampling structure of TALIS clusters teachers within schools; teachers' responses were not independent. To account for that structure, the OECD provides replicate weights that enable the construction of accurate confidence intervals from the weighted sample. We used the weights wherever possible in our analysis. However, weights fit less naturally with some statistics. For example, the number of mid-career teachers in our sample was unweighted because the sample weighting did not include years of experience in its construction.

2.3.1. Professional Development Indicators in TALIS

The TALIS 2018 survey includes several composite indicators related to professional development [52], of which we used three (Table 2). This use of composite indicators avoided relying on single questions and meant that we did not have to develop and validate scales of our own. These individual questions were weighted using confirmatory factor analysis, and the latent variable was provided by the OECD as an indicator [52].

Each of the indicators are unitless and were rescaled to have a mean of zero and a standard deviation of one.

Table 2. Professional development indicators in TALIS used in this study.

Indicator	Related Question(s)
Effectiveness of professional development	The professional development activity that had the greatest positive impact on the respondent's teaching during the previous 12 months: • It built on my prior knowledge • It adapted to my personal development needs • It had a coherent structure • It appropriately focused on content needed to teach my subjects
Professional development needs	Respondent's need for professional development in subject matter and pedagogy in five areas: • Knowledge and understanding of my subject field(s) • Pedagogical competencies in teaching my subject field(s) • Knowledge of the curriculum • Student assessment practices • Student behaviour and classroom management
Barriers to professional development	The degree to which the respondent considers the following are present barriers to their participation in professional development: • I do not have the prerequisites (e.g., qualifications, experience, seniority) • Professional development is too expensive • There is a lack of employer support • Professional development conflicts with my work schedule • I do not have time because of family responsibilities • There is no relevant professional development offered • There are no incentives for participating in professional development

2.3.2. Indicators of Retention in TALIS

In Jerrim and Sims' analysis of the 2018 TALIS findings for the English Department for Education [53], they found that attrition from the profession can be predicted by a composite indicator of teachers' satisfaction with their work environment. That indicator is composed from teachers' responses to four questions:

- I would like to change to another school if that were possible;
- I enjoy working at this school;
- I would recommend this school as a good place to work;
- All in all, I am satisfied with my job.

Therefore, in this study, we used this indicator as our central measure of job satisfaction and predictor of attrition. To supplement this, we used two additional latent measures provided by the OECD [52]:

- The OECD's composite measure of satisfaction with the teaching profession;
- The OECD's composite measure of overall satisfaction. This combines the measure of satisfaction with the work environment with the measure of satisfaction with the profession. It does not include new information beyond those two.

These indicators are, again, unitless latent variables constructed using confirmatory factor analysis.

2.4. Primary Data Collection

To complement the literature review and secondary analysis of data, we carried out a small-scale survey and a series of focus groups with teachers working in England. Survey data was collected prior to COVID-19 restrictions in England; focus groups took place via videoconferencing during a period of limited access to schools for English pupils. The study received ethical approval from Sheffield Hallam University's research ethics committee (reference ER21486548), and all participants gave their informed consent for inclusion before they participated in the study.

Using convenience sampling, members of the Chartered College of Teaching, the professional body of teachers in England, were invited to complete the survey and participate in focus groups. Respondents were asked to self-identify as "mid-career teachers," in order that we could test teachers' perceptions of this stage of their career against findings from the literature.

Survey respondents (88 in total) had an average teaching experience of 14 years, with a range of 5–29 years. Nearly three quarters (73%) had taken one to three career breaks, mostly due to parental leave, which lasted for around one year for the majority of participants. Three focus groups were carried out, including eight participants in total, who had an average teaching experience of 17 years (9–30 years).

The survey (Appendix A) drew on the outcomes of the literature review and the questions used in the secondary analysis for the three composite scores on teacher professional development. These questions were complemented further with questions focussed on issues including changes in job satisfaction over time, self-efficacy, autonomy, career progression, and retention. The focus groups used collated data from the survey as prompts for discussion.

The primary data sample was not representative of the whole teaching profession and did not seek to be, instead providing a current insight into teachers' lived experiences, which complemented the other methods. The majority of respondents to the survey taught in secondary schools (Appendix A), with the next largest group teaching in primary schools. This is unrepresentative of the number of teachers in England, which has roughly equal numbers of teachers working in primary and secondary schools [54], instead being more representative of membership of the Chartered College of Teaching. Eighty-four per cent of respondents were working full-time at the time of completing the survey and about half of those reported having caring responsibilities. The majority of respondents described themselves as female (74%) and white British (86%), which are both roughly representative of the English teaching workforce as a whole [54].

3. Findings

In this section, we present the findings from our study, starting with findings from the literature review on definitions of mid-career teachers. Next, we describe our findings from each of the three methods as they relate to teacher professional development and then to teacher retention and career progression, identifying similarities, differences, confidence and gaps in the results across the three approaches.

3.1. Defining Mid-Career Teachers

Mid-career teachers might be defined in a number of ways: through their years of experience, their professional role, their expertise and/or their dispositions to their role and practice. We begin our findings with a focus on the ways in which we can define and characterize mid-career teachers, exploring the literature which focuses on teachers' progression and development through their careers, including models of teacher career stages and phases.

While there is a sizable body of literature around teachers' careers, which give varying definitions of career and life phases or stages, the term "mid-career" is generally not used. Instead, researchers have developed various models of teacher career phases or stages, which may be collectively known as the career cycle [16]. The most common of these

is a five-stage model (Table 3), typically deriving from Huberman's study of teachers in Switzerland [55]. In these models, career stages are defined by broad categories of years of practice and by teachers' perceptions of themselves, their skills and confidence and their context.

Table 3. Overview of models of career and/or skill acquisition stages.

Dreyfus [56] [1]	Huberman [55]		Day et al. [57]
Novice	Exploration: survival and discovery	1–3 years of teaching	Launching a career: initial commitment (easy or painful beginnings)
Advanced beginner	Stabilisation: teachers choose to commit to teaching or leave	4–6 years of teaching	Stabilisation: find commitment (consolidation, emancipation and integration into peer group)
Competent	Experimentation/activism or reassessment: teachers may question their career choice	7–18 years of teaching	New challenges, new concerns (experimentation, responsibility, consternation)
Proficient	Serenity: a "gradual loss in energy and enthusiasm is compensated for by a greater sense of confidence and self-acceptance" [55]	19–30 years of teaching	Reaching a professional plateau (sense of mortality, stop striving for promotion, enjoy or stagnate)
Expert	Disengagement: either with serenity or disappointment and bitterness	31–40 years of teaching	The final phase (increased concern with pupil learning and increasing pursuit of outside interests; disenchantment; contraction of professional activity and interest, disengagement, serenity)

[1] Note that Dreyfus' skill acquisition categories do not map in terms of years onto Huberman's five career stages.

The second and third stages of the five-stage model appear to be most relevant to the professional lives of mid-career teachers. The third stage Huberman [55] calls "experimentation/activism" and Day et al. [57] "new challenges, new concerns." Within this stage, Huberman sees three interlinking "aspects" through which teachers may move:

- a teacher gains the confidence to experiment with the intention of increasing their impact;
- the teacher realises that there are barriers to experimentation and seeks to make changes in their context, by seeking promotion or becoming an "activist";
- the teacher, having become proficient, grows "stale" [58] (p. 34) and may seek new challenges.

Building on this work, Day and colleagues [58–60] identify, from their studies of teachers in England, six professional life phases, based on the number of years teaching, rather than skills or progression. In these models there are subgroups within the phases where individual teachers may take different paths.

In relation to mid-career teachers, Day [58] defines a "middle professional life phase" for those with 8–24 years' experience (Table 4). In the first part of this phase, some teachers may remain engaged while others may begin to lose motivation. Later, many teachers have additional responsibilities in school, and this, alongside rising levels of personal commitment, means that the management of work/life tensions requires substantial amounts of energy, which in turn has the potential to affect professional motivation, commitment and effectiveness. Three subgroups of teachers in this professional life phase were identified:

Table 4. Mid-career teachers' middle professional life phase, adapted from Day [58].

	Years in Teaching	Phase	Sub-Group
Middle professional life phase	8–15	Managing changes in role and identity: growing tensions and Transitions	(a) Sustained engagement (b) Detachment/ loss of motivation
	16–23	Work/life tensions: challenges to motivation and commitment	(a) Further career advancement and good results have led to increased motivation/commitment (b) Sustained motivation, commitment and effectiveness (c) Workload/managing competing tensions/ career stagnation have led to decreasedmotivation, commitment and effectiveness

- teachers whose commitment and motivation increased;
- teachers who maintained existing levels of commitment, motivation and effectiveness, while coping with competing demands;
- teachers who struggled with these, leading to decreased levels of commitment and motivation and effectiveness.

These stage and phase models suggest linear progressions through a career, albeit with varying timescales and differences for individual teachers. This is not without critique (for example, [61–63]), since teachers may move between stages depending on "a variety of influences related to personal, professional, and organizational factors" [64]. Therefore, these models may fail to adequately acknowledge an "understanding of, and in, practice" [65] and how teachers' skills and knowledge become embedded into their professional practice over time.

However, Huberman [55] acknowledges that differing, non-linear paths may be taken through the stages. Different stages in a teacher's career may or may not relate directly to years of experience, and both "harmonious" and "problematic" trajectories (p. 38) are possible, in which teachers variously experience periods of stability, progression, regression and discontinuities. In this regard, career development might more helpfully be seen as a "process not a series of events" (p. 32).

Some linear models have also been criticised for not recognising that unexpected movement is possible, for example, when teachers choose to challenge existing ideas or initiate change [65]. Therefore, rather than progressing through time, teachers may develop their skills and practice in a non-linear fashion. Some teachers might deepen their understanding of a particular area of practice, such as working with children with special educational needs and disabilities, thereby increasing their specialist expertise in this area without necessarily increasing their range of practice. Others may develop their skills across a range of areas of practice. This recognises the importance of the teachers themselves and "the way in which professionals understand and perform their practice"... "which forms the basis for professional skill and its development" (p. 406).

Finally, the literature suggests that there is not necessarily a link between time spent in the profession and teaching expertise. As Maandag et al. [66] note: "the mere length of a career does not necessarily lead to the development of expertise and improved performance and not all teachers reach high levels of teaching quality in spite of lengthy careers" (p. 8). Indeed, it appears that some teachers become less effective later in their careers as they deal with the challenges of increased responsibility and commitments in and out of school [59,66,67]. While a teacher might have acquired high levels of skills after 5–7 years, and competence a few years earlier, time and experience alone do not make an expert teacher. This highlights the need, which we shall return to later, for continuing professional development for teachers beyond the early years of their career.

In summary, from the literature review:

- there is no agreed definition of a mid-career teacher;
- mid-career teachers are typically defined by their time in the profession, but there are intersections with wider life phases and the development of expertise;
- for at least some teachers, there is non-linear trajectory between and within career phases and career progression.

3.2. Mid-Career Teachers' Professional Development

Next, we consider the second research question: how might we categorise mid-career teachers' professional development needs and how do these relate to the professional development that is typically available to teachers? We look first at findings from the literature, then from our secondary data analysis and finally from the primary data collection.

3.2.1. Literature

There is an existing international body of work on the development needs of teachers at various career stages [68]. Professional development may play a significant role in reducing the numbers of teachers leaving the profession beyond the first five years of teaching [16]. The need for teachers, at all career stages, to engage in what Donaldson [24] calls "career-long learning," and Day et al. [68] an "important professional life investment," "recharging batteries," or "renewal/refreshment" (p. 148), is widely accepted, as is a recognition that not engaging in professional learning could result in eventual disengagement and withdrawal from the profession [69]. However, there is little specific evidence on what constitutes effective professional development other than for teachers in the early stages of their career [66].

For all teachers, professional development opportunities and offers should be tailored to individual teachers' particular context and needs [16,34,62,70–72]. In common with other studies, Kyndt et al. [73] found that conditions both personal (e.g., motivation, autonomy, self-efficacy) and contextual (e.g., in-school support) are vital for professional development to be successful. Day and Gu [59] agree, noting the need for professional development to address the differing needs, influences and issues at different stages of teachers' professional lives, including motivation, commitment and outcomes. As Schwille [74], quoted in Bressman et al. [62], states: opportunities for professional development "must be substantive and relevant so that teachers connect to them" (p. 155).

Since professional learning can be influenced by multiple factors, including teachers' working environment, their professional identity and feelings of self-efficacy, home life and career ambitions [75], it seems likely that there may be key factors of relevance to mid-career teachers' professional development that differ from those for teachers at other career stages. For example, while teacher autonomy may increase during the first six years in the profession, it does not continue to increase unless a teacher moves into a managerial role [68], and a "substantial proportion of teachers report having little direction over how they intend to improve their practice and grow as professionals in future." Avidov-Ungar et al.'s study of Israeli teachers with over eight years' experience [70] found that compulsory professional development was unpopular, with teachers finding that it it was limited in scope and did not suit their needs..

Day and Gu [59] note that each of their six professional life phases (see above) requires tailored professional development, as teachers have differing needs and issues at different stages of their professional lives. For example, they describe the 8–15 years phase as a "key watershed or crossroads in teachers' professional learning and development" (p. 436), when it is vital to offer appropriate professional development because "more than in any other phases, [professional development] is likely to influence their final commitment and effectiveness trajectories" (p. 436). For these teachers, then, it is of particular importance to ensure that appropriate professional learning opportunities are available, particularly that which aims to lead to "enhanced . . . effectiveness" (p. 437). For example, those who have aspirations towards leadership will have differing professional development needs from those who wish to stay as classroom teachers. For those struggling with their work/life

balance, support could focus on "improving their self-efficacy, morale and emotional well-being" (p. 436).

Drawing on Huberman's career stages [55], described earlier, the mid-career teachers' "activist" drive to experiment, refine and diversify teaching skills, such as by combining classroom and management roles, has been recognised in other studies [73,76]. A challenge for those looking to support mid-career teachers may therefore be to consider how to effectively support teachers in moving from "stabilisation" to "experimentation," while ensuring that they do not grow "stale" and, as a consequence, choose to leave the profession. Interestingly, Day et al. [57] found that teachers with 8–15 years' experience rated professional development as more important than teachers at early or later career stages. For the next career stage (16–23 years of teaching), this study found that tensions around work/life balance become more significant, with increased leadership and management responsibilities at work potentially clashing with home lives and teachers feeling that their classroom effectiveness is at risk of compromise.

For those with caring responsibilities outside school and those who work part-time, the need for and access to professional development tailored to career stage may be particularly pertinent, and this may disproportionately affect women. Brown [77] highlighted inconsistencies in terms of access to paid professional development amongst women who worked part-time in England. In this study, an inconsistent approach was identified in terms of remuneration for women to attend professional development on days when they were not normally working; this included inconsistency in schools funding childcare whilst women were attending non-paid professional development.

In terms of the content of professional development for mid-career teachers, Louws et al.'s small-scale study [78] identified some differences in professional development preferences across career phases, with mid-career teachers more focussed on developing new areas of practice (including coaching, teaching children with special educational needs and curriculum development). The authors relate this to a desire to be challenged, motivated or to engage with new practice, which chimes with Huberman's "experimental" third career stage. In an earlier study [79], the same authors found that mid-career teachers showed less enthusiasm for learning about classroom management or learning climate than either early or late career teachers, perhaps due to a belief that they are sufficiently skilled in these areas.

Donaldson et al.'s study [80] concluded that those "second stage" teachers (defined as 3–10 years' experience) who take on "reform" roles—seeking to "change colleagues' practice" (p. 1089)—should engage in professional development to help them both promote their role and support them in encountering resistance in their context. Eros [16] notes that professional development for "second stage" teachers (defined as teachers in Huberman's "stabilization" phase of 4–6 years' experience) should emphasise topics beyond classroom teaching, moving towards leadership or wider contributions to teaching, such as in a family of schools, responding to local educational priorities, or to building new skills such as curriculum development. A study by Richter et al. [76] found that reading of professional literature, and interest in subject content, psychology and subject-specific pedagogies, increased towards mid-career. Avidov-Ungar et al. [70] found that 13 out of 14 advanced primary teachers (with 8–23 years' experience) were interested in "teaching-related courses" (p. 841) rather than those linked to management.

Experienced teachers appear to be more able to relate their learning to wider experiences of teaching than early career teachers [73]. Further, Mertler [81] identified that the chance to take part in "job-embedded, relevant, and ongoing professional development in the form of classroom-based action research" was rated more highly than other kinds of professional development by more experienced teachers (p. 44). Relating to this, Steffy et al. [69] suggest that professional development for experienced teachers could therefore include supporting others, through, for example, coaching of novice teachers, study teams and peer coaching.

Summarising, it appears, from the literature and across various national contexts, that the professional development needs of mid-career teachers differ from those of early career teachers but that they may not be easily classified as a single group. Rather, mid-career teachers have varying professional development needs, based on their career trajectories and ambitions, their expertise, and wider personal and professional priorities and contexts.

3.2.2. Secondary Data

Compared with teachers in the early stages of their careers, mid-career teachers taking part in the TALIS 2018 survey reported a reduced need for professional development, increased barriers to professional development and, notably, that the professional development they engage in is less effective (Figure 1). The change from early career to mid-career is significantly larger than that from mid- to late-career. Mid-career teachers here are those with 5–15 years of experience, whereas early-career teachers are those with less than five years of experience, and teachers with more than fifteen years are described as "experienced."

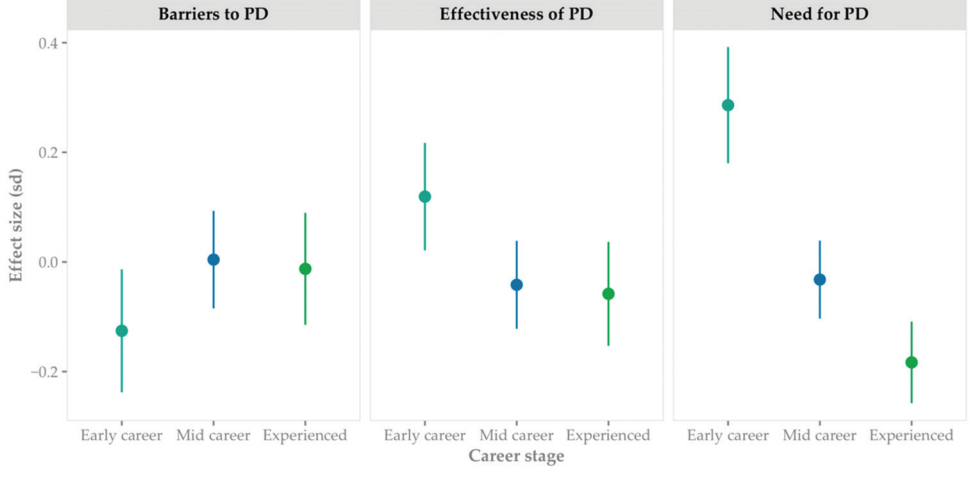


Figure 1. Teachers' experiences of professional development; source: TALIS 2018; note: weighted means and 95 per cent confidence intervals of the mean value, scaled and centred to a full-sample mean of zero and standard deviation of 1.

The data also indicate that mid-career teachers undertake fewer hours of professional development than teachers in other career stages (Figure 2), which is perhaps an indicator of their perceived lack of need and/or its lack of effectiveness.

Even though they report participation in less professional development overall, mid-career teachers engage in broadly similar types and models of professional development to teachers at other career stages (Figure 3); although there do appear to be trends across teachers' careers. For example, as teachers gain experience, they appear to attend fewer courses, conferences and networking events in person.

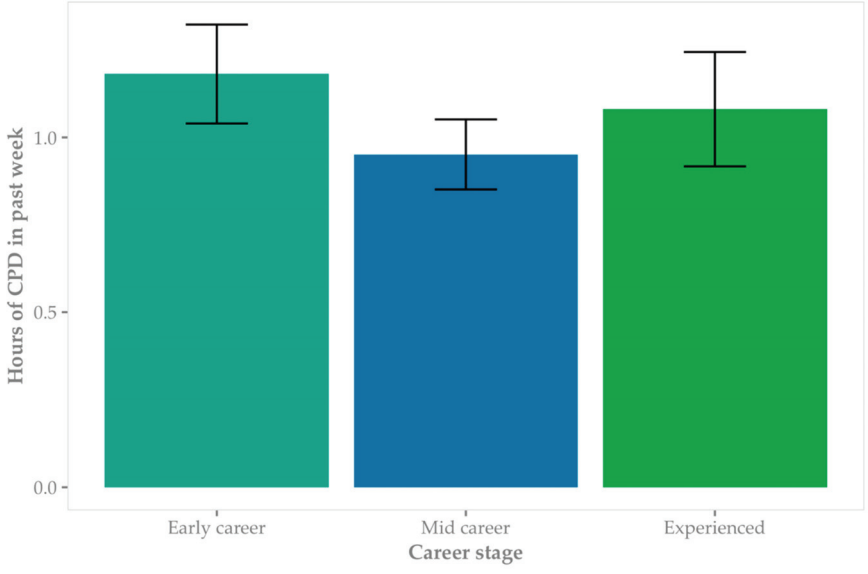


Figure 2. Teachers' time spent engaging in professional development; source: TALIS 2018; note: weighted means and 95 per cent confidence intervals of the mean value.

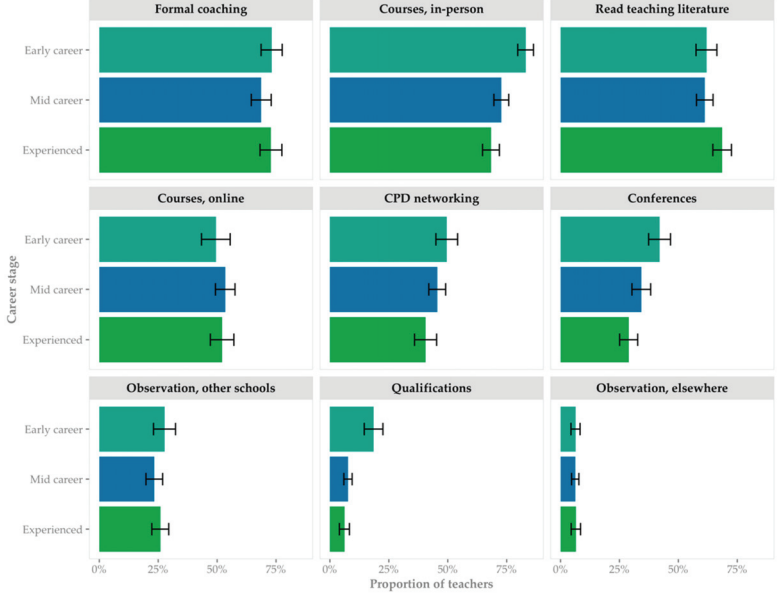


Figure 3. Teachers' engagement in different forms of professional development; source: TALIS 2018; note: weighted response to the question "during the last 12 months, did you participate in any of the following professional development activities?" Whiskers denote 95 per cent confidence intervals.

Reported barriers to professional development were roughly similar across career stages (Figure 4). In line with evidence from the literature, mid-career teachers reported conflicts with their work schedules and competing family commitments as greater barriers

to participation, compared to teachers in other stages of their careers. It is possible that the increased scheduling difficulty contributes to the observed decline in in-person attendance at professional development events indicated above and the slight increase in online participation compared to early career teachers.

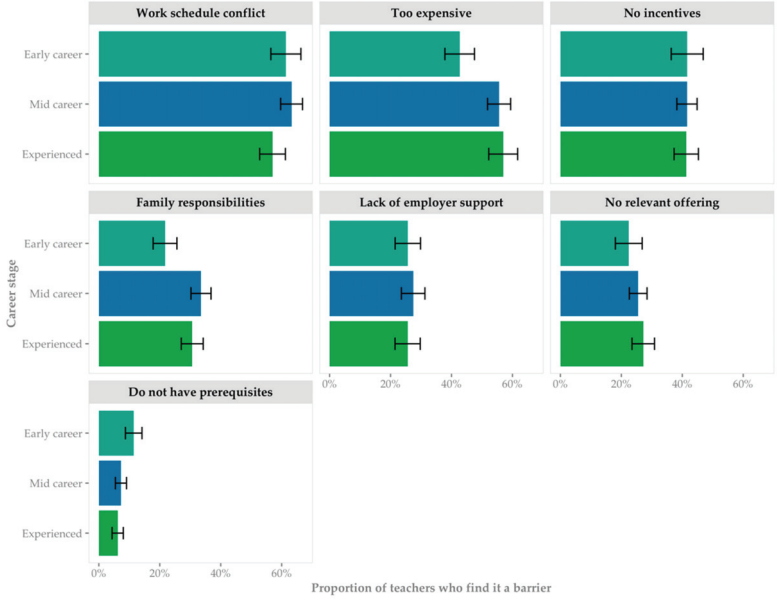


Figure 4. Teachers' reported barriers to participation in professional development; source: TALIS 2018; note: weighted proportion of respondents who agree or strongly agree that "the following present barriers to your participation in professional development?" Whiskers denote 95 per cent confidence intervals.

Looking at the data for mid-career teachers specifically and matching this to data about job satisfaction (using satisfaction with their work environment as a proxy for this) suggests that where these teachers experience more effective professional development, or when they perceive fewer barriers to this, they feel a greater degree of job satisfaction (Figure 5). The association is particularly strong for the latent variables representing barriers to professional development.

Quantifying the strength of the association using linear multiple regression confirmed its strength (Appendix B). The regression also showed the strength of the associations between mid-career teachers' professional development and job satisfaction to be indistinguishable from that of all teachers. It is possible that there is a distinction between mid-career and early-career teachers, but, from this data set, mid-career teachers were not significantly different from the average teacher participating in TALIS.

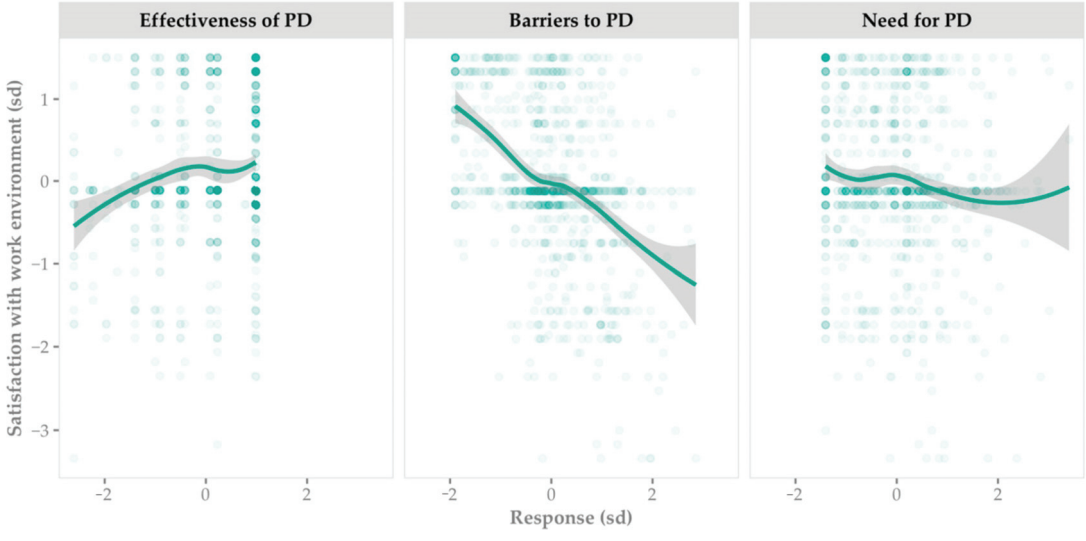


Figure 5. Mid-career teachers' experiences of professional development are correlated with job satisfaction; source: TALIS 2018; note: scatter of latent variables for aspects of professional development against latent variable for satisfaction with work environment. All variables centred and scaled. Loeoursss fit lines with 95 per cent confidence intervals.

3.2.3. Primary Data

Our primary data reveals mixed views about access to and quality of professional development for mid-career teachers. For example, whereas nearly half the participants (49%) indicated (agree/strongly agree) that they have access to high-quality professional development professional development that supports them in their career, nearly the same proportion (42%) felt that they did not. Meanwhile, most teachers in the survey reported participation in between five and 25 hours of professional development activities per academic year.

Over half of survey participants (54%) felt that the professional development they had access to supported their professional development needs, contrasted with nearly 40% who felt that it did not. In focus groups, some regional, phase- and context-specific variations relating to the availability of professional development emerged. For example, teachers in the south-east of England, particularly London, and in government-designated "opportunity areas" [82] and university cities appeared to be more able to access professional development.

Survey respondents felt overwhelmingly that the professional development activities they participated in had a positive effect on their teaching (80%), a finding which contrasts with that of the TALIS data above. Activities that built on teachers' prior knowledge were perceived to have the most impact (Table 5), a finding which indicates the importance of recognising mid-career teachers' experience and expertise in their professional development.

Table 5. Survey responses: thinking of the professional development activities that had the greatest positive impact on your teaching during the last 12 months, did it have any of the following characteristics?

CPD Feature	Percentage	95% CI
It built on my prior knowledge	83.33%	[73.33,93.33]
It provided opportunities to practise/apply new ideas and knowledge in my own classroom	56.25%	[46.25,66.25]
It took place over an extended period of time (e.g., several weeks or longer)	47.92%	[37.92,57.92]
It focused on innovation in my teaching	37.50%	[27.5,47.5]
It provided opportunities for collaborative learning	37.50%	[27.5,47.5]
It had a coherent structure	35.42%	[25.42,45.42]
It provided opportunities for active learning	31.25%	[21.25,41.25]
It adapted to my personal development needs	31.25%	[21.25,41.25]
It appropriately focused on content needed to teach my subjects	27.08%	[17.08,37.08]
It took place at my school	25.00%	[15.00,35.00]
It provided follow-up activities	20.83%	[10.83,30.83]
It involved most colleagues from my school	12.50%	[2.50,22.50]

In terms of barriers to participation in professional development, focus group participants noted that they often engaged in professional development outside working hours, at weekends or evenings, leading to conflicts with family and other commitments. Indeed, conflicting professional development and work schedules were considered by survey respondents to be the most significant barrier to attending professional development (62%), followed by the cost of professional development and lack of incentives to attend. Supporting other findings, around a third of respondents felt that available professional development was not relevant to their learning needs, and that family responsibilities were a barrier to attending professional development. For the full set of responses to this question, see Appendix A.

Some participants highlighted a lack of access to professional development that was relevant to their career stage. As one participant phrased it: "once you're the expert, there aren't any other experts." Closer connections to higher education, either through initial teacher education or postgraduate study, were suggested as potential solutions to this issue. For example, two participants referred to their experience completing master's degrees and the Chartered Teacher programme [83] alongside teaching, suggesting that these experiences were beneficial by providing opportunities to engage with research, to reflect on practice and to conduct small-scale inquiries. In one school, teachers were supported to complete master's degrees by the provision of part-time contracts for the duration of study.

Views were also split regarding schools' prioritisation of professional development. Whereas 43% of survey respondents agreed or strongly agreed that, in their school, it was a priority to ensure that teachers have access to high-quality professional development, 41% disagreed or strongly disagreed with this. In focus groups, some participants described systems where professional development is woven into the school's fabric, with an offer described as personalised, regular and subject- and career-stage specific. However, other teachers described issues such as schools' priorities clashing with their own and lack of time away from the classroom to participate in professional development, especially the case for teachers with large numbers of examination classes. Whole-school professional development was not seen as particularly relevant to participants because it appeared to be either aimed at colleagues earlier in their career and/or lacked subject specificity.

When asked how current professional development opportunities could be improved, the majority of responses referred to an increase in time allocated to professional development, more strategic planning, greater financial support and an offer extending beyond leadership development. They also mentioned increased flexibility, more bespoke offers that take teachers' subject specialism and their career stage into account, and the need to sustain professional development over time, so teachers are able to reflect on learning and develop their practice.

Some suggestions to improve the relevance of professional development programmes for mid-career teachers included: more professional development focused on teaching and learning and opportunities to combine postgraduate study with teaching, although some participants felt that such approaches may be easier towards the beginning of a teaching career when teachers tend to have fewer caring responsibilities. Further, it was suggested that coaching and mentoring could facilitate more personalised approaches to professional development, by enabling teachers to identify their goals and define their professional development priorities. In some schools, coaching and mentoring was common practice, sometimes combined with appraisals and performance management, but other schools did not offer any support for coaching or mentoring.

The need for more professional development opportunities outside the leadership track emerged from focus group discussions. Where participants reported high levels of job satisfaction, this appeared to relate to a strong school-led focus on professional development and high levels of autonomy in choosing professional development that related to their professional development needs. A majority of focus group participants emphasised that the system does not currently provide sufficient developmental opportunities for teachers who want to build their expertise in teaching, valuing their experience, rather than moving into leadership. This is illustrated in the following quote: "I'm teaching eight years now. I think after five years, teachers should be given funded or partially funded courses to go back into the research about education: time within schools to complete these courses. More work on the different pathways. I'm a second-in-department, and my traditional next step is to become a head of department, which I don't want."

Teachers' autonomy in shaping their own professional development emerged as another issue with mixed responses. Thirty-five percent of survey respondents indicated that they have little or no autonomy in choosing their professional development, contrasted with 48% who indicated that they have at least some level of autonomy in choosing their professional development. The need for more autonomy in choosing professional development was highlighted in focus group discussions, illustrated by this quote: "they would need to directly address my own personally-directed needs and wants for myself as a teacher: I know where and what I want to develop." Those participants who had more autonomy in choosing professional development were able to choose opportunities that were most relevant to them, were given time to read literature that they found to be helpful, were engaged in coaching or visited other schools.

These findings suggest that mid-career teachers, represented here by the participants in our primary data collection, have varying experiences of professional development, which appear to be influenced, at least in part, by their school contexts. To improve access and engagement, some possible actions include more flexible participation, better targeting of professional development in relation to teachers' career stages, and opportunities for professional learning outside leadership development.

3.3. Mid-Career Teachers' Retention and Career Progression

We end the findings by looking at the third research question: can we identify common factors relating to the retention and career development of mid-career teachers, and, if so, what are they? Again, we firstly consider findings from the literature, then from the secondary data analysis and finally from our primary data collection.

3.3.1. Literature

Based on the career stage models described earlier, it is likely that the reasons for mid-career teachers leaving the profession differ from those earlier in their professional lives. For example, there appears to be a group of teachers in the "middle professional life phase" [58] who may begin to lose motivation. Mertler's study of teachers in Arizona [81] showed that, in general, teachers with greater experience in the profession had higher levels of dissatisfaction with their jobs and therefore were at higher risk of leaving. The author concluded that "it is quite likely that Arizona's teachers are satisfied with the work that they perform, but not with the recognition they receive" (p. 43), particularly around the issue of pay. On the other hand, some mid-career teachers feel that they have already made a substantial commitment of time and energy to the profession and gained a significant amount of profession-specific knowledge [61]. It is not clear whether this increased "professional capital" makes teachers more likely to stay in the profession, although Borman and Dowling's meta-analysis of US studies of teacher attrition [61] found that "there is somewhat more evidence suggesting that it is the more talented rather than the less talented teachers-those who are better trained, more experienced, and more highly skilled-who tend to be lost to turnover with greater frequency" (p. 396).

A lack of autonomy, for example, around professional development or career progression, has been suggested as a reason why experienced teachers choose to leave the profession [16]. Further, as we have discussed above, experienced teachers may feel discouraged when their development needs are ignored [35], and, arguably more than teachers in other career stages, they may seek opportunities to work flexibly and may require support in balancing competing priorities. If these are unavailable, they may choose to leave the profession.

After five years in the profession, teachers are in Huberman's "stabilisation" phase, where they choose to commit to teaching or leave, and, if taking the former route, may be looking beyond the classroom to build a long-term career in teaching [59]. However, in practice, teachers are often not given a clear progression route in their careers. For example, there tend to be few opportunities for promotion [73] and, in any case, as we have seen already, not all teachers wish to take on leadership roles [24]. Therefore, in order to retain and support teachers in their second or third career stages, schools must develop "new pathways for these seasoned educators to flourish" (p. 168) [62]. In an English context, Worth et al. [84] agree, stating that "alternative career pathways which keep good teachers teaching, rather than taking up management positions" (p. 14) might offer a route to improved retention.

Coldwell [28] noted a link between satisfaction, efficacy and retention in earlier literature, stating there is "evidence of a relationship between efficacy and intention to stay in the profession such as Brouwers and Tomic (2000) that showed that higher self-efficacy was related to lower levels of teacher 'burn out' on a range of measures, and other reviews link efficacy to teacher retention, again with generally weak relationships found (Tschannen-Moran, Woolfolk Hoy & Hoy, 1998; Klassen, Tze, Betts & Gordon, 2011). They also suggest that we might usefully focus our attention on secondary teachers, where the greatest effects are found, and where job satisfaction is lower and attrition is higher." Meanwhile, analysis of the 2013 TALIS survey [85] found that "increased levels of effective professional development are associated with a reduced desire for teachers to move schools. More specifically, a one SD improvement in effective professional development is associated with a large, 63% reduction in the odds that a teacher wants to move to another school. This is almost as strong as the association with leadership quality" (p. 38).

Subsequent analysis in England linked data on teachers' attrition to questions in the TALIS 2018 survey about job satisfaction [53]. This found that secondary school (11–18) teachers have lower job satisfaction than primary (4–11) teachers and are, consequently, more likely to leave their jobs. It also found that the quality of leadership in the school has a strong association with job satisfaction and the chances that a teacher will leave the profession. While these findings are not specific to mid-career teachers, they illustrate

how variations in job satisfaction provide a proxy measure of the probability of leaving the profession.

Overall, in common with findings described above, the literature here points towards mid-career teachers as a heterogeneous group, with varying experiences, levels of commitment and job satisfaction and motivations. These can be limited by a lack of opportunities for progression that meet teachers' aspirations for their career development.

3.3.2. Secondary Data

Turning to our secondary analysis of TALIS 2018 data, teachers' satisfaction with the work environment was fairly constant across career stages (Figure 6), and once teachers are beyond the early stages of the career so is their overall satisfaction with the profession. There appears to be little difference here between mid-career teachers and other experienced teachers, suggesting that other elements of the working environment contribute to teachers' job satisfaction in addition to experience. There is wide variation in job satisfaction among teachers, but very little is related to the teacher's career stage. The variation within career stages is far greater than the variation across career stages. It should be noted that the measure of job satisfaction is standard deviations from the overall average level of satisfaction, and so it describes only relative levels of job satisfaction between the groups and says nothing about the absolute level of satisfaction.

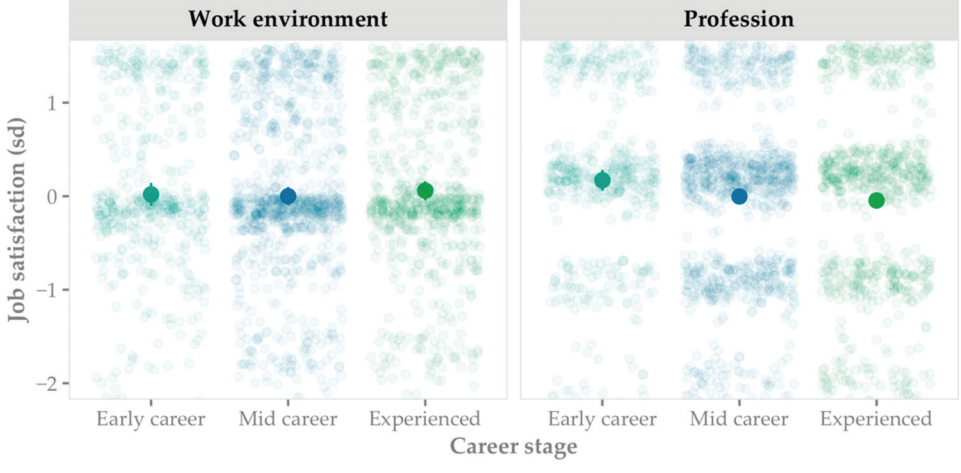


Figure 6. Teachers' satisfaction with their work environment and profession by career stage; source: TALIS 2018; note: background points are an unweighted scatter of individual respondents' satisfaction with their work environment and profession; both latent variables, centred and scaled. Large points in the foreground are the weighted means of those variables with 95 per cent confidence intervals denoted by whiskers.

3.3.3. Primary Data

Overall, respondents to the online survey displayed high levels of job satisfaction. The majority of respondents (62%) agreed or strongly agreed with the statement that they were satisfied with their job, although nearly a quarter of respondents (24%) were not satisfied. Interestingly, over half (57%) of the respondents indicated greater satisfaction now than at the start of their career, although nearly one third (29%) indicated that their job satisfaction was lower now than at the beginning of their career. This may be reflective of the evidence from the literature of the varying, individualised routes teachers take through their career stages.

Drawing on Day and Gu's descriptions of teachers' professional life phases [60], the survey included questions in which teachers rated their self-efficacy relating to various

aspects of their practice (Appendix A). Overall, respondents reported very high levels of self-efficacy, especially in the areas of teaching and learning. With participants in the survey reporting both high levels of job satisfaction and high levels of self-efficacy, these findings support positive correlations shown in other studies (for example, [85]).

Given these high levels of job satisfaction and self-efficacy, it is surprising that a large majority of survey respondents (86%) indicated that they had considered leaving the teaching profession at some point in their career, and nearly half (44%) of respondents indicated that they were considering leaving the profession at the time of completing the survey. However, our findings also indicated that this group of teachers is not planning an imminent departure from teaching.

A lack of relevant professional development opportunities, as discussed above, appears to be an important factor in teacher retention here, with a majority of respondents (57%) saying that a better professional development offer would encourage them to stay in the profession and help to keep them motivated and interested in teaching. As one survey respondent put it: "CPD [continuing professional development] is vital to teacher retention. Without innovative, forward thinking, exciting CPD, teaching could become a 'job' rather than a 'profession.'"

The lack of relevant professional development also appears to be a barrier to teachers' career progression. The majority of survey respondents (53%) felt that the professional development they participated in was not relevant to the next step in their careers. During focus group discussions, it emerged that, for some teachers, the order of professional development and career progression can sometimes be inverted, with teachers being promoted to leadership roles before completing related professional development. Interestingly, participants felt that expert teaching was not sufficiently rewarded in comparison to managerial roles, when the opposite should be the case or they should, at least, be equally rewarded.

As we saw in the literature review, parental leave and family commitments were highlighted by several female participants as obstacles to career progression. A lack of flexibility in middle and senior leadership roles (such as opportunities for part-time working and/or job sharing) could hinder teachers with caring responsibilities to progress in their careers. The workload associated with these roles was further mentioned as being incompatible with family commitments, leading three participants to consider moving into higher education, based on a perception of this allowing for more flexible work arrangements.

These findings suggest that, even though teachers' job satisfaction may increase with experience, opportunities for flexible, individualised engagement in careers and professional development are not consistently available to mid-career teachers. While the impact of this is not clear, in the long-term, it may lead to dissatisfaction and influence the likelihood of teachers' retention in the profession.

4. Discussion and Conclusions

In this study we aimed to identify, through a scoping study, the ways in which mid-career teachers might be defined or categorised, thereby to better understand their professional development, career development and retention in the profession and to identify questions for further research. We took a purposefully rapid – and therefore inevitably partial—approach, aiming to explore current understanding in the literature, from existing data and from practitioners. In places, our findings were complementary; in others, they did not fully support each other. As mentioned in the introduction, the bulk of this study took place before restrictions relating to COVID-19 were implemented in England, and so, if a further review of the literature and/or primary data collection were conducted now, our findings might take a different shape.

In this section, we bring together the findings described above, referring back to our research questions, in order to explore what we can learn from this study and what further research might be valuable.

Turning first to the ways in which mid-career teachers might be defined and categorised, we considered whether this group can be identified in terms of years in the

profession, professional role or expertise. We took different approaches to this: examining the literature to identify categorisations from previous studies, using a working definition based on years of experience for the secondary data analysis and enabling self-identification in the primary data collection.

The literature indicates that there is no widely accepted definition of what constitutes a "mid-career teacher." However, there is some agreement that teachers enter a "second stage" of teaching between around three or four and ten years in the profession. This period of teachers' professional lives, which follows the early career stage, covers, from the literature, Huberman's stabilisation and experimentation/activism or reassessment phases [55]; the advanced beginner and competence stages from Dreyfus [56]; the stabilisation and new challenges, new concerns stages from Day et al.'s five career stages [57]; and, from Day and Gu [60], the second part of the early professional life phase and the first part of the middle professional life phase. Therefore, for the secondary data analysis, we tested a definition of mid-career teachers as those with between five and fifteen years' experience. We found that this broadly aligned with the years of experience of those who self-identified as mid-career teachers in the primary data collection, suggesting that teachers enter what might be called their "mid-career" after around five years in the profession.

Focusing on whether mid-career teachers might be categorised in other ways, we identified some similarities in this group of teachers. Specifically, a key characteristic of mid-career teachers is the management of competing priorities and different aspects of professional and personal lives. This is brought about by a lack of time caused by the pull between family and work, and, for many, a lack of attractive development opportunities outside a "traditional" route to leadership.

Overall, mid-career teachers appear to be a less homogeneous group than early career teachers, who have relatively consistent development needs as beginning practitioners, although with individual and contextual variation related to subject, phase, school and individual needs and aspirations [29]. Mid-career teachers enter and leave this phase of their career in differing ways, with diverse professional experiences, varying trajectories within the phase, and changing priorities. These depend on a variety of factors including teachers' professional contexts and opportunities for development and progression and their own attitudes, beliefs and personal lives. Therefore, in order to better understand the retention and professional development of mid-career teachers, we need to consider their experiences, aims, interests and priorities, both within and outside work [14]. These are dynamic, varying over time alongside the ongoing development of expertise and/or changing professional or personal contexts.

This leads us to consider the second research question, on the professional development needs and experiences of mid-career teachers and whether or how these differ from those of other teachers. There are few studies that focus specifically on the development needs of mid-career teachers, other than those relating to school and subject leadership, which—as we have seen—mid-career teachers may not necessarily see as attractive development options. Teachers in the mid-career stage appear, from our primary data collection, to feel confident in their classroom practice and may perceive a lower need for some forms or types of professional development, compared to early career teachers. This may indicate a different set of professional development needs or indicate, as we have seen in our study, a lack of opportunities to develop in their contexts, and this may be linked to the correlation between teaching experience and decreased participation in in-person courses. Further, to these teachers, the concept of career progression can appear to be overly focussed on leadership roles rather than the further extension and development of expertise, practice or specialist interests.

However, as already noted, mid-career teachers are a heterogeneous group; in fact, our findings suggest that between-group differences (such as those between early career teachers and mid-career teachers) are smaller than differences between those in the mid-career group. Mid-career teachers as a group have varied, complex experiences and, importantly, differing—and dynamic—aspirations for their careers. This firstly means

that their professional development needs are varied, based on individual ambitions, motivations, expertise, circumstances and wider life orientations. The ways in which mid-career teachers choose or opt to participate in professional development activities are also likely to differ. For example, some teachers may wish to develop through formal, accredited postgraduate routes; others through inquiry-led approaches. Competing priorities between work and home lives mean that opportunities should also be offered in ways that are flexible and adaptable. There is a need to identify and recognise the individual needs of mid-career teachers, the types and areas of potential development needs, and to acknowledge that these are dynamic and as such may change over time or context.

The final research question relates to the retention of mid-career teachers. Workload is the most commonly cited reason for leaving the profession for teachers across career stages [14]. Looking across our three methods, we found that for mid-career teachers, workload plays a significant role, both in relation to retention directly and indirectly as a factor that restricts engagement in professional development. This is in line with earlier work developing "path models" [86] and sits alongside other factors specific to this group of teachers. In particular, our sources of data agree that competing priorities, particularly a lack of time caused by a pull between family and work, are key characteristics of mid-career teachers. This suggests that supporting mid-career, and indeed all, teachers to manage their workload, through flexible models of working, career progression and professional development opportunities, may increase retention. Interestingly, in England, the Department for Education has recently announced an initiative to promote flexible working for teachers [87].

Looking at job satisfaction, as a measure of potential retention in the profession, our findings suggest that for many teachers beyond the early stages of their careers, job satisfaction is relatively high: the secondary data analysis shows that teachers beyond the early stages of their career have the highest levels of job satisfaction, and, in our primary data collection, a majority of respondents indicated greater job satisfaction in their mid-careers than in its early stages. The literature review suggests that for some teachers, there are signs of increased or maintained commitment and professional skills gained through experience ("professional capital"). Increased autonomy is generally perceived to be a positive aspect of increased experience, leading to greater job satisfaction, and a lack of autonomy appears to contribute to demotivation and lower satisfaction. However, our findings suggest that, for some teachers, increased autonomy can lead to a feeling of being left unsupported, with limited availability of opportunities for development or progression.

Taken together, our findings suggest that mid-career teachers are a group of teachers that are easily overlooked: largely committed, skilled and intending to stay in the profession, but often feeling undervalued, juggling competing priorities and lacking opportunities for development. Further, we can see, although with somewhat lower confidence, that within the group of mid-career teachers there are variations in professional development needs, career trajectories and aspirations and levels of commitment to the profession.

Therefore, considering how policy makers might support mid-career teachers, and to open up possibilities for further investigation, we tentatively propose four broad, and potentially overlapping, groups of mid-career teachers:

- Career climbers: those moving up the traditional career ladder to leadership roles;
- Satisfied stalwarts: largely satisfied, often skilled practitioners, committed to the profession, without the intention to develop their careers in terms of leadership roles or awareness of how to do so;
- Family focussers: those maintaining, or trying to maintain, a dual focus on family life and career;
- Dissatisfied stayers: those that see little alternative to teaching, with few career ambitions and dissatisfaction with their role.

Our findings suggest that teachers move between these groups as their careers, personal lives and contexts change and that these trajectories may be influenced by the support they receive. Meanwhile, each of these groups, and those of multiple subgroups within

them, is likely to have different development and learning needs, based on their experience, interests, motivations and working patterns. With this in mind, effective models of development should be based on the identification of particular learning needs, should build on and recognise teachers' expertise and should provide attractive progression routes that look beyond leadership opportunities. At a policy level, flexible models of working and of development, such as the English Department for Education's Specialist NPQs [33], appear to be potentially valuable opportunities to support mid-career teachers. A further avenue to consider is the potential for a career and/or progression framework for teachers beyond the early stages of their careers, so that varying career development routes are available and valued.

Given the nature of this project as a scoping study, inevitably these suggestions are tentative and require further consideration. Therefore, potential lines of research to meet the gaps in the research evidence include the following:

- investigate the four proposed groupings of mid-career teachers via primary research;
- examine different models of in-school support including flexible working and development models;
- further examine the relationships between development and support models and retention and/or career satisfaction of mid-career teachers.

To conclude, this scoping study indicates that mid-career teachers are a heterogeneous group who would benefit from an environment that is supportive in terms of professional development, flexible in terms of balancing competing priorities and that offers help with identifying opportunities for learning without being overly directive. This type of environment will offer teachers autonomy and agency in identifying their own areas for learning and where they choose to spend their time and effort in terms of professional development or other activities that they feel to be of benefit professionally. We offer some suggestions to enable this to happen. In order to provide the most appropriate support, and therefore for the profession to benefit most from this group of teachers, further rigorous research following the lines of enquiry developed in the scoping study is required.

Author Contributions: Data curation, J.B., M.C., L.-M.M. and J.Z.; Formal analysis, L.-M.M. and J.Z.; Funding acquisition, J.B., M.C., L.-M.M., E.P. and J.Z.; Methodology, J.B., M.C., L.-M.M., E.P. and J.Z.; Project administration, E.P.; Validation, J.Z.; Visualization, J.Z.; Writing—original draft, J.B., M.C., L.-M.M., E.P. and J.Z.; Writing—review & editing, J.B., M.C., L.-M.M., E.P. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Sheffield Hallam University's allocation from the Research England QR Strategic Priorities Fund 2020.

Institutional Review Board Statement: The study received ethical approval from Sheffield Hallam University's research ethics committee (reference ER21486548, approved January 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The secondary data presented in this study are openly available from the OECD at https://www.oecd.org/education/talis/talis-2018-data.htm (accessed on 15 June 2021).

Acknowledgments: The authors acknowledge the contribution of Suzanne Brown to the original review of literature.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Primary Data Collection

Survey Questions

Job Satisfaction

Please indicate how far you agree with the following statements.

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
Overall, I am satisfied in my job.					
Overall, I am more satisfied in my job now than I was at the start of my career.					

Self-Efficacy

Please indicate below how far you agree with the following statements.

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
I can control disruptive behaviour.					
I can calm a student who is disruptive or noisy.					
I can help students to value learning.					
I can formulate good questions for my students.					
I can get students to follow classroom rules.					
I can get students to believe that they can do well in school.					
I can establish a classroom management system with each group of students.					
I can use a variety of assessment strategies.					
I can provide alternative explanations or examples when students are confused.					
I can assist families in helping their children do well in school.					
I can implement alternative teaching strategies in my classroom.					

CPD

Please indicate how far you agree with the statements below.

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
I have access to high-quality CPD which supports me in my career.					
Ensuring teachers have access to high-quality CPD is a priority in my school.					
The CPD I participate in supports my individual development needs.					
I have autonomy in choosing CPD that is relevant to me.					

Have you done any of the following? Please tick all that apply.

- acted as a mentor for ITE [initial teacher education] students
- acted as a mentor for newly qualified teachers
- taken on a leadership role in your school
- taken on a leadership role across a wider group of schools

In the last 12 months, did you participate in any of the following professional development activities? Tick all that apply.

- Courses/seminars attended in person
- Online courses/seminars
- Education conferences where teachers and/or researchers present their research or discuss educational issues
- Formal qualification programme (e.g., degree)
- Observation visits to other schools
- Observation visits to business premises, public organisations or non-governmental organisations
- Peer and/or self-observation and coaching as part of a formal school arrangement
- Participation in a network of teachers formed specifically for the professional development of teachers
- Reading professional literature
- Other (please specify)

For the professional development in which you participated in the last 12 months, did you receive any of the following? Please tick all that apply.

- Release from teaching duties for activities during regular working hours
- Non-monetary support for activities outside working hours (e.g., reduced teaching time, days off, study leave)
- Reimbursement or payment of costs
- Materials needed for the activities
- Monetary supplements for activities outside working hours
- Non-monetary rewards (e.g., classroom resources/materials, book vouchers, software/apps)
- Non-monetary professional benefits (e.g., fulfilling professional development requirements, improving my promotion opportunities)
- Other (please specify)
- Increased salary

Thinking of all the professional development activities during the last 12 months, did any of these have a positive impact on your teaching practice?

- Yes
- No

Thinking of the professional development activities that had the greatest positive impact on your teaching during the last 12 months, did it have any of the following characteristics?

- It built on my prior knowledge
- It adapted to my personal development needs
- It had a coherent structure
- It appropriately focused on content needed to teach my subjects
- It provided opportunities for active learning
- It provided opportunities for collaborative learning
- It provided opportunities to practise/apply new ideas and knowledge in my own classroom
- It provided follow-up activities
- It took place at my school

- It involved most colleagues from my school
- It took place over an extended period of time (e.g., several weeks or longer)
- It focused on innovation in my teaching

Approximately how many hours of CPD do you participate in, on average, per academic year?

- 0–5
- 5–25
- 26–50
- over 50

How strongly do you agree or disagree that the following present barriers to your participation in professional development?

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
I do not have the pre-requisites (e.g., qualifications, experience, seniority)					
Professional development is too expensive					
There is a lack of support from my school leaders					
Professional development conflicts with my work schedule					
I do not have time because of family or other responsibilities					
There is no relevant professional development offered					
There are no incentives for participating in professional development					

What are your professional learning and development priorities for the next couple of years?

What could be improved about the current CPD offer in your school to better support you in your career?

Career Progression

How would you rate your career progression so far?

- Much better than expected
- Better than expected
- As expected
- Worse than expected
- Much worse than expected

Please elaborate on your choice above.

How far do you agree with the following statements?

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
I have sufficient opportunities for career progression within my school/group of schools.					
I have sufficient opportunities for career progression in the education system more widely.					
I know what I need to do to achieve the next step in my career.					
The career paths that are currently available in the system are relevant to me.					
The CPD I attend is relevant to the next step in my career.					

Specialist NPQs

How likely are you to take up the new specialist Teacher Developer NPQ once it is rolled out?

- Very likely
- Likely
- Unlikely
- Very unlikely
- I don't know
- I haven't heard of it

What content should the new specialist "Teacher Developer" NPQ or other specialist NPQs include?

Retention

Have you ever considered leaving the teaching profession?

- Yes
- No
- Prefer not to say

Are you currently considering leaving the teaching profession?

- Yes
- No
- Prefer not to say

When are you planning on leaving the teaching profession?

- Within this academic year
- After the end of this academic year
- Within the next 1–5 years
- Not sure
- Prefer not to say

Would better CPD opportunities encourage you to stay in the profession?

- Yes
- No

What would these CPD opportunities need to fulfil to encourage you to stay in the profession?

What role, if any, does CPD play in encouraging you to stay in the profession?

Personal Information

In this section we will ask you some information that will help us to contextualise your answers. This information will only be analysed on a group level.

Which of these best describes the setting you currently teach in?

- FE college
- Sixth form college
- Secondary (10–18)
- Secondary (10–16)
- All-through, with sixth form
- All-through, without sixth form
- Middle
- Primary
- Junior
- Nursery
- Infant
- Other (please specify)

How long have you been teaching for?

Have you had any career breaks?

- Yes
- No
- Prefer not to say

How many?

For how long in total?

Reason/s for your career breaks. Tick all that apply.

- Parental Leave
- Other caring responsibilities
- Sabbatical
- Health problems
- Prefer not to say
- Other (please specify)

Is your position full-time?

- Yes
- No

Do you currently have any caring responsibilities? (e.g., young children, elderly relatives, etc.)

- Yes
- No

What is your gender?

- Female
- Male
- Other
- Prefer not to say

What is your ethnicity?

- White British
- White Irish
- White Other
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other mixed background
- Indian

- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background
- African
- Caribbean
- Any other black background
- Any other ethnic group

53. How old are you?

- 20–25
- 26–30
- 31–35
- 36–40
- 41–45
- 46–50
- 51–55
- 56–60
- over 60

Would you like to participate in phone interviews or online focus groups to provide further information on this topic?

- Yes
- No

Table A1. Survey responses: school phase.

Respondent's School Phase	Respondents (%)
Sixth form (16–18)	8%
FE College	2%
Secondary (11–16)	27%
Secondary (11–18)	35%
All-through school (3–18)	8%
All-through school (3–16)	2%
Primary (3–11)	14%
Junior	2%
Other	2%

Table A2. Survey responses: teachers' self-efficacy: please indicate below how far you agree with the following statements.

Question	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
I can control disruptive behaviour	33.33%	59.42%	4.35%	2.90%	0.00%
I can calm a student who is disruptive or noisy	26.09%	69.57%	4.35%	0.00%	0.00%
I can help students to value learning	31.88%	59.42%	7.25%	1.45%	0.00%
I can formulate good questions for my students	46.38%	50.72%	2.90%	0.00%	0.00%
I can get students to follow classroom rules	44.93%	50.72%	4.35%	0.00%	0.00%
I can get students to believe that they can do well in school	34.78%	53.62%	10.14%	1.45%	0.00%
I can establish a classroom management system with each group of students	39.13%	55.07%	5.80%	0.00%	0.00%
I can use a variety of assessment strategies	50.00%	47.06%	2.94%	0.00%	0.00%
I can provide alternative explanations or examples when students are confused	60.29%	36.76%	2.94%	0.00%	0.00%
I can assist families in helping their children do well in school	26.47%	47.06%	23.53%	2.94%	0.00%
I can implement alternative teaching strategies in my classroom	36.23%	60.87%	2.90%	0.00%	0.00%

Table A3. Survey responses: barriers to professional development—how strongly do you agree or disagree that the following present barriers to your participation in professional development?

Question	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
I do not have the prerequisites (e.g., qualifications, experience, seniority)	47.54%	27.87%	16.39%	4.92%	3.28%
Professional development is too expensive	9.84%	13.11%	18.03%	47.54%	11.48%
There is a lack of support from my school leaders	18.03%	22.95%	19.67%	18.03%	21.31%
Professional development conflicts with my work schedule	4.92%	13.11%	19.67%	45.90%	16.39%
I do not have time because of family or other responsibilities	19.67%	27.87%	21.31%	27.87%	3.28%
There is no relevant professional development offered	31.15%	22.95%	9.84%	22.95%	13.11%
There are no incentives for participating in professional development	11.48%	18.03%	11.48%	37.70%	21.31%

Appendix B. Regression of Job Satisfaction on Mid-Career Teachers' Experiences of Professional Development

Table A4 reports the coefficients and standard errors (in parentheses) of regressing the dependent variable of a teacher's satisfaction with their work environment on several independent variables discussed in the main text. All variables were centred and standardised so coefficients can be interpreted similarly to an "effect size," although they show associations rather than causal effects. Four specifications are reported.

The association between job satisfaction and the barriers to professional development remains the strongest by some margin (0.36 sd) even after conditioning on the other measured aspects of professional development. Including a measure of the number of hours of professional development does not alter the strength of the association. Including measures of school leadership quality reduces it by a third but it remains strong (0.25 sd) and comparable in size to the largest association with leadership (0.29 sd).

Table A4. Regression of job satisfaction on mid-career teachers' experiences of professional development.

	All Teachers	Mid-Career Teachers 1	Mid-Career Teachers 2	Mid-Career Teachers 3
Effectiveness of PD	0.076 ** (0.023)	0.103 ** (0.033)	0.100 ** (0.034)	0.107 ** (0.034)
Barriers to PD	−0.362 ** (0.025)	−0.358 ** (0.037)	−0.358 ** (0.037)	−0.254 ** (0.041)
Need for PD	−0.075 ** (0.024)	−0.060 (0.045)	−0.058 (0.046)	−0.085 * (0.045)
Hours of PD	No	No	Yes	Yes
Quality of leadership	No	No	No	Yes
n				

* $p < 0.1$, ** $p < 0.01$.

References

1. See, B.H.; Morris, R.; Gorard, S.; Kokotsaki, D.; Abdi, S. Teacher recruitment and retention: A critical review of international evidence of most promising interventions. *Educ. Sci.* **2020**, *10*, 262. [CrossRef]
2. See, B.H.; Gorard, S. Why don't we have enough teachers? A reconsideration of the available evidence. *Res. Pap. Educ.* **2020**, *35*, 416–442. [CrossRef]
3. Sutcher, L.; Darling-Hammond, L.; Carver-Thomas, D. *A Coming Crisis in Teaching? Teacher Supply, Demand, and Shortages in the U.S.*; Learning Policy Institute: Palo Alto, CA, USA, 2016.
4. Dee, T.S.; Goldhaber, D. *Understanding and Addressing Teacher Shortages in the United States*; Hamilton Project Report 2017-05; The Brookings Institution: Washington, DC, USA, 2017.
5. Fisher, M.H.; Royster, D. Mathematics teachers' support and retention: Using Maslow's hierarchy to understand teachers' needs. *Int. J. Math. Educ. Sci. Technol.* **2016**, *47*, 993–1008. [CrossRef]
6. Weldon, P. Early career teacher attrition in Australia: Evidence, definition, classification and measurement. *Aust. J. Educ.* **2018**, *62*, 61–78. [CrossRef]
7. Mason, S.; Matas, C.P. Teacher attrition and retention research in Australia: Towards a new theoretical framework. *Aust. J. Teach. Educ. Online* **2015**, *40*, 45–66. [CrossRef]
8. Birch, P.; Balcon, M.-P.; Bourgeois, A.; Davydovskaia, O.; Tremosa, S.P. *Teaching Careers in Europe: Access, Progression and Support*; Eurydice Report; Publications Office of the European Union: Luxembourg, 2018.
9. Ingersoll, R.; May, H. Recruitment, Retention and the Minority Teacher Shortage. 2011. Available online: https://www.cpre.org/sites/default/files/researchreport/1221_minorityteachershortagereportrr69septfinal.pdf (accessed on 30 April 2021).
10. Grudnoff, L. All's well? New Zealand beginning teachers' experience of induction provision in their first six months in school. *Prof. Dev. Educ.* **2012**, *38*, 471–485. [CrossRef]
11. Den Brok, P.; Wubbels, T.; Van Tartwijk, J. Exploring beginning teachers' attrition in The Netherlands. *Teach. Teach.* **2017**, *23*, 881–895. [CrossRef]
12. Gallant, A.; Riley, P. Early career teacher attrition in Australia: Inconvenient truths about new public management. *Teach. Teach.* **2017**, *23*, 896–913. [CrossRef]
13. Towers, E.; Maguire, M. Leaving or staying in teaching: A 'Vignette'of an experienced urban teacher 'Leaver'of a London primary school. *Teach. Teach.* **2017**, *23*, 946–960. [CrossRef]
14. Perryman, J.; Calvert, G. What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *Br. J. Educ. Stud.* **2020**, *68*, 3–23. [CrossRef]
15. Tricarico, K.M.; Jacobs, J.; Yendol-Hoppey, D. Reflection on their first five years of teaching: Understanding staying and impact power. *Teach. Teach.* **2015**, *21*, 237–259. [CrossRef]
16. Eros, J. The career cycle and the second stage of teaching: Implications for policy and professional development. *Arts Educ. Policy Rev.* **2011**, *112*, 65–70. [CrossRef]
17. Sibieta, L. *Teacher Shortages in England: Analysis and Pay Options*; Education Policy Institute: London, UK, 2020.

18. Department for Education. Initial Teacher Training: Trainee Number Census 2020 to 2021. Department for Education: London, UK, 2020. Available online: https://www.gov.uk/government/statistics/initial-teacher-training-trainee-number-census-2020-to-2021 (accessed on 15 June 2021).
19. Worth, J.; McLean, D.; National Foundation for Educational Research (NFER) (UK). *The Impact of Covid-19 on Initial Teacher Training*; National Foundation for Educational Research (NFER): Slough, UK, 2020.
20. Department for Education Teacher Recruitment and Retention Strategy. 2019. Available online: https://www.gov.uk/government/publications/teacher-recruitment-and-retention-strategy (accessed on 30 April 2021).
21. Allen, R.; Sims, S. Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools. *Oxf. Rev. Educ.* **2018**, *44*, 441–458. [CrossRef]
22. Donaldson, M.; Johnson, S.M. The price of misassignment: The role of teaching assignments in teach for America teachers' exit from low-income schools and the teaching profession. *Educ. Eval. Policy Anal.* **2010**, *32*, 299–323. [CrossRef]
23. Organisation for Economic Co-operation and Development (OECD). *TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals*; Organisation for Economic Co-operation and Development (OECD): Paris, France, 2020.
24. Donaldson, G. *Teaching Scotland's Future Report of a Review of Teacher Education in Scotland*; The Scottish Government: Edinburgh, UK, 2010.
25. Ronfeldt, M.; Loeb, S.; Wyckoff, J. How teacher turnover harms student achievement. *Am. Educ. Res. J.* **2012**, *50*, 4–36. [CrossRef]
26. Atteberry, A.; Loeb, S.; Wyckoff, J. Teacher churning: Reassignment rates and implications for student achievement. *Educ. Eval. Policy Anal.* **2016**, *20*, 1–28. [CrossRef]
27. Boyd, D.; Grossman, P.; Ing, M.; Lankford, H.; Loeb, S.; Wyckoff, J. The influence of school administrators on teacher retention decisions. *Am. Educ. Res. J.* **2011**, *48*, 303–333. [CrossRef]
28. Coldwell, M. Career orientations and career cultures: Individual and organisational approaches to beginning teachers' careers. *Teach. Teach.* **2016**, *22*, 610–624. [CrossRef]
29. Ovenden-Hope, T.; Blandford, S.; Cain, T.; Maxwell, B. Retain early career teacher retention programme: Evaluating the role of research informed continuing professional development for a high quality, sustainable 21st century teaching profession. *J. Educ. Teach.* **2018**, *44*, 590–607. [CrossRef]
30. Gordon, A.L. Educate-Mentor-Nurture: Improving the transition from initial teacher education to qualified teacher status and beyond. *J. Educ. Teach.* **2020**, *46*, 664–675. [CrossRef]
31. Johnson, B.; Down, B.; Le Cornu, R.; Peters, J.; Sullivan, A.; Pearce, J.; Hunter, J. Promoting early career teacher resilience: A framework for understanding and acting. *Teach. Teach.* **2014**, *20*, 530–546. [CrossRef]
32. Department for Education Early Career Framework. 2019. Available online: https://www.gov.uk/government/publications/early-career-framework (accessed on 30 April 2021).
33. Department for Education National Professional Qualifications (NPQs) Reforms. 2021. Available online: https://www.gov.uk/government/publications/national-professional-qualifications-npqs-reforms/national-professional-qualifications-npqs-reforms (accessed on 30 April 2021).
34. Taylor, M.; Yates, A.; Meyer, L.H.; Kinsella, P. Teacher professional leadership in support of teacher professional development. *Teach. Teach. Educ.* **2011**, *27*, 85–94. [CrossRef]
35. Kirkpatrick, C.L. To invest, coast or idle: Second stage teachers enact their job engagement. In Proceedings of the American Education Research Association Annual Conference, Chicago, IL, USA, 9–13 April 2007.
36. Hartshorne, R.; Baumgartner, E.; Kaplan-Rakowski, R.; Mouza, C.; Ferdig, R.E. Special issue editorial: Preservice and inservice professional development during the COVID-19 pandemic. *J. Technol. Teach. Educ.* **2020**, *28*, 137–147.
37. Organisation for Economic Co-operation and Development (OECD). *Creating Effective Teaching and Learning Environments: First Results from TALIS*; TALIS, OECD Publishing: Paris, France, 2009. [CrossRef]
38. Levac, D.; Colquhoun, H.; O'Brien, K.K. Scoping studies: Advancing the methodology. *Implement. Sci.* **2010**, *5*, 69. [CrossRef] [PubMed]
39. Mays, N.; Roberts, E.; Popay, J. Synthesising research evidence. In *Methods for Studying the Delivery and Organisation of Health Services*; Fulop, N., Allen, P., Clarke, A., Black, N., Eds.; Routledge: London, UK, 2001.
40. Sharma, U.; Rangarajan, R. Teaching students with autism spectrum disorders in South Asia: A scoping study and recommendations for future. *Int. J. Dev. Disabil.* **2019**, *65*, 347–358. [CrossRef]
41. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
42. Munn, Z.; Peters, M.D.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **2018**, *18*, 1–7. [CrossRef]
43. Spilsbury, K.; Pender, S.; Bloor, K.; Borthwick, R.; Atkin, K.; McCaughan, D.; Watt, I.; Adderley, U.; Wakefield, A.; McKenna, H. Support matters: A mixed methods scoping study on the use of assistant staff in the delivery of community nursing services in England. *Health Serv. Deliv. Res.* **2013**, *1*, 1–146. [CrossRef] [PubMed]
44. Schofield, B.; Voss, S.; Proctor, A.; Benger, J.; Coates, D.; Kirby, K.; Purdy, S.; Booker, M. Exploring how paramedics are deployed in general practice and the perceived benefits and drawbacks: A mixed-methods scoping study. *BJGP Open* **2020**, *4*. [CrossRef] [PubMed]

45. Anderson, S.; Allen, P.; Peckham, S.; Goodwin, N. Asking the right questions: Scoping studies in the commissioning of research on the organisation and delivery of health services. *Health Res. Policy Syst.* **2008**, *6*, 7. [CrossRef]
46. Greenhalgh, T.; Robert, G.; Macfarlane, F.; Bate, P.; Kyriakidou, O.; Peacock, R. Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. *Soc. Sci. Med.* **2005**, *61*, 417–430. [CrossRef]
47. Allen, R.; Benhenda, A.; Jerrim, J.; Sims, S. New evidence on teachers' working hours in England. An empirical analysis of four datasets. *Res. Pap. Educ.* **2020**, 1–25. [CrossRef]
48. Jerrim, J.; Sims, S. *The Teaching and Learning International Survey (TALIS) 2018*; UCL Institute of Education: London, UK, 2019.
49. Burns, D.; Darling-Hammond, L. *Teaching Around the World: What Can TALIS Tell Us*; Stanford Center for Opportunity Policy in Education: Stanford, CA, USA, 2014; Available online: https://edpolicy.stanfor.edu/sites/default/files/publications/teaching-around-world-what-can-talis-tell-us_3.pdf (accessed on 15 June 2021).
50. Freeman, C.; O'Malley, K.; Eveleigh, F. *Australian Teachers and the Learning Environment: An Analysis of Teacher Response to TALIS 2013*; Australian Council for Educational Research: Melbourne, Australia, 2014.
51. Jenkins, A.; Ueno, A. Classroom disciplinary climate in secondary schools in England: What is the real picture? *Br. Educ. Res. J.* **2017**, *43*, 124–150. [CrossRef]
52. Organisation for Economic Co-operation and Development (OECD). *TALIS 2018 Technical Report*; Organisation for Economic Co-operation and Development (OECD): Paris, France, 2019; Available online: https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf (accessed on 15 June 2021).
53. Sims, S.; Jerrim, J. *TALIS 2018: Teacher Working Conditions, Turnover and Attrition*; Department for Education: London, UK, 2020; p. 59.
54. Department for Education. *School Workforce in England: November 2019*; Department for Education: London, UK, 2020.
55. Huberman, M. The professional life cycle of teachers. *Teach. Coll. Rec.* **1989**, *91*, 31–57.
56. Dreyfus, S.E. The five-stage model of adult skill acquisition. *Bull. Sci. Technol. Soc.* **2004**, *24*, 177–181. [CrossRef]
57. Day, C.; Stobart, G.; Sammons, P.; Kington, A. Variations in the work and lives of teachers: Relative and relational effectiveness. *Teach. Teach.* **2006**, *12*, 169–192. [CrossRef]
58. Day, C. New lives of teachers. *Teach. Educ. Q.* **2012**, *39*, 7–26.
59. Day, C.; Gu, Q. Variations in the conditions for teachers' professional learning and development: Sustaining commitment and effectiveness over a career. *Oxf. Rev. Educ.* **2007**, *33*, 423–443. [CrossRef]
60. Day, C.; Gu, Q. *The New Lives of Teachers*; Routledge: London, UK, 2010.
61. Borman, G.D.; Dowling, N.M. Teacher attrition and retention: A meta-analytic and narrative review of the research. *Rev. Educ. Res.* **2008**, *78*, 367–409. [CrossRef]
62. Bressman, S.; Winter, J.S.; Efron, S.E. Next generation mentoring: Supporting teachers beyond induction. *Teach. Teach. Educ.* **2018**, *73*, 162–170. [CrossRef]
63. Fullan, M.G. Why teachers must become change agents. *Educ. Leadersh.* **1993**, *50*, 12.
64. Fessler, R.; Christensen, J. *The Teacher Career Cycle: Understanding and Guiding the Professional Development of Teachers*; Allyn and Bacon: Boston, MA, USA, 1992.
65. Dall'Alba, G.; Sandberg, J. Unveiling professional development: A critical review of stage models. *Rev. Educ. Res.* **2006**, *76*, 383–412. [CrossRef]
66. Maandag, D.; Helms-Lorenz, M.; Lugthart, E.; Verkade, A.T.; Van Veen, K. *Features of Effective Professional Development Interventions in Different Stages of Teachers' Careers*; University of Groningen: Groningen, The Netherlands, 2017.
67. Berliner, D.C. Expert teachers: Their characteristics, development and accomplishments. In *De La Teoria . . . a L'aula: Formacio del Professorat Ensenyament de las Ciències Socials*; Batllori i Obiols, R., Gomez Martinez, A.E., Oller i Freixa, M., Pages i Blanch, J., Eds.; Universitat Autònoma de Barcelona: Barcelona, Spain, 2004; pp. 13–28.
68. Day, C.; Sammons, P.; Stobart, G. *Teachers Matter: Connecting Work, Lives and Effectiveness*; McGraw-Hill Education: London, UK, 2007.
69. Steffy, B.E.; Wolfe, M.P. A life-cycle model for career teachers. *Kappa Delta Pi Rec.* **2001**, *38*, 16–19. [CrossRef]
70. Avidov-Ungar, O.; Herscu, O. Formal professional development as perceived by teachers in different professional life periods. *Prof. Dev. Educ.* **2020**, *46*, 833–844. [CrossRef]
71. Mertler, C.A. Classroom-Based action research: Revisiting the process as customizable and meaningful professional development for educators. *J. Pedagog. Dev.* **2013**, *3*, 39–43.
72. Klassen, R.M.; Chiu, M.M. Effects on teachers' self-efficacy and job satisfaction: Teacher gender. *Years Exp. Job Stress J. Educ. Psychol.* **2010**, *102*, 741–756.
73. Kyndt, E.; Gijbels, D.; Grosemans, I.; Donche, V. Teachers' everyday professional development: Mapping informal learning activities, antecedents, and learning outcomes. *Rev. Educ. Res.* **2016**, *86*, 1111–1115. [CrossRef]
74. Schwille, S.A. *Never Give Up: An Experienced Teacher Overcomes Obstacles to Change*; IAP: Charlotte, NC, USA, 2016.
75. Maurer, T.J.; Tarulli, B.A. Investigation of Perceived environment, perceived outcome, and person variables in relationship to voluntary development activity by employees. *J. Appl. Psychol.* **1994**, *79*, 3. [CrossRef]
76. Richter, D.; Kunter, M.; Klusmann, U.; Lüdtke, O.; Baumert, J. Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teach. Teach. Educ.* **2011**, *27*, 116–126. [CrossRef]
77. Brown, S. Part-Time Women Teachers and Their Career Progression: A Life History Approach. Ph.D. Thesis, Sheffield Hallam University, Sheffield, UK, 2019.

78. Louws, M.L.; Meirink, J.A.; van Veen, K.; van Driel, J.H. Understanding teachers' professional learning goals from their current professional concerns. *Teach. Teach.* **2018**, *24*, 63–80. [CrossRef]
79. Louws, M.L.; Meirink, J.A.; van Veen, K.; van Driel, J.H. Teachers' self-directed learning and teaching experience: What, how, and why teachers want to learn. *Teach. Teach. Educ.* **2017**, *66*, 171–183. [CrossRef]
80. Donaldson, M.L.; Johnson, S.M.; Kirkpatrick, C.L.; Marinell, W.; Steele, J.L.; Szczesiul, S.A. Angling for access, bartering for change: How second-stage teachers experience differentiated roles in schools. *Teach. Coll. Rec.* **2008**, *110*, 1088–1114.
81. Mertler, C.A. Should I stay or should I go? Understanding teacher motivation, job satisfaction, and perceptions of retention among Arizona teachers. *Int. Res. High. Educ.* **2016**, *1*, 34–45. [CrossRef]
82. Department for Education; Michelle Donelan, M.P. Opportunity Areas Programme to Support Young People Hit Hardest by Pandemic. Available online: https://www.gov.uk/government/news/opportunity-areas-programme-to-support-young-people-hit-hardest-by-pandemic (accessed on 25 April 2021).
83. Chartered College of Teaching Chartered Status. Available online: https://chartered.college/chartered-teacher/ (accessed on 25 April 2021).
84. Worth, J.; Bamford, S.; Durbin, B. *Should I Stay or Should I Go? NFER Analysis of Teachers Joining and Leaving the Profession*; NFER: Slough, UK, 2015.
85. Sims, S. *TALIS 2013: Working Conditions, Teacher Job Satisfaction and Retention*; Department for Education: London, UK, 2017.
86. Coldwell, M. Exploring the influence of professional development on teacher careers: Developing a path model approach. *Teach. Teach. Educ.* **2017**, *61*, 189–198. [CrossRef]
87. Speck, D. Revealed: The 8 Lead Schools for Teacher Flexi Working. Available online: https://www.tes.com/news/revealed-8-lead-schools-teacher-flexible-working-part-time (accessed on 26 April 2021).

education sciences

MDPI

Article

Training Teachers in China to Use the Philosophy for Children Approach and Its Impact on Critical Thinking Skills: A Pilot Study

Caiwei Wu

School of Education, Durham University, Leazes Road, Durham DH1 1TA, UK; caiwei.wu@durham.ac.uk

Abstract: Philosophy for Children (P4C) is an educational approach that helps children question, reason, construct arguments, and collaborate with others. This approach to teaching is new to Chinese teachers and students who have traditionally relied on rote learning and dissemination of knowledge. Independent thinking and questioning are rarely encouraged. This article reports on a pilot study aimed at training teachers in one school in mainland China to use P4C to promote thinking skills. Six year 7 classes (age 12–13) and their teachers were randomly assigned to receive P4C training ($n = 90$ pupils) or to a control group ($n = 88$). The intervention ran for 4 weeks. The study found that teachers appreciated the P4C methods but were concerned about using the method in their regular curriculum. An impact evaluation shows that students who were taught P4C experienced a small improvement in thinking skills, measured using a composite of validated critical thinking tests.

Keywords: philosophy for children; thinking skills; teaching pedagogy development

Citation: Wu, C. Training Teachers in China to Use the Philosophy for Children Approach and Its Impact on Critical Thinking Skills: A Pilot Study. *Educ. Sci.* **2021**, *11*, 206. https://doi.org/10.3390/educsci11050206

Academic Editors: Beng Huat See and Rebecca Morris

Received: 8 February 2021
Accepted: 20 April 2021
Published: 27 April 2021

1. Introduction

The promotion of higher-order thinking skills is not new but has been increasingly highlighted as skills necessary for the 21st century, especially with the rapid growth and dissemination of information globally, not all of which is trustworthy [1]. Following the dramatic changes that have taken place in Chinese society, the target of education has shifted to all-round development of citizens. To be competitive globally in business, education, science, and technology, China has emphasised the promotion of critical thinking rather than simply imparting knowledge and fulfilling exam requirements.

In 2010, the Chinese government released the 'The National Guidelines for Medium and Long-Term Educational Reform and Development (2010–2020)', which emphasised skills such as independent inquiry, cooperation, communication, and problem solving as well as fostered cognitive skills [2]. To meet the aims of the reform, professional training was needed to help teachers access new pedagogical approaches and skills.

Although the government has provided some training, researchers have criticised it as inefficient [3,4]. They were very theoretical and consisted largely of hours of listening to government policy and regulations [3,4]. They reported that teachers complained of having to listen to theory-based lectures in large groups [4]. There was little training in practical application of the new approaches, and specific guidance for practical learning activities was lacking [5]. For example, in the training programme, teachers were simply told the importance of thinking skills and why they are beneficial for the development of children's cognitive skills but were not given any practical examples of how to develop thinking skills in students. Therefore, despite the provision of teacher training, implementing new pedagogies is still challenging for many teachers.

An effective teacher development programme should be one where teachers are exposed to the types of activities and lessons that would be performed in the classroom [6]. Similar perspectives also presented that teacher professional training should have high applicability, a practical orientation, participation, and interactivity [7]. It should be a process

that interlaces and complements theory and practice. This was not the case in China; the education reform was introduced without an accompanying effective professional development programme. It is possible that those responsible for the training were ill-equipped to teach the new curriculum, not having been exposed to critical thinking themselves.

Against this background, there is now recognition for a more practical approach to teacher training. One programme that has been considered potentially useful for promoting critical thinking and for developing higher-order thinking skills is Philosophy for Children (P4C). Philosophy for Children (P4C) is an educational approach focused on philosophical enquiry and dialogic teaching. The aim of the intervention is to help children develop their abilities to reason, question, construct arguments, and communicate collaboratively with others [1,8]. It emphasises carefully constructed dialogue between teachers and students, and among students [9]. The philosophical dimension of teacher inquiry depends largely on teachers' knowledge of philosophical issues as well as on pedagogical content knowledge [10]. Training teachers is therefore essential for successful implementation of the programme.

2. Background

2.1. Philosophy for Children

P4C was first developed by the American Philosopher Matthew Lipman in 1972. While Lipman was teaching philosophy at Columbia University, he claimed that the educational system failed to help students become independent thinkers. Under this background, Lipman envisaged programmes for teaching critical thinking and informal logic. Borrowing from Peirce and Dewey, who were initially put forward and broadened the scope of 'Community of Inquiry', Lipman systematically applied the concept of 'community of enquiry' to the educational setting and combined it with teaching thinking skills. He argues that 'classroom is a type of community of inquiry, which leads to questioning, reasoning, connecting, deliberating, challenging, developing problem-solving' [11].

Lipman also wrote novels that included various philosophical topics for students. These materials were based on discussions of philosophical issues in line with Piaget's approach to cognitive development. He hoped that students might raise questions according to the philosophical novels and might explore issues through reasoning, asking for examples, questioning hypotheses, and testing hypotheses.

P4C teaching is not a skill that teachers can easily pick up and deliver. To facilitate philosophical inquiry, teachers need to be trained both in content knowledge and pedagogical skills. This is even more necessary with Chinese teachers whose approaches to teaching are highly teacher-centred and have been primarily one of disseminating knowledge.

This philosophical approach to teaching and learning was supported by the Institute for the Advancement of Philosophy for Children (IAPC), which develops curriculur materials for schools. These materials are designed to engage children in thoughtful discussions regarding the epistemological, ethical, social, and aesthetic dimensions of philosophical experience to support them in making informed choices. According to the IAPC website, a typical P4C session would involve children reading aloud or acting out philosophical stories and ponder on issues that are relevant to their experience. The role of teachers is to facilitate these dialogues, for example, by encouraging students to share their questions and ideas with their classmates. Teachers also model examples of good thinking such as clarifying terms, giving good reasons, offering counter arguments (explanations), drawing inferences, and challenging assumptions. The social aspects of dialogue such as listening to each other and building on each other's ideas are reinforced. Teachers guide the discussion and share their perspectives about the issues [12].

In the UK, the Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE) was established to provide training courses since 1992. The SAPERE's training methods are different from that of IPAC. SAPERE suggests various materials, which include stories, poems, films, images, pictures, and books, rather than asks for specially written philosophical novels. Currently, SAPERE is a leading service provider,

and its sister organisation, named P4C, in China is the only organisation that can provide P4C training in mainland China.

2.2. The P4C Intervention

The impact of P4C on thinking skills and academic outcomes has been evaluated in several studies, most of which were conducted in the US and in the UK. The earliest evaluation by Lipman and colleagues [13] included 20 pupils from two schools in the Montclair District. The matched comparison study reported significant gains in logical reasoning and reading. However, the results are worrying because their research design cannot be considered effective. The sample was small, and the students only received intervention for a short duration [14]. A long-term study that aimed to assess cognitive skills was conducted in 2007: 177 students attended, and they were measured using the CAT test. The results indicated that the treatment students performed better than the control group two years later although the attrition rate was high (35%). In 2015, an independently evaluated large-scale efficacy trial in England [15] found positive effects on primary school children's reading and math. However, the latest published large-scale randomised control trial presented different results [16]. There is no statistical evidence to provide evidence that the treatment group performed better than the control group. Using questionnaires, they supported the conclusion that P4C can promote classroom engagement, improve the level of respect for others' opinions, and the ability to express views clearly.

In 2004, a systematic review that included ten studies was conducted. Eight studies reported a positive impact on cognitive and noncognitive skills [1]. However, it received some criticism that the employed studies were small scale and used different measurements. A more recent meta-analysis was conducted in 2018 and employed ten studies [17]. The findings showed a moderate, positive influence on cognitive outcomes and an apparent positive effect on reasoning skills. In summary, most of the research results in Western countries that use P4C as an intervention are positive, although they focused on different abilities.

P4C is not new in China. In 1997, the earliest P4C practice was developed in Yunan Province. In this project, teachers' feedback on P4C was quite positive. Assessed via a questionnaire, 89% of teachers believed that P4C could promote students' thinking and 83% of teachers believed that their dialogic skills improved. However, there were no standardised test results to prove whether students made progress [18]. In the last 10 years, the number of studies on P4C in Chinese educational contexts has increased. However, most of them have made theoretical contributions, for example, to explore the possibility of combining Confucian dialogue and the community of inquiry [19] and to analyse what types of themes and topics are suitable for the Chinese classroom [20]. In addition, some empirical studies were developed in Hong Kong and Taiwan. A study involving 28 first-year secondary students in Hong Kong showed that P4C students made bigger gains on the New Jersey Test of Reasoning Skills, with an increase of 27.9% from the pretest compared to the control group, which achieved only 13.3% improvement [21]. The treatment group was taught by the researcher, who had attended a 10-day training workshop run by IAPC. The control group was taught by an English language teacher at the same school. Since there were only two teachers, it is not possible to attribute the results to the intervention alone. The second study was a quasi-experimental study conducted in Taiwan, which randomly assigned two classes of secondary school students to P4C or not ($N = 62$) [22]. P4C was delivered as part of an after-school activity. The results showed that the control group made bigger gains between pre- and posttest than P4C students. Since students were randomised by classes ($N = 2$), the study cannot rule out class and teaching effects. The impact of P4C on Chinese students is therefore inconclusive.

2.3. P4C Training

With education and teaching pedagogy reform, we cannot forget to train the teachers with the new teaching skills.

A medium-scale empirical research was conducted in the UK [23] and presented the complete training process, which included how they introduced the community of enquiry, what exercises were demonstrated to improve dialogue, and what books and resources they suggested that the teachers read. Similarly, a study related to developing dialogic teaching was presented in 2018 [24]. In addition to basic theoretical training, the follow-up training service in that project showed potential as a model of learning. After the teachers received theoretical training, the training team also provided mentors to each school. They shared real video records to help teachers find similar problems within their class. They also provided guided planning, target-setting, and review. SAPERE has conducted two large-scale P4C research studies in the UK since 2015 [14,15]. They run an independent training course system that can meet the needs of different levels of difficulty. For example, in the Level 1 foundation course, they introduce the theory, provide tools to develop questioning and thinking skills, connect P4C to the curriculum, demonstrate P4C in practice, and present resources and sources for teaching materials. In the Level 2A course, they focus on facilitating P4C enquiries with more flexibility and reflection.

P4C was introduced in China nearly 30 years ago. However, most Chinese teachers do not have P4C teaching experience. The earliest P4C practice provided seven-day training to practitioners. [18]. However, the authors discussed neither what methods they used nor what types of activities were part of the training. They only mentioned general views about P4C using a questionnaire. For example, 'Is P4C suitable for our country?' and 'Can a community of inquiry improve the efficiency of teaching and learning?' Their work makes a poor reference for future research in P4C professional development. In small- and medium-scale studies, researchers are more likely to employ experienced P4C teachers rather than to train teachers who have never experienced P4C before [21,22,25,26]. Therefore, P4C professional development is a relatively weak part of the academic research field in China [27].

In summary, previous literature provides evidence that P4C has a positive impact on various abilities, especially reasoning skills. However, most of them were conducted within a Western educational background. Although researchers who focused on Chinese content have contributed to the promotion of P4C theory, the number of empirical studies is limited and there are few studies on teachers' professional development in this field.

3. Aims and Objectives

The aims of this pilot are as follows:

- Test the feasibility of delivering P4C lessons in a regular Chinese school;
- Evaluate the professional development model developed for training of the teachers in the delivery of P4C;
- Test the training resources (i.e., lesson plans) developed by the researcher;
- Test the measurement tools (i.e., the critical thinking test) to see how long it will take and if the questions are appropriate for the educational level of the students;
- Identify potential barriers/challenges in the staff training and classroom delivery of P4C.

4. Methodology

4.1. Trial Design

This pilot study was conducted in one state school in Fushun, Liaoning Province, in northeast China. This is an average region in terms of economic and educational development. The study design was a two-group pretest posttest randomised controlled trial. Randomisation was at the class/teacher level. Random assignment of the participants ensured that any confounding variables were equally distributed between the two groups. To avoid bias, randomisation was blinded to the participants (teachers and students) and

group allocation was revealed only after the pretest. There is research evidence suggesting that knowledge of group assignment could influence the way students perform on the test [28].

The purpose of the pretest was to establish baseline equivalence. The pretest was administered before randomisation and before the first session was delivered. A posttest was taken after the last P4C lesson. The primary outcome was students' critical thinking skills. Impact was estimated using the gain scores between the pretest and posttest, expressed as Hedges' g effect size. This is the difference in the mean gain scores between the two groups divided by the pooled standard deviation.

In addition, a process evaluation was carried out to monitor fidelity to treatment and to identify challenges or barriers faced by teachers in the implementation of P4C as well as to collect feedback on the teacher training. The process evaluation was conducted using classroom observations and interviews with the teachers.

4.2. Participants

The participants included six year 7 classes for a total of 178 students (aged 12–13 years) and two teachers. The teachers of three classes were randomly assigned to the experimental group ($N = 90$; 1 teacher), and the other three classes were signed to the control group ($N = 88$; 1 teacher), in which they continued regular lessons. The students in the experimental group received seven P4C sessions during the four weeks of the study.

Year 7 classes were chosen because the curriculum pressure was relatively small compared to that of year 8 and year 9 classes. The school was secular, and the school population was ethnically homogeneous (all were Chinese). The students were predominantly (more than 90%) from middle- or working-class families.

It is not standard practice to ask young children (age 11–12) for consent in this kind of research. Consents, where required, are normally sought from parents. The lessons were delivered as part of the school curriculum where consent from the school leaders have already been sought. No personal sensitive data was collected in this study, and consent from parents was mainly to ask for permission to use pupil test data.

4.3. Procedure

Before the trial, the pilot teacher attended two days of P4C training (Level 1) in Shanghai. The training was provided by *P4C in China*, which is SAPERE (UK)'s sister organisation in China and uses standardised SAPERE methods.

The trial started on 1 September 2019. The P4C lessons, each taking 40 min, were taught twice a week for four weeks. The lessons were taught during English class. A total of seven lessons were observed. They were recorded by handwritten field notes and video. In this way, not only can the researcher record key segments but also the video can be helpful when reviewing the dialogue. The P4C session proceeded as follows: the teacher first presented a stimulus. The stimulus could be a story, a picture, or a video. Here, we used the story of Cinderella as an example. Then, students were given individual thinking time and they were allowed to take turns sharing with their peers what they found interesting about the stimulus. The teacher helped and guided students in raising philosophical questions. After that, they chose a question that was supported by the majority and discussed it as a group. The question could be 'Should parents ask their children to do housework?' This is not a factual question. Students were allowed to present different opinions. They were encouraged to think hard about what reasons or examples could be used to support their ideas and to critique others. In contrast to a traditional Chinese classroom, the teacher worked as a facilitator rather than as an authority.

Students in the control group were taught by another teacher at the same school. To avoid contamination, the control group teacher did not teach any topics related to philosophy and thinking skills. Instead, the teacher taught the normal curriculur content using traditional methods.

Finally, both groups were administered the posttest on 29 September 2019. The teacher and students in the experimental group were interviewed informally at the end of the trial to obtain feedback regarding their experience with the P4C lessons and the challenges faced as well as to identify optimal ways to overcome these challenges.

5. Teacher Training

Prior to the trial, the teachers were trained in the delivery of P4C lessons. The training was provided by P4C in China (Shanghai), which is SAPERE (UK)'s sister organisation in China. SAPERE (Society for the Advancement of Philosophical Enquiry and Reflection in Education) is the UK charity supporting P4C. Its mission is to train teachers in P4C. P4C in China (Shanghai) also provides standardised SAPERE training methods. As mentioned above, traditional teacher training provided by the government is theoretical and lacks practical guidance. For this study, therefore, more detailed guidance was provided to the teacher. In addition to imparting theoretical knowledge, the trainer demonstrated the entire process of the P4C programme several times, allowing the teacher to experience how the new approach works in a real classroom situation.

The Level 1 P4C course lasted two days. Eighteen participants from different provinces attended the course. Twelve of them were kindergarten teachers, and the rest of them were primary and secondary school teachers. The training sessions ran from 9:00 to 12:00 and from 13:00 to 16:00 each day. The training included the following elements:

- Introduction of P4C, key principles and methods;
- Demonstration of P4C lessons and key practice;
- Sharing of available resources and teaching material;
- Advice and support.

5.1. Principles of Philosophy for Children

Generally, a P4C class has a relatively standard structure. There are five main steps in a P4C session:

- **Start the inquiry**: students and teachers sit in a circle so that everyone can easily hear each other.
- **Share a stimulus to prompt inquiry**: the American educational reformer John Dewey believed that all enquiry begins with a problematic situation. Therefore, presenting a challenging stimulus is the basis for a successful P4C class. Stimuli may include stories, poems, picture books, videos, and news articles. The purpose of the stimulus is to introduce the topic and to generate interest.
- **Encourage students to think**: this step allows time for individual thinking and public reflection. After a couple of minutes, the members of the discussion group are encouraged to share their responses to the stimulus.
- **Question and discussion**: after the students share their ideas about the stimulus, a philosophical question that can be discussed during the remainder of the session is put forward. During this time, the students listen to others; question their peers; and present their arguments with reasons, evidence, and examples. The teachers should play a supportive role. They can employ a Socratic questioning model to promote deeper discussion and to guide students' thinking.
- **Have the students evaluate, build ideas, and summarise**: this step allows students to express their final thoughts. Ideally, everyone provides an evaluation and summary of the discussion.

5.2. The Key Practice

To help teachers have a better understanding of how P4C works in the classroom, the trainer demonstrated and modelled a P4C lesson focused on creating questions and developing discussion.

5.2.1. The Stimulus

At the beginning, the participants were trained to prepare the stimuli. Picking out a stimulus with significant concepts helps encourage pupils to lead the discussion to deeper levels. During the training, the participants prepared an image as the stimulus. They shared their thoughts about what interests them or puzzles them, or what they find significant. By comparing participants' evaluation and feedback of different images, the teachers can find out what types of stimuli are suitable for the P4C class. The trainer also gave some suggestions that could be used for stimuli such as the following:

- *Literature: love, democracy, fairness, justice*
- *Humanities: justice, truth*
- *Citizenship: rights, duties, freedom, welfare*

5.2.2. Questioning and Dialogue

Unlike the traditional Chinese pedagogy of cramming and memorisation, P4C encourages students to think about questions on their own rather than to receive answers directly from their teacher. To achieve this goal, the students were given enough space and time. Question-creation may include talking to partners, writing questions down, and displaying questions.

In P4C lessons, dialogue is a key element of the programme. It promotes deeper engagement and a level of understanding. In a P4C class, the dialogue generally includes *asking questions, giving examples and evidence, discussing and critiquing, summarising, and evaluating*. To improve the quality of the dialogue, the Socratic method of questioning was suggested. For example, '*Do you mean . . .'* to clarify their questions and '*Do you think your question is similar or different from others*' to help summarise. The discussion could focus on providing alternative points of view, giving examples, examining reasons, and establishing logical statements. Some cards with Socratic questions were provided in the training. It could be used for both teachers and students during the discussion. Here are some examples:

Information process questions

- *Could you explain what you mean?*
- *Can someone give an example?*

Reasoning Questions (expanding and probing)

- *What are your reasons for saying that?*
- *Do we have any evidence?*

Enquiry Questions

- *So you agree/disagree with . . . ?*
- *Is that always the case or only sometimes?*

Creative Thinking Questions

- *What if. . .*
- *Does.. imply. . . ?*

Evaluation Questions

- *Have we reached any conclusion?*
- *Can anyone summarise what we have said so for?*

5.3. Sharing of Available Resources and Teaching Material

In this pilot study, the participants were given the opportunity to practice the steps and the questioning techniques. They were also provided instructional materials as a theoretical supplement. These were available in print and as online videos. The following were included:

- *A handbook provided by P4C in China that includes the aims, process, details, and P4C theoretical knowledge;*

- *A booklet containing lesson plans and transcripts of lesson extracts;*
- *Video material: videos of recorded P4C classes in a Taiwan secondary school of P4C lessons in real classroom situations from which the teachers can learn and review real classroom practices.*

5.4. In-School Support

In addition to the introductory P4C training, additional monitoring support was offered. First, the researcher helped the pilot teacher prepare lesson plans. The philosophical topics and stimuli were selected in consultation with the teacher. The question enquiry and discussion parts were first designed by us, with additional materials supplemented by the teacher. The researcher then observed the delivery of the P4C session each week and provided feedback on the lesson. The pilot teacher was encouraged to reflect upon each lesson.

In summary, the training course introduced the theory and practice of P4C, provided tools to develop questioning and thinking skills, demonstrated philosophical enquiry in practice, and presented P4C resources for compiling teaching materials. The training provided the teacher with the skills, knowledge, tools, and resources to teach P4C.

6. Data Collection

The following three instruments were used to collect data for this study.

6.1. Thinking Skills Test

The primary objective was to improve students' critical thinking skills. The choice of an appropriate test was a challenge as the test had to be validated and appropriate for the language ability of first-year secondary school students for whom English was not their first language. An English–Chinese edition was provided for students to help them understand the questions.

Their critical thinking skills were measured using a bespoke test made up of components from the three commonly used tests for critical thinking: CAT, the Cornell Test Level 1 [29], and the Waston Glasser test [30]. All of the tests have limitations, however. They are either too long, too expensive, or not age-appropriate. As a compromise, the researcher selected appropriate questions from these existing tests and developed a test that is suitable for the students in the study. It includes tests on inferences, assumptions, deduction, interpretation, arguments, verbal and nonverbal reasoning, spatial reasoning, quantitative reasoning, analogical reasoning, and inductive reasoning. The tests included 12 questions to be completed within 30 min. All questions were in multiple-choice format. An example of the test questions is presented below (see the pretest questions in Appendix A). As the test is a test of critical thinking and not language skills, the researcher prepared the questions in both Chinese and English.

6.2. Classroom Observation

In addition to collecting data from the pre- and posttests, all P4C lessons were observed. The observations were recorded as field notes by the researcher. For future training and for feedback to the teacher, all sessions were also audio-recorded. It mainly aims to document the interaction among students and the teacher during discussions in the classroom. The level of student engagement in terms of asking and answering questions is noted. It helps to determine whether students demonstrated critical thinking skills after exposure to P4C and teachers' use of the P4C protocol.

6.3. Informal Interview

As part of the process evaluation, nearly ten students and the pilot teachers were asked for their opinions of P4C lessons. This included questions about whether they enjoy the sessions, which part of P4C they liked or disliked, the quality of dialogue between peers and with the teacher, and what challenges they faced. The interviews were unstructured

and informal. The purpose of the informal conversation was to gather a general overview of the pupil's attitudes towards P4C.

7. Findings

7.1. Findings from the Impact Evaluation

In order to answer the research questions, it is necessary to evaluate the impact of P4C on students' critical thinking between the experimental and control groups. A significance test and its variance (e.g., confidence intervals and *p*-value) are not appropriate in this study. What the *p*-value in significance tests shows is how likely it is to obtain our results, assuming that there is no difference between groups. It does not tell whether within the given results there is a difference between the groups. Therefore, Hedges' g effect size was used to estimate the gain scores between the pretest and posttest. This is the difference in the mean gain scores between the two groups divided by the pooled standard deviation. The formula is *Effect size = (Mean treatment gain score − mean control gain score)/pooled standard deviation (average SD).*

At the beginning, the experimental group had a slightly higher pretest score (M = 5.27) than the control group (M = 5.24). After one month of intervention, the posttest mean score of both groups improved. The experimental group showed a greater percentage increase (5%) than the control group (4%). This indicates that students who received P4C intervention made greater progress in the thinking tests than those who did not.

The overall results show that students who were taught P4C made slightly bigger gains than those in the control group who received the regular lessons (Tables 1–3). This suggests that P4C has a small positive impact on students' critical thinking skills after only one month of intervention (effect size +0.03).

Table 1. Descriptive statistics for pretest scores.

Group	Pre-Score Mean	SD	ES
Experimental group (N = 86)	5.27	1.29	
Control (N = 87)	5.24	1.34	
Overall (N = 173)	5.25	1.35	0.02

Table 2. Descriptive statistics for posttest scores.

Group	Post-Score Mean	SD	ES
Experimental group (N = 86)	5.57	1.47	
Control (N = 87)	5.49	1.31	
Overall (N = 173)	5.53	1.39	0.05

Table 3. Overall gain score.

Group	Gain Score	SD	ES
Experimental group (N = 86)	0.31	1.59	
Control (N = 87)	0.25	1.85	
Overall (N = 173)	0.28	1.72	±0.03

7.2. Classroom Observation

In addition to using standardised tests to examine the role played by P4C in developing students' thinking skills, classroom observations of theP4C lessons were used to determine the changes in classroom dialogue. In the four weeks of the intervention, it was observed that students' engagement in the classroom and the willingness to express themselves increased. The teacher gradually handed over control of the class discussions to the students. Students talked more, and the lessons became more student-oriented and less teacher-focused.

Chinese students are typically quiet and passive learners. For most of the time in class, they are more likely to receive knowledge from the teacher as an authoritative figure than to contribute to the discussions. Students' voices are rarely heard. The role of the teacher is to disseminate information, and that of the students is to receive the information. They are not encouraged to question the information imparted upon them. The traditional teacher-centred pedagogical style in China provides limited opportunities for students to express themselves. In this study, when the teacher adopted the role of a facilitator rather than as a figure of authority, the students initially found it difficult to engage. As the lessons progressed, they became more willing to participate and to contribute their ideas.

To demonstrate the progression, we selected two short dialogues between the teacher and students, one from the first week and one from the second week of the intervention. They show how the teacher was better able to probe and ask open-ended questions to get the discussion going. The students similarly show their willingness to provide more expanded answers. Other students also joined in the discussion, and they were keen to provide a reason for their responses. This was something that was never explicitly encouraged.

Week 1
Teacher: *Why do you study?*
Student: *Because I want to go to university.*
Teacher: *Is this the only reason you want to study?*
Student: *Erm. . . [keeps quiet]*

Week 2
Teacher: *Can you enjoy learning as you do gaming?*
Student A: No!
Teacher: *Why not?*
Student A: *I don't know. But I know I love playing computer games but I don't like studying.*
Teacher: *Why do you like playing games?*
Student A: *I get satisfaction from the game, especially when I win or complete a difficult level.*
Teacher: *Why can't you get satisfaction from learning?*
Student B: *I can. I love maths—it challenges me—I feel satisfaction when I solve difficult questions.*
Student A: *I don't agree with you. You're good at maths. . . I work hard but still can't perform well in maths. But I can win the game if I try a few more times.*

At the beginning, the students felt confused and unwilling to talk. During the second week, however, their answers became longer than in the first week. In addition, the conversation was not only between the teacher and a single student: other students also participated in the discussion. During the first two weeks, however, there was not much evidence of higher-order thinking. Changes were also observed in the teacher's questioning techniques in the third and fourth weeks, a dialogue of which is shown below:

Teacher: *Why do you think Tony and Bill are friends? Can you give me some reasons?*
Student A: *Because they play together.*
Teacher: *Do you mean people who play together are friends?*
Student A: *Erm. . . no.*
Teacher: *Oh? How do you define 'friends'?*
Student A: *I don't know.*

This was an example during the fourth week when the teacher applied Socratic questioning more frequently than before. The questions she put forward were more effective in facilitating thinking. For example, 'Do you mean . . . [repeat student's point]?' helps students find contradictory points in their answers. Asking students to give a definition is also a good way to practise summarising. In the discussion, the students also displayed their thinking processes, such as identifying contradictions, giving counter examples, and summarising.

Student B: *No, friends can play together, but I can't be sure that people who play together are definitely friends.*
Student C: *I agree. I don't think Tony and Bill are friends.*
Teacher: *Why?*
Student C: *Although Tony plays with Bill, Tony is unhappy.*
Teacher: *Why is Tony unhappy?*
Student C: *Because Bill broke Tony's new bike and his toys—he doesn't respect Tony.*

Similarly, during this example, when the students were challenged to give deeper answers, the teacher asked 'Why?' twice rather than giving the answers directly. In a traditional Chinese classroom, students expect the teacher to give them the answer, and the teacher is often quick to offer the answer as they think that this is the role of the teacher. In the four weeks of intervention, the teacher was observed to consciously become more of a facilitator. She tried to reduce her control of the classroom and gave students more opportunities to think independently.

Overall, throughout the four weeks of P4C intervention, the teacher and students made progress. The students changed from passive receptors of knowledge to active participants in classroom discussions. They started to engage in more logical and reasonable dialogue rather than simply providing 'yes' or 'no' answers. Moreover, the teacher was more likely to apply Socratic questioning techniques and to put forward high-quality questions as a facilitator in the class.

7.3. Interview Findings

In addition to collecting data from the thinking test results and classroom observation, informal interviews were conducted to capture both the teacher's and students' general feedback towards the P4C intervention. The primary target of the pilot study was to test the feasibility of the training model and to identify potential barriers that may challenge our study in the future. Therefore, the interview questions that were prepared for the teacher revolved around the following themes: What is your impression of the training? What do you like or dislike about P4C? Are there any challenges to applying P4C in your classroom? The students were asked "Do you like P4C lessons?"

Firstly, the pilot study teacher expressed her appreciation of the P4C training as follows:

I think the training is good. It not only introduced the principles of the P4C approach, but a large part of the training included classroom simulations. The trainer worked as a facilitator modelling the delivery of P4C, while I acted as a student. This allowed me to see how the P4C pedagogy works in practice. It also helped me to think about the actual situation and the challenges we may encounter in the classroom.

However, the teacher was concerned about how to implement P4C into regular Chinese classes and whether the P4C pedagogy can be generalised and incorporated into other subjects within the curriculum. Although the training provided a few ideas on how to go about doing this, the teacher commented that: *I used a specific time each Friday to deliver P4C in place of the regular lessons, but after the project was completed, I would have to revert to the national curriculum and its lesson plans. Thus, it may be challenging for us to integrate the P4C pedagogy into the regular curriculum content.*

When the researcher asked the teacher what her impression of the P4C pedagogy was, she provided positive feedback: *Within the span of one month, I think the most positive impact was classroom engagement. I can see that the students' participation has greatly increased during this time. There was a clear difference between their active performance in my P4C class and their tendency to remain silent in the past. And I think that is a great beginning for facilitating thinking because if they talk more, they will think more.*

For the student interviews, the researcher randomly selected five students and invited them to share their opinions. Three key words that were prominent among the students' feedback were 'relaxed', 'respectful', and 'learn from others'. Here are some of their comments:

- *I like the P4C class because I have a lot of amazing idea to share with my classmates.*
- *I'm very happy because the topics are interesting in the P4C class. It also makes me relaxed because the teacher does not judge me at the moment.*
- *I progressed a lot. I learned how to present my ideas and theories, and how to use evidence to support them.*
- *I like hearing how other classmates think and talk. I am curious about other people's mind.*

However, while the students enjoyed the new programme and teaching style and appreciated the opportunity to learn from one another and to exchange opinions, some students doubted whether P4C classes could help them reach their academic goals. As one student expressed: *I don't understand what the relationship between the P4C lessons and exam requirements is*.

In summary, both the teacher and students expressed their critical attitudes towards P4C intervention. The teacher agreed that the P4C training was helpful to her teaching skills, but she also hopes to receive guidance on how to integrate it into subject teaching. Moreover, while the P4C approach increased the level of energy and engagement in the classroom because it provided students with opportunities to share their thoughts, the students doubted whether it is ultimately beneficial to their academic performance.

8. Discussion

This pilot study reports an attempt to incorporate P4C in a Chinese secondary school classroom. The outcomes are positive. The pilot teacher was able to complete the P4C lessons, and the students gave positive feedback on the new teaching methods. However, these results should not be accepted without criticism.

8.1. P4C Training

In this pilot study, the P4C organisation in China provided the training, which followed the SAPERE method. Few previous studies revealed the details of their P4C training. Gao [27] also suggested that there should be more research focusing on the professional development of training in P4C. That author introduced their training content and activities. It included how to select stimuli, how to facilitate questioning and dialogue skills, and what materials were provided to the teacher. However, the two-day P4C training is not sufficient for teachers who do not have experience with P4C. It is necessary to provide follow-up training as a supplement.

One possibility is a monthly seminar with a P4C trainer. Professional trainers can help teachers strengthen their theoretical knowledge and can increase the depth of content. Additionally, they can address the problems that teachers face in practice from the previous stage. Another possibility is to invite the experimental group teachers for peer discussions. There should be opportunities and discussion forums for teachers to exchange experiences, ideas, and challenges. They may encounter similar problems in practice. Peer discussions allow for sharing solutions with each other.

8.2. The Application of P4C

The trial lasted for a month. Both the teacher and students gave positive feedback on P4C. The teacher learned a new teaching pedagogy and improved her dialogic skills; the students improved their thinking and expression. However, some barriers were found.

The first barrier is the design of the P4C lesson. So far, there is no P4C textbook that is specially focused on Chinese content. Most materials were translated directly from the textbooks of Lipman and IPAC, which were created based on Western educational background [12,31]. This may lead teachers to think that the content of P4C is not related to Chinese curriculum requirements and is thus not suitable or helpful for Chinese students. In this study, the training provided skills about how to design module and materials provided rich P4C topics to the teacher. The researcher and the teacher strove to develop the localization of the P4C programme by choosing topics from the curriculum and designing lesson plans on their own. However, it is not ideal to create the lesson plan independently.

For example, due to traditional teaching habits, the teacher was more likely to choose stimuli relating to factual knowledge than controversial topics. From the feedback of the interview, students agreed that the modules are helpful for their talking and thinking, while the teacher hoped to be closer to the textbook.

To solve these problems, more materials are needed. On the one hand, it is necessary to provide more lesson plan templates, especially to present what stimuli are appropriate for P4C lessons. On the other hand, the materials need more integration with Chinese curriculur content, Chinese teachers' teaching habits, and the interests of Chinese students.

Another challenge is the application of classroom dialogic skills. Philosophy for Children is a new pedagogy completely opposed to the traditional Chinese authoritative style of teaching. It is not easy for teachers to abandon the role of authority, to change their dialogue habits in the classroom, and to create an open learning environment. In China, thus far, there are few instructions available on how teacher's dialogue skills can be effectively promoted [32]. To improve teachers' dialogue skills, it is insufficient to only learn theory—the knowledge and techniques that one acquires must be effectively put into practice in the classroom.

Therefore, helping teachers set short-term goals and providing targeted training in dialogue skills are necessary. Any reform will need to be instituted in stages. In practice, for example, the first step may be to ask teachers to move slowly from a teacher-centred to a student-centred approach and to then introduce Socratic questioning techniques to encourage students to engage in class. Teachers could then be gradually encouraged to develop high-quality discussions, giving students the opportunity to apply more higher-order thinking skills.

8.3. Limitations and Future Research

Although the results showed a slight improvement in the experimental group, due to the time being too short and the sample being small, our conclusions are not sufficiently convincing.

This pilot trial lasted only four weeks. Evidence from other studies shows that students who received 24 weeks of P4C intervention performed better than those in the 10 week intervention group [23]. Therefore, four weeks may be too short of a period for the impact to be realised.

The sample size was relatively small as well. The trial involved only two teachers. Any differences between the two teachers could therefore have accounted for the difference in results between the experimental and control groups. The results cannot therefore be generalised. For future work, to increase the generalisability of the results, the sample size should be increased and the duration of the intervention should be extended.

Most importantly, we suggest that academic performance should be assessed in future research because exam scores are irreplaceable to China's education system. If these P4C modifications result in greater academic achievement, it is more likely that P4C will be more widely promoted.

9. Conclusions

The results of this study achieved the research aims and objectives posed in the Introduction. This study demonstrated that it is feasible to train and deliver P4C lessons in Chinese classrooms, but to embed P4C in the curriculum is a challenge. Unlike previous studies that focused more on kindergartens and primary schools, this study indicates that students in secondary school are willing to accept this new teaching pedagogy.

In this study, the pilot teacher received training from P4C in China and used standardised SAPRER methods. Moreover, the training content included essential theoretical knowledge, opportunities for practice, and some useful resources. As a result, the pilot teacher positively affirmed the training outcomes.

In terms of testing the measurement tools, the modified English–Chinese version of the Critical Thinking Test was found to be appropriate to the language ability and age of

the students. They were able to understand the meaning of the questions and to complete the test within the specified time. The test results showed that the experimental group made some improvements in their critical thinking compared to the control group.

Funding: This research received no external funding.

Institutional Review Board Statement: No applicable.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Sample Questions in the Pre-Test for Critical Thinking Assessment

Pretest (Pilot Study)

- Class
- Name
- Gender

Which one of the following conclusions is definitely true based on the statement?
1. All birds are animals and all chickens are birds

A. All chickens are animals
B. No chickens are animals
C. Some chickens are animals
D. Some chickens are not animals
E. No valid conclusion

2. No dogs are pets. Some pets are cats

A. All cats are dogs
B. No cats are dogs
C. Some cats are dogs
D. Some cats are not dogs
E. No valid conclusion

3. Xiaofei says he rides a bicycle every day. One day you went to his house and you saw some bicycles with flat tyres in his garden. When you see this, you

A. You know that Xiaofei rides his bicycle every day
B. You do not know if Xiaofei rides a bicycle every day
C. You know that all the bicycles in his garden are Xiaofei's

4. Reading the passage below and answer the questions
4a. A group of explorers had gone to a village, called Nicoma, and disappeared. You took a group of soldiers to find out what had happened to them. You found some stone huts put up by the first group. You went into the first hut and everything was covered by a layer of dust. You called out but nobody answered. One of your members said: 'Maybe they are all dead.' Do you agree with his conclusion?

A. Yes
B. No
C. There is not enough information to decide

4b. You send two of your soldiers to explore the area and check if the water is safe to drink. The soldier A looked at the stream by the village and reported, 'The water looks clear, it is therefore safe to drink.' Soldier B said, 'We can't tell yet if the water is safe to drink'. Which soldier is more believable?

A. Soldier A
B. Soldier B
C. Neither

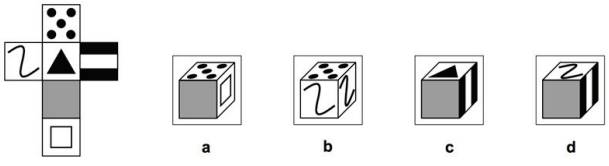
5. Children who go to private schools do better in exams than children who go to public schools. Mimi goes to a public school but her friend Caicai goes to a private school. This means that

A. Mimi will do better than Caicai in exams
B. Caicai will not do well in exams
C. Mimi might do better than Caicai in exams

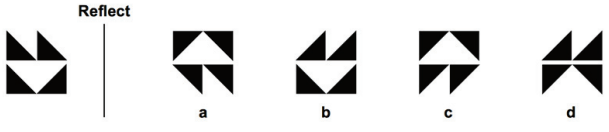
6. These words are similar in some way. Decide how they are the same. Then choose the answer choices that goes with the example word. Hen: Egg

A. Dog: Bark
B. Cow: Milk
C. Peacock: Feathers
D. Swan: White

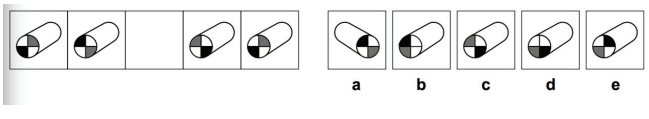
7a. Work out which of the six cubes can be made from the left figure



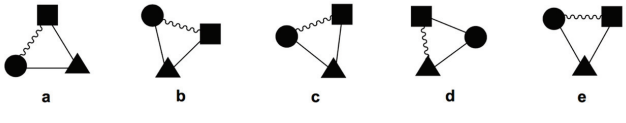
7b. Work out which option would look like the figure on the left if it was reflected over the line



8. In each of these questions, these figures are similar in some way. Decide how they are the same and then choose the figure from the answer choices that goes with them.



9. Find the figure in the row that is most unlike to the other figure.



10. Lili, Wang Lei and John all walk to school each day. Lili leaves home at 8:30 a.m. and takes twenty minutes to reach school. John arrives at school five minutes after Mark, who arrives at school five minutes after Lili. Mark takes 5 min to get to school. Using this information decide which of the following statements is true.

A. Wang Lei leaves home at 8:45

B. Wang Lei leaves home at 8:50
C. Wang Lei arrives at the same time as John
D. LiLi arrives at school last

References

1. Trickey, S.; Topping , K. 'Philosophy for children': A systematic review. *Res. Pap. Educ.* **2004**, *19*, 365–380. [CrossRef]
2. Yeung, S. Is student-centered pedagogy impossible in Hong Kong? The case of inquiry in classrooms. *Asia Pac. Educ. Rev.* **2009**, *10*, 377–386. [CrossRef]
3. Yan, C. 'We can only change in a small way': A study of secondary English teachers' implementation of curriculum reform in China. *J. Educ. Chang.* **2012**, *13*, 431–447. [CrossRef]
4. Jin, H.; Wei, X.; Duan, P.; Guo, Y.; Wang, W. Promoting cognitive and social aspects of inquiry through classroom discourse. *Int. J. Sci. Educ.* **2016**, *38*, 319–343. [CrossRef]
5. Shan, D. Kegai Tuidong Suzhi Jiaoyu (Curriculum reform aimed at promoting Suzhi Jiaoyu). *Kao Shi Zhou Kan* **2002**, *3*, 24–27.
6. Harcombe, E.S. *Science Teaching/Science Learning: Constructivist Learning in Urban Classrooms*; Teachers College Press: New York, NY, USA, 2001; Volume 14.
7. Jovanova-Mitkovska, S. The need of continuous professional teacher development. *Procedia-Soc. Behav. Sci.* **2010**, *2*, 2921–2926. [CrossRef]
8. Vansieleghem, N.; Kennedy, D. What is philosophy for children, what is philosophy with children-After Matthew Lipman? *J. Philos. Educ.* **2011**, *45*, 171–182. [CrossRef]
9. Gorard, S.; Siddiqui, N.; See, B.H. Can 'Philosophy for Children' Improve Primary School Attainment? *J. Philos. Educ.* **2016**, *51*, 5–22. [CrossRef]
10. Gao, Z.Y. Philosophy for Children in China: Teacher Knowledge and Teacher Development. In *Autobiography and Teacher Development in China. Curriculum Studies Worldwide*; Hua, Z., Pinar, W.F., Eds.; Palgrave Macmillan: New York, NY, USA, 2015.
11. Lipman, M. *Thinking in Education*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2003.
12. Gao, Z.Y. Philosophy for Children: An Analysis of the IAPC Textbooks and Other Diversified Curriculum Materials. *J. Zhejiang Norm. Univ. Soc. Sci.* **2010**, *2*, 41–45.
13. Lipman, M.; Sharp, A.M.; Oscanyan, F.S. *Philosophy in the Classroom*; Temple University Press: Philadelphia, PA, USA, 2010.
14. Gorard, S. *Research Design: Creating Robust Approaches for the Social Sciences*; Sage: Thousand Oaks, CA, USA, 2013.
15. Gorard, S.; Siddiqui, N.; See, B.H. *Philosophy for Children: Evaluation Report and Executive Summary*; EEF: London, UK, 2015.
16. Lord, P.; Dirie, A.; Kettlewell, K.; Styles, B. *Evaluation of Philosophy for Children: An Effectiveness Trial*; EEF: London, UK, 2021.
17. Yan, S.; Walters L.M.; Wang, Z. Meta-analysis of the effectiveness of philosophy for children programs on students' cognitive outcomes. *Anal. Teach. Philos. Prax.* **2018**, *39*, 13–33.
18. Boqin, L.; Peng, D. Development of the Children's Thinking and the Reform of Education in China: A Pioneering International Workshop on Philosophy for Children in Kunming, China. *Anal. Teach.* **1999**, *19*, 124–130.
19. Gao, Z. Confucian dialogue and the reconstruction of the community of inquiry in Philosophy for Children. In *Philosophy for Children in Confucian Societies: In Theory and Practice*; Routledge: London, UK, 2019.
20. Xiurong, G. On the thematic features of Philosophy for Children Inquiry Activities. *J. Beijing Inst. Educ.* **2017**, *6*, 6–12
21. Lam, C.M. Continuing Lipman's and Sharp's pioneering work on philosophy for children: Using Harry to foster critical thinking in Hong Kong students. *Educ. Res. Eval.* **2012**, *18*, 187–203. [CrossRef]
22. Tian, S.; Liao, P.F. Philosophy for children with learners of English as a foreign language. *J. Philos. Sch.* **2016**, *3*, 40–58. [CrossRef]
23. Fair, F.; Haas, L.E.; Gardosik, C.; Johnson, D.D.; Price, D.P.; Leipnik, O. Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a Philosophy for Children program. *J. Philos. Sch.* **2015**, *2*, 18–31. [CrossRef]
24. Alexander, R. Developing dialogic teaching: Genesis, process, trial. *Res. Pap. Educ.* **2018**, *33*, 561–598. [CrossRef]
25. Wang, J.C.S. Creating moral winds and nurturing moral growth in a P4C classroom community in Taiwan. In *Philosophy for Children in Confucian Societies: In Theory and Practice*; Routledge: London, UK, 2019.
26. Yang, P.M.H.; Yang, J.P. No one uses chopsticks to drink soup! In *Philosophy for Children in Confucian Societies: In Theory and Practice*; Routledge: London, UK, 2019.
27. Gao, Z. Thiry Years of Chinese Philosophy for Children Research: Review and Prospect. In *Educational Development Research*; Education Science: Hangzhou, China, 2019.
28. Kendall, J. Designing a research project: randomised controlled trials and their principles. *Emerg. Med. J. EMJ* **2003**, *20*, 164. [CrossRef] [PubMed]
29. Ennis, R.H.; Millman, J.; Tomko, T.N. *Administration Manual: Cornell Critical Thinking Tests*; The Critical Thinking Co.: Seaside, CA, USA, 2005.
30. Watson, G.; Glaser, E.M. *Critical Thinking Appraisal: Manual*; Psychological Corporation: New York, NY, USA, 1980.
31. Leng, L.; Gao, Z. The Development and Contextualization of Philosophy for Children in Mainland China. *Teach. Philos.* **2020**. [CrossRef]
32. Hong, F. Review and Prospect on the Research of Philosophy for Children. *J. Educ. Sci. Hunan Norm. Univ.* **2011**, *1*, 2–12.

education sciences

MDPI

Article

Delivering Music Education Training for Non-Specialist Teachers through Effective Partnership: A Kodály-Inspired Intervention to Improve Young Children's Development Outcomes

Lindsay Ibbotson * and Beng Huat See *

School of Education, Durham University, Durham DH1 1TA, UK
* Correspondence: lindsay.ibbotson@durham.ac.uk (L.I.); b.h.see@durham.ac.uk (B.H.S.); Tel.: +44-1748-519054 (L.I.)

Abstract: A priority area identified by the Department of Education (England) and the Economic Social and Research Council is the development of teachers, especially in primary music education where the limited opportunities for training offered by teacher training providers have raised concerns. This paper reports on an evaluation of a collaborative partnership training of non-specialist teachers, using a Kodály-inspired pedagogy to teach music in a classroom setting. Participants included 54 teachers (and 1492 pupils, aged 5–6), selected from 55 schools, as part of a large randomised control trial (RCT) in the north of England. This paper presents findings from a qualitative study that was conducted alongside the RCT, but which was not part of the RCT itself or the RCT's implementation and process evaluation. Results from the study, which include a pre-post survey of teachers, focus group interviews and reflective journals, suggest promising effects on teachers' pedagogical skills, their self-efficacy and competence, and children's self-confidence and disposition for learning.

Keywords: music education; Kodály approach; teacher professional development; partnership; music hubs; workplace mentoring; close to practice research

Citation: Ibbotson, L.; See, B.H. Delivering Music Education Training for Non-Specialist Teachers through Effective Partnership: A Kodály-Inspired Intervention to Improve Young Children's Development Outcomes. *Educ. Sci.* **2021**, *11*, 433. https://doi.org/10.3390/educsci11080433

Academic Editor: Gary McPherson

Received: 28 June 2021
Accepted: 9 August 2021
Published: 16 August 2021

1. Introduction

Increasing concerns about the declining numbers of students taking music in school in the UK have led many to call for arts and music to be given priority in schools because of the purported benefits, such as improved self-confidence, behavioural and social skills and academic attainment (e.g., [1] Schellenberg 2004; [2] DfE 2011; [3] Linnavalli et al., 2018; [4] Hallam 2015; [5] Hanson 2003). The 2019 *Music Education: State of the Nation* report recommended that the "government should encourage all schools to embed a culture of singing via classroom teaching". [6] (p. 19) This is especially so with the current need to address the challenges created by the COVID-19 pandemic, with increasing calls for arts and music to be made available to mitigate the effects of isolation and aid the recovery programme [7] (Education Endowment Foundation [EEF] 2021).

The introduction of the National Curriculum and the 1988 Education Reform Act in UK has made it compulsory for music to be taught in schools. This has put a huge demand on generalist teachers, who are now expected to teach music, many of whom have no prior knowledge of music nor the confidence to do so. Because of the lack of specialist teachers in primary schools, the role of non-specialist teachers has become even more important ([8] DfE 1988; [9] Hallam et al., 2009). Some have argued that music is best taught by the class teacher as it helps children to appreciate music as part of the curriculum ([10] Mills 1989). In England, there have been concerns that primary teachers teachers do not necessarily have the skills to teach what is required of the National Curriculum. [9] Hallam et al.'s (2009) survey of primary teacher trainees shows that only half said they were confident

in teaching music and suggested that more time should be devoted to training to prepare them for the curriculum. With the publication of the new Model Music Curriculum [11] (Department for Education [DfE] 2021), the aim of which is to ensure a universal provision of music education for all pupils, there is a pressing need to find ways to train teachers to prepare them for the new curriculum so that they can be confident in incorporating music into their core practice.

The Henley Report [2] [DfE] 2011 (p. 7), has identified inadequate teacher training, patchy delivery, and lack of government support and funding as key barriers to effective delivery of the music curriculum in England. A report on "Understanding and Developing Creativity in Early Years Settings" [12] (Pascal. C et al., 2005) also stressed "the continuing emphasis on early literacy and numeracy in many settings, which can lead to a 'bolt on' approach to work on creativity," and "a poor understanding of the role of creativity, or even how to use basic arts skills, by many early years practitioners and providers" (p. 2).

The present position is not much improved, despite initiatives such as 'Sing Up' and the Voices Foundation, demonstrating the power of singing in schools. A thorough review of music training for teachers of young children highlighted the importance of positive perceptions of music self-identity and self-efficacy. It emphasised the need for these to be encouraged and supported throughout a teacher's continuous professional development, beginning with initial teacher education [13] (Digby 2020, (p. 6)).

However, in a recent article, [14] Welch et al. (2020) lament the current situation, particularly in early years education, attributing the unequal access to music education to limited pedagogical experience of early years' teachers. One suggestion they made for overcoming this lack of pedagogical experience in generalist teachers is effective, class-based mentoring by specialist music educators (p. 9). They also suggest partnership between music specialists and early years colleagues in 'close-to-practice research' [15] (Wyse, Brown, Oliver, & Poblete, 2018) and the incorporation of the principles of effective music pedagogy when working with young children (p. 10).

There is a body of work that highlights the need to provide non-specialists with effective long-term training and support to develop their music skills, subject knowledge and confidence (e.g., [16] Barrett, Zhukov & Welch, 2019; [17] Bautista, Toh and Wong, 2016; [18] Holden & Button 2006; [14] Welch 2020). This concern is not unique to England or indeed the UK. Similar concerns were raised in Australia (again, Barrett, Zhukov & Welch, 2019), Indonesia ([19] Julia, Hakim & Fadilah, 2019), the USA, Namibia, Ireland and South Africa ([20] Russell-Bowie, 2009)

Several studies have consistently found that teachers with little or no music background are (not surprisingly) less confident and uncomfortable in teaching music (e.g., [9] Hallam et al., 2009; [21] Hennessy, 2017; [10] Mills, 1989; [22] Poulter & Cook, 2020). They are thus less likely to be able to provide high quality musical experiences and might even pass down their own discomfort and fear of music [23] (Siebenaler, 2006).

Recent studies suggest that effective mentoring and induction may improve teachers' job satisfaction and thus retention ([24] See et al. (2020); [25] Allen & Sims, 2017; [26] DeAngleis, Wall & Che, 2013; [27] Glazerman & Seifullah, 2013; [28] Gold, 1987; [29] Ronfeldt & McQueen, 2017), and this could be particularly important in a subject area where there may be less confidence. [16] Barrett, Zhukov & Welch's (2019) study of Australian teachers suggests that a collaborative approach to workplace mentoring for generalist teachers can help strengthen music provision in early childhood education. This study adopts such an approach, helping to address the gap between research and practice ([30] Glushankof (2007)) and exploring the potential for collaboration between teachers and music specialists. It aims to develop 'knowledge for use' to improve musical experience and learning, [31] (Young 2016), with trainers, specialists and teachers looking not only at the 'what' of teaching—such as the new Model Music Curriculum—but also the 'how' of teaching those skills ([32] Bremmer, 2015). This paper presents an evaluation of the professional development training of 54 non-specialist primary teachers who participated in a large randomised controlled trial (RCT) of the First Thing Music (FTM) intervention,

adopting a Kodály-inspired approach to music. The trial was conducted in the north of England. Although the trial was funded, (by the Education Endowment Foundation and the Royal Society for the Arts), to test the impact of FTM on young children's outcomes, the focus of this paper is on the collaborative approach to mentoring in the professional development of non-specialist teachers. The evaluation activities and findings presented in this paper were not part of the RCT study or its implementation and process evaluation and were not conducted with input from the EEF, but this study focuses on the group of teachers who participated in the RCT trial. While the RCT trial was mainly focused on pupil outcomes (as this was what the project was funded for), a large part of the FTM intervention involved the training, mentoring and regular support of generalist teachers. This aspect of the intervention should not be ignored as it has huge implications for teacher development. Any improvement in children's outcomes can only come about if teachers have the knowledge and understanding of music, the pedagogical skills and the confidence to deliver it.

The training programme itself was delivered through partnership with the British Kodály Academy (BKA), the Tees Valley Music Service (TVMS) and the Sheffield Music Hub. The BKA developed the teaching modules and the pedagogical approach to the training, while the TVMS and the Sheffield Music Hub provided the mentors who were, in turn, trained to support teachers in the school.

2. Background to the Kodály Approach

The Kodály approach on which FTM is based is described by [5] Hanson (2003) as "a sequential process, by which (a culture's folk) songs and active, authentic singing games are implemented to teach rhythm, melody, harmony, form, timbre, texture, and expression, in addition to the skills of singing, listening, moving, reading and writing notation, and the analysis of music (p. 7)".

Although it was developed by Zoltan Kodály, (1882–1967) in his native Hungary, it used the 'solfege' technique created in England by Sarah Glover and developed by John Curwen [33] (Landis 1972). It should be noted that FTM uses the term 'Kodály' to describe a Kodály-inspired approach which also incorporates theories of Dalcroze [34] (Choksy 1981)—internalizing musical patterns through movement. The Kodály approach has an international appeal, with well-established research and practice both in America [35] (Houlahan and Tacka 2016), and Australia, (the Kodály Music Education Institute of Australia), as well as in Spain. The British Kodály Academy is affiliated with the International Kodály Society and runs courses to support and promote quality music education in the UK.

The approach developed for FTM was chosen for simplicity and accessibility, being low-cost, involving no instruments, and designed to be delivered by generalist teachers during curriculum hours. Importantly, it made no assumptions of previous musical knowledge, taking the trainees through the same steps that they would follow with their classes—everything to be experienced and embodied.

Most Kodály music sessions in previous studies have been delivered by music specialists, and while there have been studies evaluating the impact of the Kodály approach to music on young children's learning and developmental outcomes, no studies have actually evaluated the use of this approach for training non-specialist teachers. And, as far as we know, prior to the EEF trial, there have been no studies that evaluated the impact of the Kodály trained non-specialist teachers on pupil outcomes. Our study is the first to do both.

3. The Model of Teacher Training Delivery

The teacher training programme in this study was developed based on the Kodály's philosophy to music, which emphasises intuitive interactive learning. One core principle of Kodály is that it should be taught in a logical and sequential manner. The techniques engage learners in singing, body movement and group exercises. Teachers were trained in the way they would teach the children following the core principles of Kodály. In other

words, the trainers model for the teachers how the music lessons would look, and teachers, in turn, model the singing and the movement for the children. It is well established that modelling is one of the most powerful teaching and learning strategies ([36] Bassanezi, 1994; [37] Cruess, Cruess and Steinert 2008; [38] Muijs and Reynolds, 2017).

In order to facilitate training on a Kodály-inspired approach, BKA trainers developed a printable Resources Pack, containing an introduction, and 30 songs and rhymes arranged in incremental progression (see Appendix C), with explanations of how the underlying pedagogical theory could be applied. This was augmented by recordings of the training sessions themselves, all made available on a password-protected website, (https://firstthingmusic.co.uk/) accessed from 30 November 2018. No instruments or equipment were necessary.

The teachers' physical experience of sound, movement and voicing laid the foundations for the development of musical skills and use of notation, which they in turn could deliver to the children in their own classrooms. The underlying philosophy—'Preparation; Presentation; Practise'—meant that everything was 'prepared' or introduced over time as physical experience, before 'presentation' as a conscious cognitive concept, (i.e., giving something a name, e.g., 'beat', a very brief process). This would then be followed with plenty of 'practise' of the use of the concept in creative ways. Critically, this applied as much to the teachers who went through this experiential process themselves as it would later to the children, with the same incremental progression being mirrored in the classroom stage by stage. For example, if we were looking at creating a sense of a steady beat, typically with a child's experience of being rocked or bounced preceding the developing control of movement in the upper and then lower body ([39] (Greenhalgh 2014, (p. 34)); [40] (Forrai, 1988)) then the trainers used the songs of the repertoire to give a similar sequence of experiences to the teachers. The sequential progression was central to the training and was reflected in the unfolding content of each session.

Training of Teachers

It should be noted that the EEF/RSA trial and the study presented in this paper were conducted in 2018/19 before the COVID-19 pandemic. All the FTM training was 'in-person' during that time. The lead trainers were both well-established tutors from the British Kodály Academy. Figure 1 shows the teacher training delivery model.

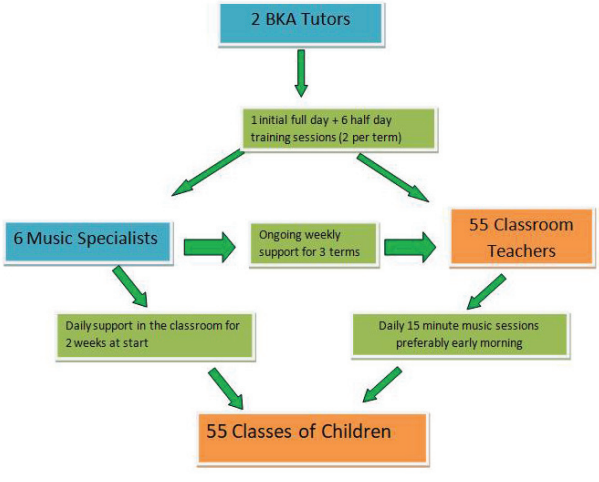


Figure 1. Training delivery.

The training was practical, consisting of one whole day in mid-September, plus one 3-h session for each of six half-terms over the academic year. All participants came together for

the first full day, but the 3-h sessions were often repeated in up to three optional locations to accommodate the wide regional spread of the participating schools—Teesside, the North East and Sheffield. In addition, each teacher was assigned a music practitioner/mentor, who spent weeks 1 and 3 of delivery visiting the school on a daily basis, leading the 15 min sessions at first, but gradually handing over the session leadership role to the teachers themselves as they gained confidence. After these intensive early weeks, the practitioners continued to support the teachers, but began reducing the contact time to one visit or phone call per week as necessary; sometimes more or less as required.

The profile of the group ranged from newly qualified teachers to those with three decades of teaching experience. Four of the 54 were male. Generally, the group sessions were held in an informal circle, enabling participants to take an active role in the songs and games, whilst simultaneously learning about the pedagogical approach of the Kodály-based activities. After the introductory full day on how to prepare experiences of steady beat for the children, the following half days covered:

- Making beat conscious;
- Experiencing rhythm;
- Making rhythm conscious and how to physically represent and visualise this, leading to analogical notation, (e.g., 1 large figure for a quarter note/crotchet and 2 smaller co-joined figures for eighth notes/quavers);
- Pitch matching—singing in tune;
- Pitch shape and melodic contour—introducing simple pitch notation.

At the end of the trial, all the teachers came together for a celebration event where feedback was collected from them about the success and challenges in delivering this model of teacher training.

The music sessions in schools were to be delivered every day for 15 min over the academic year. This paper reports only the results of the CPD evaluation, which collected data on the teachers' observed and reported self-efficacy, competence and teaching practice, as well as the perceived impact on children's behavioural and learning outcomes.

4. Aims

Although the FTM trial was to test the impact of this Kodály-inspired approach to music on children's outcomes, this paper presents a study conducted separately from the RCT trial, particularly focused on the model of training, which uses the Kodály-inspired pedagogy as described above. This approach to training involves the collaboration of a number of stakeholder partnerships. This kind of partnership between music specialists and teachers in 'close-to-practice research' [15] (Wyse, et al., 2018), has been suggested as an effective way to overcome the lack of pedagogical skills among generalist teachers. The aim of this paper is, therefore, to evaluate the feasibility of this kind of partnership in training non-specialist primary teachers in delivering the Kodály-inspired approach to music in a typical classroom setting, and the result of this training on teacher outcomes. As this was the first time that such a model of training has been used with the Kodály-inspired approach, we think it would be useful to identify some of the challenges and barriers to delivering the training.

Research questions:

1. How feasible is a collaborative partnership with music specialists in delivering Kodály-inspired music training to generalist teachers?
2. Does this kind of training appeal to schools, teachers and trainers?
3. What are the challenges and barriers to delivering this model of training?
4. To what extent is the Kodály-inspired approach to training associated with improvement in teachers' confidence and competence in delivery of music sessions and their attitude towards music?
5. What is the perceived impact on young children's learning and developmental outcomes?

5. Method

The First Thing Music training was designed as part of a two-armed randomised controlled trial (RCT) where 108 teachers across 55 schools were randomised to First Thing Music (the treatment group, n = 54) or a business-as-usual control, (n = 54) to test the impact of the Kodály-inspired approach to music on Year One (aged 5–6 years) children's academic and behavioural outcomes. For this paper we report the findings from the process evaluation, which documents the training of 54 non-specialist teachers in delivering music, and the reported impact on both teacher and pupil outcomes.

The data for the process evaluation was collected using a combination of strategies, including a pre-post survey (Appendix A), observation records from mentors, written feedback from teachers and headteachers, focus group feedback from teachers and school leaders and a reflective journal.

The questionnaire for the survey was designed by members of the FTM team, consisting of 12 items across four domains measured using 5-point Likert-type scales from Strongly Agree to Strongly Disagree, looking at the teachers' self-assessment of their personal musicality, how they saw the place of music in the classroom, attitudes to teaching music and workload and, lastly, attitudes to curriculum and testing in a music education context. All the teachers taking part in the training completed the survey on arrival on the first training day (n = 54), and the teachers were asked to repeat the process with identical 'post project' surveys, (n = 47). Analysis was performed on only the 41 teachers who completed both pre- and post-surveys. Categorical data from the surveys were analysed using odds ratios to see if teachers had become more or less positive in their perceptions of their musical competence after the intervention. We calculated the odds of teachers answering "strongly agree" and "agree" before and after the intervention. Odds ratios are calculated by dividing the after-intervention results by the results before intervention, giving the odds of how likely teachers are to report they agree or strongly agree with each of the items after the intervention.

Observational data by mentors who supported the teachers in the delivery of the programme were collected as weekly logs, producing a series of external points of reference throughout the formative process. The logs document the progress of the teachers assigned to that particular mentor.

Teachers and headteachers also provided written feedback at mid-point in the trial. In addition, verbal feedback from focus groups with teachers and school leaders was collated by an independent evaluator on the last training day of the project.

Fourteen teachers provided further insight into the effect on their teaching practice and pedagogy in their reflective journals. These journals were voluntary and were only required for those that wished to complete extra assignments to qualify for the accredited BKA Foundation Level Certificate.

6. Findings

6.1. Is the Collaborative Partnership with Music Specialists a Feasible Approach to Delivering Music Training?

This study was the result of a collaboration between several stakeholders, leading to the development of practical modules for training teachers in music and music specialists in mentoring skills. The question is whether this collaborative partnership is feasible. The fact that music specialists were able to develop a working syllabus from scratch, successfully train teachers to deliver the programme in schools, that teachers felt supported and that the programme was well received is evidence of the feasibility of this partnership. It is an example of the potential of such collaboration in providing diverse opportunities for professional musicians, music education organisations and schools to work together to support teachers in music education. As the descriptive data in 6.4 show, this approach was deemed to be not only feasible but also an effective way to introduce music experience to teachers with accompanying Kodály-inspired pedagogy. This supports the findings of [16] Barrett, Zhukov and Welch's (2019) research, in their evaluation of the workplace

mentoring of generalist teachers in early childhood education settings in Australia. It also confirms the claims by [15] Wyse et al. (2018) of the value of this kind of partnership in a 'close-to-practice research' as an effective way to develop pedagogical skills among generalist teachers.

6.2. What Do Teachers, Schools and Trainers Think of the Collaborative Kodály-Inspired Approach to Training?

Teachers received regular support in their own settings from specialist mentors. The Kodály-inspired approach, involving repetition of song/rhyme repertoire, helped to embed learning, as did the layering of games within each item. The *First Thing Music* Resource Pack plus the website videos developed by the music specialists supported teachers throughout their training. Teachers reported that the informal support and camaraderie through regular shared professional development sessions had been an important factor in helping to develop their confidence. This echoes the effects of the interactive songs and games in supporting the learning within the classes of children. All the teachers thought the intervention model was effective in developing their own music skills. The training did not require special equipment or instruments and might therefore be attractive to schools where the budget is tight. The playful approach also helped to develop a more positive teacher–class relationship, thus facilitating classroom management. In addition, teachers reported enjoying the music sessions (See Section 6.4).

The regular support of trained mentors with knowledge and expertise in this Kodály-inspired approach was crucial in providing that initial handholding on which the initially less confident teachers could rely. The participatory training where teachers and trainers together practised the songs and movements along with their colleagues helped build up that confidence. Trainers modelled how the songs were to be introduced, so teachers could also model and lead the singing and actions in class. This chimes with research on teaching and learning, which has consistently found role modelling to be an effective strategy (e.g., [38] Muijs and Reynolds 2017). All this has implications for in-service teacher training, school mentoring and induction programmes and the continuing professional development of teachers, not only in music teaching, but also in the wider curriculum.

6.3. What Are the Challenges and Barriers to Delivering This Model of Training?

Not all teachers found that the pace of the training suited them, especially when there were gaps between the CPD and the chance to deliver content in the classroom—a 'little and often' would have been preferred. Lack of leadership support, timetabling priorities such as phonics and maths and staff absences and turnover were other barriers. However, these kinds of barriers are not specific to the approach of training itself. Some teachers felt that some of the difficulties experienced were because the training was introduced as a 'one class at a time' intervention, interrupting the flow of normal shared activities within year groups. They suggested that a whole-school approach might be better.

6.4. Is There Any Relationship Between Training and Teacher Outcomes

6.4.1. Development of Teacher Musicality

Teachers reported observable benefits to their own practice and their ability to teach music to their classes. Of particular significance is the development of teachers' knowledge and understanding in delivering music lessons. Teachers had also developed confidence and self-efficacy in teaching children through singing and movement, and development in their musicality.

All teachers reported that they enjoyed listening to music before and after the intervention (Table 1). This is not surprising. However, comparing teachers' responses before and after the training, there were obvious positive changes in teachers' musicality. They were now almost six times more likely to sing to themselves often and three times more likely to report having a good sense of rhythm than before. After the training, teachers were over four times more likely than before to report greater confidence in performing

music in front of other people. A common theme that came up frequently was the progress teachers made in building their confidence in delivering music to their children. Feedback from teachers suggested that some struggled to find the confidence to sing in front of others earlier in the project. For example, one reflected in their journal:

> *I was a bit apprehensive at first, but actually I think I was just over-thinking it.*
>
> Journal Teacher 1

Another teacher also reflected on how the training had empowered them in their confidence in delivering music:

> *The project has inspired non-musical teachers to feel confident to deliver music within their settings which is a huge positive for the world of primary education. Having taken part in the* **First Thing Music** *project, I would highly recommend this approach to be used within schools, enabling all children to see themselves as 'musicians' regardless of ability and experience.*
>
> Journal Teacher 7

Observations from the FTM team also noted the teachers' developing reactions to the activities and their growing confidence. As the teachers relaxed and began to concentrate on the games, their confidence grew. Below are some examples of observations from the FTM research assistant:

> *Some teachers quite reluctant in the warm-up–"I can't do this"; I don't know what I'm doing!" or shy when stamping or jumping–little enthusiasm. This tended to ease when teachers worked together more,* e.g., *when pairing up and trying to 'break record', or working on 'Queen Queen Caroline' in a group.*
>
> *Enjoying the training, laughing, smiling at mistakes; becoming more confidence as training continues.*

As the year progressed, and the training covered the difference between beat and rhythm and how to notate simple rhythmic patterns, there were some real 'light bulb moments', as observed by the research assistant:

More than half originally said they expected you have to be musical to read music, and this reduced to none after the activity with the te-te ta. (This refers to activities that moved from physical representation of the number of sounds on a beat, using teams of children and floor spots, to simple analogical notation. Password-protected examples are on the training website: https/firstthingmusic.co.uk/rhythm-people/-Password available on request).

Table 1. Proportion of teachers agreeing with the statements before and after the training.

Teachers' Musicality	Agree	Not Agree	Odds Ratio
I enjoy listening to music			
Before	100	0	
After	100	0	0
I often sing to myself			
Before	78.1	21.9	
After	95.1	4.8	5.6
I have a good sense of rhythm			
Before	70.7	29.3	
After	87.8	12.2	2.98
I am confident to perform music in front of other people			
Before	19.6.	80.4	
After	51.2	48.8	4.3

6.4.2. Development of Teachers' Music Practice in the Classroom

The teachers' practices in the classroom have also seen positive changes (Table 2).

Table 2. Proportion of teachers who agreed with each of the statements before and after the training.

Teachers' Music Practice in the Classroom	Agree	Not Agree	Odds Ratio
My normal teaching practice does not incorporate singing with my class.			
Before	29.2	70.8	
After	9.8	90.2	* −3.8
I am happier using a pre-recorded music lesson than leading a music lesson myself.			
Before	68.3	31.7	
After	39.0	61.0	−3.4
Music can be a useful behavioural tool in the classroom			
Before	75.6	24.4	
After	85.4	14.6	1.1
I have a strong understanding of how to teach music to my class.			
Before	14.6	85.4	
After	73.2	26.8	15.98

* Negative signs indicate teachers were more likely to disagree with the statement after than before.

The biggest improvement was seen in the teachers' knowledge of teaching music. Teachers were now almost 16 times more likely than before to report a strong understanding of teaching music. This has implications for initial teacher training and also for professional development. To illustrate, we have an extract from a teachers' journal where they said:

> *Over the course of the First Thing Music project, I feel that I have developed my own skills in teaching and inspiring a love and understanding of Music education and how it can be taught through a process which is joyful and fun for the children. I believe strongly that the children in my class have come on a journey with me which has helped us to develop a secure grounding in life long, transferable musical skills.*
>
> Journal Teacher 5

Most teachers were grateful to the project for developing their deeper understanding of how to teach music in KS1, with some quite moving testimonies about the realisation that they could understand music in a new way themselves:

> *A stave of music appeared on the screen with traditional graphic notation and I could hear it clearly inside my head! I learnt to read graphic notation, in the traditional way as a child when learning to play the violin. During this process I was never encouraged to imagine what the music would sound like in my head, before "squeaking" it out on my instrument. This flash of inspiration made it clear how the Kodály(*-inspired*) method/*approach *prepares the learner for actually reading music, like we learn to read texts.*
>
> Journal Teacher 2

The mentor observation logs also recorded that teachers were making much progress, and by the end of the project, all had covered beat and rhythm as well as the use of notation. Of the 54 teachers, 28 had moved on to pitch by the third term.

To illustrate, we document the progress made by two teachers, as observed by their mentor. The teacher in School B had less confident children to begin with, but by the end of the week, the children were participating fully, and the teacher showed confidence in leading the songs in the second week. By the end of the Spring term, the mentor noted that

both children and their teacher had made good progress, working on items 11–13, (Engine Engine, Listen Listen and Apple Tree) in the repertoire (see Appendix C).

In School C, all the teaching assistants were fully involved in all the sessions, working with the staff and children as a large team. By the end of the Spring term, the teacher had covered items 13–15, (Apple Tree, Mary Ann and Hot Cross Buns) in a week, and the children had secured the beat and the rhythm, showing the ability to internalise. Teacher C was then able to move quickly to work on notation for these songs, including the addition of a 'rest', and some of the more complicated group games, such as the 'jump' in Hot Cross Buns and the hand drum/passing apples in Apple Tree. As an example, one mentor reported this:

> *Seeing the development in the class teachers, some from a quiet and unsecure place to now leading and writing an end of year sharing story involving many of the singing games—just fabulous!*
>
> Mentor M

Extracts from teachers' reflective journals provide clues as to how this was achieved. For example, one teacher noted how the support of their mentor enabled them to make the progress from Week 1 to Week 30.

> *Week 1: What have I been signed up for? I am not a musical person and I have spent a whole day this week singing. I don't know where to start. I am so pleased that we have a mentor coming in for a week so that she can show me what I actually need to do.*
>
> *Week2: This doesn't seem so bad after all. I'm not sure I see where this is going, but I can deliver it to the children and they really enjoy doing music every morning. The focus is on the steady beat and some of them are really getting it.*
>
> *Week 22: Mentor came in this week–It's now got to the point where I feel like we are just showing her what we have been up to. I feel much more confident singing in front of all of the children and even our mentor coming in doesn't bother me anymore. I've come to a realisation that I enjoy singing with the children and it doesn't matter if I am pitch-perfect, they just laugh it off with me. They love singing and now I actually love singing.*
>
> *Week 30: I introduced 'Charlie over the ocean' to the children today and the pure delight in their faces was lovely to see. I found that my foot was keeping the beat and when I discussed this with the children, they could explain to me exactly what I was doing. It's amazing to think that they have come so far since September with the language they not only know but understand. It isn't only the children that have come so far; I feel like my understanding of how to teach early music and my passion to do so has improved also.*
>
> Journal Teacher 1

Another teacher reflected in their journal:

> *As a teacher, I believe I have covered the music curriculum in greater depth this year. The support from my mentor, and the training sessions have given me the confidence and the vocabulary to ensure clear progression has been made, and the children are not only achieving, but are completely engaged in their learning. I hope to continue to use my training in subsequent years to ensure effective music teaching.*
>
> *The range of ability in the cohort is vast, with some children working well below age related expectation. However, this intervention provided appropriate challenge and support for all involved. It was amazing to see that specific children who are working at EYFS level are able to access the same lesson as their peers, with one child in particular excelling and becoming one of the more able in music.*
>
> Journal Teacher 13, Male

Positive improvements were also made in teachers' self-efficacy and confidence. Teachers were less likely to say that they did not incorporate music in their class. They were also less likely to want to use pre-recorded music, suggesting that they were now happier

leading the music session themselves. This is an indication of teachers' growing confidence and self-efficacy. Teachers' reflections in their journals also provide evidence of these changes:

> *The programme has been inspirational to myself, as a Year 1 teacher and my Teaching Assistant—both with a lack of musical ability. It has also been extremely encouraging when other members of staff pass on comments when working with the class, regarding their good listening skills, confidence, timing, good behaviour and positive attitude. I must also mention that the support given during the half termly sessions with the research team and the regular mentor support visits to school helped to alleviate my apprehension with delivering the programme.*
>
> Journal Teacher 4

> *I felt very under-confident and reluctant at the beginning of the process, and actually fely [sic] disappointed that my colleague hadn't been chosen for the project—now I'm going to lead Kodály sessions for all of KS1 next year.*
>
> Journal Teacher 3

Headteachers also reported changes in their staff's confidence in delivering music:

> *I have observed both classes in a session which I found very exciting. All children were on task and both members of staff delivered the sessions with confidence.*
>
> Headteacher 2

These findings concur with those of many previous studies showing how modelling, reflecting and acting was crucial in developing teachers' confidence and pedagogical knowledge (e.g., [37] Cruess, Cruess & Steinert, 2008, [21] Hennessy 2017; [22] Poulter and Cook, 2019).

Although teachers were less certain about music as a behavioural tool in the context of the survey (Table 2), several teachers reported an improvement in the children's punctuality and enthusiasm within the classroom, and this led to many commenting on the improvements that they had seen in behaviour. One teacher reported less behavioural incidents over the last year compared to the previous three years. Others reported how children were now better at turn taking and listening.

An earlier study [41] (See and Ibbotson, 2018) of the Kodaly-inspired approach suggests impact on children's social, emotional and behavioural development, showing an effect size of +0.71. This was a small, randomized study where children's outcomes were measured using the Early Learning Developmental Goals (ELGs) set out in the National Curriculum. One possible reason for the divergent findings could be that the current study was based on teacher observation, and children's behavioural outcomes were not monitored and measured to register progress. It is also possible that previous studies reporting positive effects of music on behaviour were typically conducted on children with autism and other socio-emotional and behavioural difficulties where music may be an effective intervention. Therefore, it is not surprising that no obvious changes in behaviour were observed among the majority of children. In any case, a large systematic review of experimental studies on the impact of music as a behavioural therapy has shown mixed results [42] (See and Kokotsaki, 2016). Future research, therefore, could look at the impact on the behavioural and learning outcomes of subgroups of children, such as children with special educational needs, those whose first language is not English as well as children in receipt of free school meals (i.e., those who would normally not have access to private music lessons).

6.4.3. Teachers' Attitudes to Teaching Music

The collaborative training programme has helped in the development of teachers' attitude to teaching music, many of whom started with a lot of apprehension and uncertainty. Teachers were more likely (2.4 times) to disagree that music does not make teaching enjoyable after the training than when they first started (Table 3). Teachers were slightly

more positive about being part of the research project at the end of the trial than at the beginning.

Table 3. Pre–post comparison of teachers' attitudes to teaching music.

Teachers' Attitude to Teaching Music	Agree	Not Agree	Odds Ratio
Musical activity does not make teaching more enjoyable			
Before	29.3	70.7	
After	14.6	85.4	−2.4
Delighted to be involved in research			
Before	73.2	26.8	
After	80.4	19.6	1.5

The written feedback and journal reflections from teachers show that teachers had enjoyed singing and were looking forward to more of the sessions:

> *We've thoroughly enjoyed our singing this term and are looking forward to more in the new year.*
>
> Written feedback Teacher 4

> *This is, without a doubt, my favourite part of the day. And what's even better is that it's the children's favourite part too. It's so nice to see the children learn through play, as they should, without numbers and targets to reach.*
>
> Journal Teacher 8

Feedback from focus group discussions echoed similar sentiments, where teachers said that the music sessions had brought 'enrichment', 'enjoyment' and 'pleasure' to the school day. This aligns with one of the core principles of the Kodály philosophy, which is that there should be pleasure in learning music. To this end, the strategy adopted in this study can be said to have achieved this.

The positive change in teachers' attitude to music could be the result of a number of factors. For example, the regular mentor and peer support and the opportunity to work in groups with peers who, like themselves, were at first rather diffident, meant that they became less worried or afraid of making mistakes. There is also the opportunity to reflect—all this may have contributed to this change in attitude. It is also the case that the songs introduced were simple and easy to learn—teachers did not feel that they needed a 'musical background' to be able to teach the music sessions.

6.4.4. Teachers' Perception of Music in the Curriculum

The survey results show there was a slight, positive shift in teachers' view of music as a core activity in the curriculum (Table 4). However, teachers were less sure if music has positive impact on children's attainment at KS1.

Table 4. Proportion of teachers who agreed with each of the statements before and after the training.

	Agree	Disagree	Odds Ratio
I do not see music as a core activity in the national curriculum			
Before	29.2	70.8	
After	12.2	87.8	−1.07
I think that music helps to improve children's scores in KS1 tests.			
Before	31.7	68.3	
After	41.5	58.5	1.5

While teachers were very positive about the impact of the training on their own skills and knowledge in delivering music, they were less sure about its impact on children's learning and its relevance to the curriculum. However, feedback from school leaders at the Evaluation Day discussions shows that some head teachers recognised the potential in Kodály (i.e., this Kodály-inspired approach to teaching music) in speech and language development and children's oracy. They saw this as a way to 'push on the national curriculum'. They said that they would like to see the curriculum broaden and take into greater account the simple enjoyment of the school day in order to improve learning in speech and language. Several teachers reported noticing improvements in children's fluency and expression when reading.

There have been several advocates for the Kodály-inspired approach to teaching to be integrated into the school curriculum. Some have even gone so far as to argue that music should be at the heart of the curriculum, a core subject used as a basis for education [34] (Chosky et al. (2001). Scott (2009) [43] explained how the collaborative inquiry stance of Kodály-inspired music education can be applied to the classroom across the curriculum where the instructional sequences and materials associated with the ideas of Kodály may be valuable guides for instruction. In the Kodály-inspired approach, teachers model the learning behaviour, and children learn through active engagement with teachers and their peers in a collaborative learning environment.

Previous studies have suggested that integrating music in the curriculum can facilitate learning, but only for primary school aged children. Some have shown positive effects on children's reading (e.g., [44] Bryant 2013; [45] Cochran 2009; [46] Lyons 2009; [47] Peters 2011; [48] Register et al., 2007; [49] Walton 2013) and maths ([50] An 2013; [51] Courey et al., 2012).

Some felt that the professional development of the project had 'left a legacy' of an increased value of music within their schools. They expressed that they had been inspired to continue music beyond the project as a result of taking part in *First Thing Music*, whilst others felt that this had opened up a new passion for them:

> *First Thing Music has certainly given me a passion for music and teaching early music I never imagined I'd have.*
>
> Journal Teacher 1

> *I can see that the children are really enjoying the sessions and they were fun for me to join in with them—think they enjoyed that too. I have also noticed that the children are growing in confidence when singing both together and individually and skills like their rhythm and pulse are also improving. I am a huge believer in the impact of music/song/rhythm on other areas of the curriculum and a child's development.*
>
> Headteacher 3

The following comment from a headteacher in the northeast highlighted a revival of interest in the potential benefits of music practice in the school curriculum:

> *I have really enjoyed participating in the First thing music over this term. This is something we used to do a lot of a few years ago and reminded me of some of our key music practice we've had in the past that we need to revisit and on a more regular basis—we are looking at this.*
>
> Headteacher 3

In another comment, a headteacher from a Middlesbrough school commented on how the music sessions had benefitted staff and pupils, especially those with special educational needs and those with behavioural issues, as well as those for whom English is not their first language:

> *Having attended a session this term I feel the following sums up my opinions:*
>
> *The sessions are well planned and engaging*
>
> *Children are actively engaged despite any barrier they present, e.g., EAL (English as an additional language), SEND, (special educational needs), or behavioural issues*

The opportunities for rhythm and patterns including dynamics are ripe

The structure of the sessions are sequential, stimulating and present opportunities for interleaved learning that benefits all children

I would love to attend more and more sessions and roll it out across my year 1 team

I feel the staff, children and any volunteers are benefiting greatly by the active learning of core songs, language development and musical opportunities

Headteacher 4

6.5. Perceived Impact on Children's Outcomes

Although the paper is about teacher development and its impact on teachers, the effect on children's outcomes should not be ignored. Here, we report only on teachers' observations of the impact on children's outcomes. The impact evaluation of the RCT will be reported elsewhere. At the time of writing, the results of the impact evaluation are not yet available.

6.5.1. Improvement in Children's Confidence

In their mid-point feedback, focus groups and journals, teachers reflected on the effect of the music sessions on the children in their classes. A common theme that came up very often was the improvement in children's confidence and self-esteem. This was repeated so many times by teachers, mentors and headteachers. Comments such as, "It makes them feel worth listening to", and "They don't normally have the confidence to raise their hand in maths, but now they do—after music" were heard during the November training day. Children were more ready to follow instructions and 'have a go':

The confidence of the children has improved greatly when singing and has also encouraged others with behavioural and social interaction difficulties to participate and end the project confident and happy to sing in front of others without concern. Children had opportunities to focus on their listening skills, develop their social interactions with others and also communicate their ideas in a concise way supported by their own experience evidence.

Journal Teacher 7

It was particularly heartening to hear that the music intervention had also made a big impact on children with special educational needs. Several teachers pointed out that the music sessions had a levelling effect—where very confident children might 'take a step back' and be 'a team member rather than a team leader'. The high achievers seemed to benefit from this access to group work—appearing to be more focused and enjoying the team work:

From the observations and from feedback and discussions with the intervention staff, we all felt that the children had become more used to taking turns and choosing different partners to those that they would usually choose. I noticed that the children were encouraged to make eye contact which was especially helpful for those children who find this difficult. One child who is normally very shy and extremely quiet, beamed throughout the session as it gave her the opportunity to find her voice. In both classes I saw the children really listening to the instructions which were required for each activity.

Headteacher 2

6.5.2. Perceived Impact on Children's Language Development

Teachers reported marked development in children's vocabulary, particularly among EAL (English as an additional language) children:

Working in a multilingual school with children from many nationalities and various levels of English, the First Thing music programme has been a brilliant way to get all the children involved. So far I have had two entirely non- English speaking children join my class and one of them had never even been to a school setting before. At first he used to cry and scream at having to come to school, wouldn't sit down on the carpet and

> *definitely wouldn't join in with social times such as playtime or lunch with the other children but gradually he would come and sit at the back of the hall during our music and then (with a little encouragement from my TA) he joined the circle. After a short while he was willing to join in with our welcome song and then he was the postman with 'Early in the Morning', which he did on his own! He now comes into the hall along with the others and participates with the others despite having no English. As for classroom behaviour, he is still struggling to sit and listen to stories and he'll wriggle etc but if I sing his name or sing an instruction he will turn and correct himself! This is an absolute difference to when he first arrived.*
>
> Journal Teacher 9

The Evaluation Day discussions yielded an account from a headteacher who felt that the daily music sessions had been invaluable in their school. They explained that their school community was 77% EAL, with many of their pupils coming from a Roma background. The School Leadership Team had noted improvements in fluency, writing skills, rhythm and hand–eye coordination. It was felt that the music had assisted these children in 'putting language into context' and improved their ability to infer:

> *The music subject lead and I went in to see the First Thing Music session on Friday and were really impressed with how it is going–confidence, concentration, social skills, speaking and listening skills, musicality, enjoyment were all evident.*
>
> Headteacher 1

Teachers also observed improvements in children's fluency, expression in reading and literacy:

> *With one child in particular who has speech and language issues, I have seen a great improvement. Before the First Thing Music project, the child was very hesitant to speak but he now joins in class discussions, shares answers and news and has clearer speech. This has been a wonderful development and huge achievement for this child.*
>
> Journal Teacher 11

> *The children have a great understanding of music and can keep the steady beat, rhythm, change pitch and volume. They are even able to read music which is amazing. The children in my class performed better in the Year 1 Phonics Screening test than the control group. My class had a 91% pass rate whereas the control group had a 82% pass rate. Both classes have received the same teaching of phonics, the same scheme followed (Letters and Sounds) and the same resources used.*
>
> Journal Teacher 11

Previous studies have reported positive effects of music in general on pre-school children's creativity [52] (Duncan, 2007), spatial-temporal ability [53] (Gromko and Poorman 1998), IQ scores [54] (Kaviani et al., 2014; [55] (Nering, 2002) and reading and language [56] (Harris, 2011; Myant et la., 2008). For primary school children, positive effects were reported for speech (e.g., [57] François et al. 2013; [58] Moreno et al., 2009), brain development (e.g., [59] Degé et al., 2014, [60] Olson, 2010; [1] Schellenberg 2004; [61] Schlaug et al., 2005), academic outcomes (e.g., [62] Harris, 2008; [63] Piro and Ortiz, 2009) and other cognitive skills (e.g., [64] Costa-Giomi, 1999; [65] Roden et al., 2014). However, the evidence is not conclusive, primarily because much research in this area has been weak and mostly based on teachers' reports.

While the benefits of music on children's learning have been widely acknowledged, the impact specific to the Kodály-inspired music education has not been robustly evaluated. One very early study compared two groups of children, one group receiving the Kodály music instruction and another acting as control [66] (Hurwitz et al., 1975). The study reported that the children in the experimental group performed significantly better than the children in the control group on standardised reading tests. This was a small study consisting of 40 non-randomised children. Comparison children were from two different schools.

See and Ibbotson (2018) [41] conducted a small, randomised control study where 56 pre-school children were individually randomised into treatment and control conditions. The results showed that Kodály-trained children outperformed the control children in number skills and spatial concepts after one term of exposure to the intervention. Gains in literacy (reading and writing) were only seen after two terms of exposure. These small-scale studies can only provide tentative answers. Larger-scale, more robust studies are needed to confirm the impact on children's academic outcomes. A systematic review including 17 studies that evaluated the impact of music found that the only promising approaches with e positive effects on very young children's cognitive development were the Kindermusik, Orff and Kodály methods [42] (See and Kokotsaki, 2015).

6.5.3. Perceived Impact on Children's Disposition for Learning

Schools reported positive developments in a number of aspects in children's disposition for learning. Teachers, for example, commented on how the music sessions had helped to ensure consistent behaviour for learning. They were more attentive and confident. Children who normally would not speak were now volunteering to speak. The music sessions were also seen as an equaliser where all children could join in, suggest ideas and perform. The perceived impacts included improved teamwork and cooperation, turn-taking and reduced incidents of misbehaviour. The teachers mentioned that punctuality had improved at the start of the school day and school drop-offs were easier, as the children appeared more willing to learn:

Children were happy and excited to come into school and start the day with singing. Developing this positive mind-set can only be constructive in supporting all learning and well-being.

Journal Teacher 13

Kodály inspired music has given the children confidence to perform amongst their peers and unites them in the fun of music. No matter how they enter school, in the morning, once the singing starts their attitude changes and any negative emotions disappear.

Journal Teacher 12

Headteachers also reported observing a range of pupil outcomes as a result of joining in with the music activity:

Having attended a couple of sessions it felt wonderful to be a part of the programme with a class of our Y1 children. I really enjoyed the games and the children sang really well. I think this kind of activity is having an impact as the children appear to be listening more attentively. Many children are able to keep a "steady beat" which I know will impact on many areas of the curriculum.

Headteacher 5

The children are participating more with all aspects of the music and learning. Confidence is developing and I am particularly impressed with levels of focus and concentration which is transferring to other sessions.

Headteacher 7

I have attended a number of music sessions over the term with squirrel class. I have noticed a real improvement in terms of their understanding of beat and also engagement of some of the harder to reach children.

Headteacher 8

Positive feedback was also received from parents and Ofsted inspectors:

OFSTED Inspector X visited—was impressed that 5 yr old children were reading musical notation, as was a parent who was assessing local private education options, and changed their mind on the spot!

Journal Teacher 6

One parent had this to say:

R never really settled into school well and would often cling to me when dropping off. It is only since R started the First Thing Music that she's been looking forward to arriving to school and doing some singing.

7. Conclusions

This study has shown that the collaborative partnership with stakeholders has worked in supporting the delivery of the training for non-specialists in schools, which has led to the development of teachers' skills, knowledge, self-efficacy and musicality. This has, in turn, contributed to children's enjoyment of music and the development of their learning according to teachers' and headteachers' reports. The findings from this study are promising. It shows that this collaborative approach to delivering training to non-specialist teachers is feasible and has proved to be very appealing to schools and music hubs, (which saw this as an investment in enhancing appreciation of value of their music services to schools in the future). For the BKA, the training represented an opportunity to broaden the access to its structured, sequential approach to learning musicianship through singing and movement and to incorporate contributions from other methodologies such as Dalcroze.

In summary, the findings of this study based on teacher perceptions could be represented as below, in Figure 2:

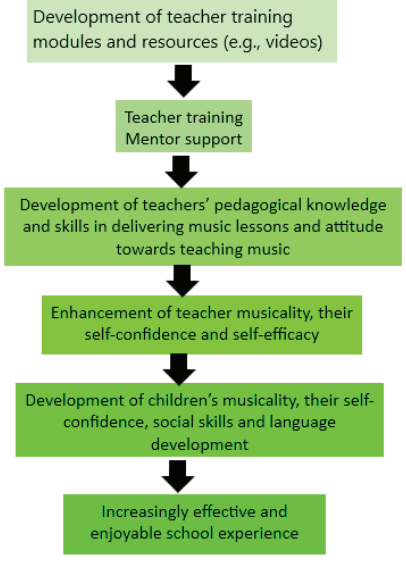


Figure 2. Synthesis of findings, based on teacher perceptions.

There were limitations to this study, as it was running alongside an RCT. This meant that participants generally had to work separately from their colleagues in school to avoid 'cross contamination' of effects in the impact evaluation. Developments since has enabled year group colleagues within a school to work together, which seems to be opening new possibilities and increasing the 'take-up' and effect of the training. It would be interesting to compare the findings of this study with future evaluations of the training, without the need for separate intervention and control groups.While the model of teacher training using the Kodály-inspired principles to music appears to work with early years primary school teachers and children, it may not be as useful for older primary or secondary teachers. This has never been tested, so perhaps the next phase could be the development of new modules for training teachers in the upper primary and secondary schools.

Since the end of the trial, the delivery team has worked on the professional development model in response to the continued enthusiasm from participating schools. A hybrid model combining online and in-person training has been developed to enable teachers to continue to be trained and supported during the pandemic. There has also been a lot of interest from initial teacher training specialists. This training model could be developed as a core approach for all teacher training.

One of the biggest challenges to delivery was fitting the 15-minute music sessions into an already crowded timetable. On the other hand, some teachers felt that focus could be achieved more quickly through 'singing rather than shouting', also making progress through other required elements of the curriculum more achievable. Certainly, any school wishing to introduce this form of teacher development needs to ensure that they have strong leadership support to provide the flexible time and space for music sessions. Whether this should be 'first thing' in the school day or not requires further investigation.

Two important lessons were learnt that could help to address the challenges of wider delivery with the currently musically under-skilled workforce. First, at the teacher professional development level, teachers needed to have the knowledge and confidence to be able to deliver the intervention in their own settings. This Kodály-inspired approach to music training, and working in partnership with music specialists, has shown to be a promising way to provide that knowledge and confidence to generalist teachers. Teachers in this study remarked on the importance of regular meetings, support and practical sharing with colleagues. Simply being given an introduction to some resources was not enough. A chance to try things out within a practical forum, having regular feedback and the opportunity to try musical games out together before introducing them to a class of children were all important in developing that confidence.

Second, this new model of teacher training using the Kodály-inspired approach can be developed to provide a practical way for music specialist mentors to share the foundations of music education with non-specialists. One key factor in the success of this approach is the expertise of the training team from the BKA. Music specialist mentors had gained an enormous amount from being part of a team that was learning 'on the job', developing their own pedagogy through this embodied approach and also discovering ways of supporting teachers and students at the post-graduate level.

Finally, we would like to add that while the findings of the process evaluation provided interesting insight into the challenges, benefits and perceived impact on teachers' and children's learning experiences, they should be interpreted with caution as they are based on participants' self-reports and observations. Nevertheless, important lessons can be learnt from the process, which suggest wider implications for the professional development of teachers, teacher–pupil relationship and job satisfaction—all of which have important policy implications for teacher training and retention, as well as for children's experience in school.

Author Contributions: Conceptualisation: L.I.; Methodology: B.H.S. and L.I.; Formal analysis–B.H.S.; Resources–L.I.; Writing—review and editing: L.I. and B.H.S.; Project administration, L.I. All authors have read and agreed to the published version of the manuscript.

Funding: The project was funded by the Education Endowment Foundation and the Royal Society of Arts.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as it involved only adults who all gave informed consent for participation. No sensitive personal data were used in the study. All data was anonymised and de-identified.

Informed Consent Statement: All participants agreed to their anonymised data being used for the research.

Data Availability Statement: Not applicable.

Acknowledgments: British Kodály Academy trainers: Lucinda Geoghegan, and Zoe Greenhalgh. Hub delivery partners: Tees Valley Music Services—Susan Robertson, Service Manager; supported

by Sheffield Music Hub. With thanks to the team of mentors who also helped to collect and collate the Teacher observations: Jane Harland, Zoe Greenhalgh, Rebecca Denniff, Clare Gale, Caroline Bell, Naomi Haigh, Colette Dutot, and Research Assistant, Olivia Wallis.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

First Thing Music – Teacher Baseline Questionnaire

Thank you for taking part in the First Thing Music Programme. As part of our evaluation, we are interested in teacher's experience, confidence and attitudes about music.

The aim of this questionnaire is to encourage you to reflect upon your musicality, prior experience of teaching music, and attitudes. The survey will take around 8 minutes to complete.

The survey is **strictly confidential**. The collection of this data is for research and reporting purposes only. Once collected, the data will be kept on secure databases and all personal identifiers will be removed. By completing this survey, you are consenting to take part in this research. If you would like to withdraw your consent at a later date, or if you have any questions about the survey, please email

Thank you for reading this.

Lindsay Ibbotson, Project Lead.

Name:

School:

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Section A – personal musicality					
1.I enjoy listening to music.					
2. I often sing to myself, (perhaps when I'm alone/when no-one is listening).					
3. I have a good sense of rhythm, (enjoy dancing, clapping/tapping to music).					
4. I am confident to perform music in front of other people, (as part of a group, or individually).					
Section B – music in the classroom					
5. My normal teaching practice does not incorporate singing with my class.					
6. I am happier using a pre-recorded music lesson than leading a music lesson myself.					
7. Music can be a useful behavioural tool in the classroom.					
8. I have a strong understanding of how to teach music to my class.					
Section C – Attitude to teaching/workload					
9. I do not feel that musical activity makes my teaching more enjoyable.					
10. I am delighted to be involved in this research project.					
Section D – Attitude to curriculum & testing					
11. I do not see music as a core activity in the national curriculum.					
12. I think that music helps to improve children's scores in KS1 tests.					

Please list any academic music qualifications you hold (GCSE; A-Level, Degree):

Please list any music grades you hold:

Are you involved in any musical activities outside work?

Figure A1. Teacher Self-Assessment Survey.

Appendix B. Teacher Training Package–Further Details/Examples

Timings	TRAINING DAY WITH FIRST THING MUSIC TEAM AND PRIMARY TEACHERS 11th September 2018
9.30am	Welcome and Introduction to the First Thing Music Team
9.45	Warm- up with singing
10.00	Session One: MUSICAL SKILL BUILDING IN EARLY YEARS
11.00	Coffee
11.20	Choosing appropriate Repertoire – • Helping children to sing in tune • Why we have chosen the songs in the pack? • How to teach the songs in the pack – trainers lead five activities from the pack
12.15pm	Why Music? – Discussion
12.45	Lunch
1.45	Working in groups with the First Thing Music Mentors. • Introductions/play name game • Mentors to lead three of the songs.
2.30	Join together for Learning of the final two songs and a listening activity
3.00	Quiz time – in teams with Mentors • Recall of all ten songs – from words/from humming/what's the next line? Etc
3.15	Round up/questions/final comments.

Figure A2. Schedule for the initial training day.

Table A1. Content for CPD October 2018—The 'Lego' Approach—Lindsay Ibbotson & Zoe Greenhalgh.

Content	Notes
Introduction Theme of day—*The Lego Approach* or How to make the most of some basic materials/recycle/build a new game	The most basic components of music begin with • Beat • Rhythm • Pitch For this term, our main focus will be on 'beat'—developing a shared sense of the musical 'heartbeat' underlying all the songs and rhymes that we are using. This does not mean that we will not pick up on things such as dynamics, (louder/quieter) or tempo (faster/slower) as they occur, but the fundamental quality of music involves that thing that makes us tap our toes along with the music! What can we make using this particular part of the lego set?
Surgery session—A large circle of 30-ish people, led by Zoe/Lindsay, discussing experiences, both positive and negative that have arisen over early weeks of the project. Examine any issues, and then split into groups with relevant music practitioners or specialists on subject.	Everyone has now had at least one week of daily music sessions in their classrooms and some experience of leading their own sessions (at least for a couple of days). • When sessions were being led by the specialists from FTM, what observations did class teachers make about how the children responded? • How did it feel to be taking over? • How are the children engaging with the musical activities?
New song—See the Candle Light (?)	Learn new song—(currently an extra, just for seasonal use, but later to teach concept of the 'rest' over the beat); incorporate in workshop below.

Table A1. *Cont.*

Content	Notes
Exploring the 'Lego' aspect of a Kodály-inspired approach in more depth . . . Three stages interweaving throughout the 'fabric' of the activities: • Preparation • Presentation • Practice At this stage, everything we are doing is about the Preparation stage, helping the children to 'feel' and experience a particular aspect of music (in this case, 'beat') BEFORE introducing the underlying concept formally.	You do not necessarily need loads of new toys in the box—you can be endlessly creative with very little. Pedagogically, this means that learning can be about concepts and creativity rather than ticking off lists of material used. Musical behaviour is playful and being playful increases the chance that everyone will be enjoying themselves. When children are feeling happy and secure, this is likely to support the learning process. • Social interaction—listening, being heard, understanding about 'taking turns', taking pleasure in being synchronised with the group. • Trying out new skills as part of a game, in a safe environment—singing, moving on a beat, using gross and fine motor skills. • Getting instant feedback within the game, speeding up the process. • Developing evaluation skills—e.g., Did we all finish 'together' or 'nearly together'? Was that a speaking voice or a singing voice? Were we faster or slower that time? • Group motivation and support—Can we get this 'just right'? • Creativity—once the basics are learned, the teacher can encourage the children to contribute their ideas around the game.
Workshop Group work, with each group undertaking to find new, creative ways of playing with two songs from the currently available repertoire. Share afterwards, with demonstrations. Pick out aspects of the games which correlate with the six points above.	There is always more than one way to play games with this material. It is almost like playing with Lego—the bricks will take deconstructing and reconstructing many times. The main thing for the teacher is to understand the underlying concept that we're practising, and to keep 10 min of playing 'on track' with this in mind.

Appendix C. Extract from the Introductory Pack and the 30-Item Repertoire List

See Also 87910 RESOURCE FILE_Layout 1

Kodály-based musical pedagogy is built on the following sequence:

Preparation:

Just sing and play the games! The children will be subconsciously experiencing music and will be building a song bank from which concepts can be introduced later. The preparation stage can take a long time but is a crucial stage in ensuring that the child understands. Just like learning to read, the child catches the language before reading it.

- Presentation:

This happens only when the children are secure with a particular aspect of the activities—then is the time to give the concept a name!

- Practise:

Open to a lifetime of creative uses of a skill—e.g., 'feeling the beat'; counting people in to the beginning of a phrase or song; being able to identify whether a piece of music is in 3:4 or 4:4; making up dance steps or lyrics to the music, etc.

First Thing Music : Repertoire list (2018/19)

Number	Title	Song/Rhyme	Concept
1	*Swing Me Over the Water*	**S**	Beat
2	*Copy Cat*	**S**	Beat
3	*Touch Your Shoulders*	**S**	Beat (Rhythm)
4	*Ickle Ockle, Bluebottle*	**S**	Beat (Rhythm) Dynamics
5	*Feet Feet*	**R**	B & R
6	*Early in the Morning*	**S**	Beat & Tempo
7	*Sally Go Round the Sun*	**S**	Beat & Dynamics
8	*1234*	**R**	B (Phrasing) (& R)
9	*Queen Queen Caroline*	**R**	B (& R) Phrasing
10	*Jack Be Nimble*	**S** & chant	Beat
11	*Engine Engine No 9*	**R**	B & R
12	*Listen Listen*	**S**	B & R
13	*Apple Tree*	**S**	B & R (Pitch)
14	*Mary Anne*	**S**	B & R (Pitch)
15	*Hot Cross Buns*	**S**	B & R *(Pitch - **not** m-s-l)*
16	*Rain is falling down*	**S**	B & R *(Pitch - **not** m-s-l)*
17	*Mix a Pancake*	**S**	B & R (Pitch)
18	*Row Boys Row*	**S**	B & R (Pitch)
19	*Bow Wow Wow*	**S**	B & R *(Pitch - **not** m-s-l)*
20	*Pease Pudding Hot*	**S**	B & R (Pitch)
21	*Doggie Doggie*	**S**	Pitch Matching
22	*Here I Come*	**S**	Pitch Matching
23	*Charlie over the Ocean*	**S**	Pitch Matching
24	*Who's that Tapping?*	**S**	PM/Melodic Shape
25	*Andy Pandy*	**S**	Melodic Shape
26	*Jack in the Box*	**S**	Melodic Shape
27	*Hi Lo Chickalo*	**S**	Melodic Shape
28	*Cuckoo Cherry Tree*	**S**	Melodic Shape : So/Mi
29	*No Robbers out today*	**S**	Melodic Shape : So/Mi
30	*I, I, Me O My!*	**S**	Melodic Shape : So/Mi

Figure A3. The 30-Item Repertoire List.

References

1. Schellenberg, G. Music lessons enhance IQ. *Psychol. Sci.* **2004**, *15*, 511–514. [CrossRef] [PubMed]
2. Henley, D. *Music Education in England: A Review for the Department of Education and the Department for Culture, Music and Sport*; Department for Education: London, UK, 2011.
3. Linnavalli, T.; Putkinen, V.; Lipsanen, J.O.; Huotilainen, M.; Tervaniemi, M. Music playschool enhances children's linguistic skills. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef]
4. Hallam, S. *The Power of Music: A Research Synthesis of the Impact of Actively Making Music on the Intellectual, Social and Personal Development of Children and Young People*; Music Education Council: London, UK, 2015.
5. Hanson, M. Effects of Sequenced Kodaly Literacy-Based Music Instruction on the Spatial Reasoning Skills of Kindergarten Students. *Res. Issues Music Educ.* **2003**, *1*, 4. Available online: http://ir.stthomas.edu/rime/vol1/iss1/4 (accessed on 28 April 2021).
6. Daubney, A.; Spruce, G.; Annetts, D. Music Education: State of the nation. Report by the All-Party Parliamentary Group of Music Education, the Incorporated Society of Musicians and the University of Sussex. (Page 19). 2019. Available online: https://www.ism.org/images/images/State-of-the-Nation-Music-Education-WEB.pdf (accessed on 28 April 2021).
7. EEF March 21 EEF Publishes New Analysis on Impact of COVID-19 on Attainment Gap | News | Education Endowment Foundation | EEF. Best Evidence on Impact of COVID-19 on Pupil Attainment | Education Endowment Foundation | EEF. Available online: https://educationendowmentfoundation.org.uk/eef-support-for-schools/covid-19-resources/best-evidence-on-impact-of-school-closures-on-the-attainment-gap/ (accessed on 21 March 2021).
8. DfE. *Teaching: High Status, High Standards (Circular 4/98)*; HMSO: London, UK, 1998.
9. Hallam, S.; Burnard, P.; Robertson, A.; Saleh, C.; Davies, V.; Rogers, L.; Kokotsaki, D. 'Trainee primary school teachers' perceptions of their effctiveness in teaching music. *Music Educ. Res.* **2009**, *11*, 221–240. [CrossRef]
10. Mills, J. The Generalist Primary Teacher of Music: A Problem of Confidence. *Br. J. Music Educ.* **1989**, *6*, 125–138. [CrossRef]
11. Model Music Curriculum. 26th March 2021 Teaching Music in Schools—GOV.UK. Available online: www.gov.uk (accessed on 27 March 2021).

12. Pascal, C.; Bertram, T. *How to Catch a Moonbeam and Pin It Down: Creativity and the Arts in Early Years*; Amber Publications and Training: Birmingham, UK, 2017.
13. Digby, J. Teacher Confidence to Facilitate Children's Musical Learning and Development in the Reception Year at School. Ph.D. Thesis, University College London, London, UK, 2020.
14. Welch, G.F. The challenge of ensuring effective early years music education by non-specialists. *Early Child Dev. Care* **2020**, 1–13. [CrossRef]
15. Wyse, D.; Brown, C.; Oliver, S.; Poblete, X. *The BERA Close-to-Practice Research Project: Research Report*; British Educational Research Association: London, UK, 2018. Available online: https://www.bera.ac.uk/researchers-resources/publications/ (accessed on 28 April 2021).
16. Barrett, M.S.; Zhukov, K.; Welch, G.F. Strengthening music provision in early childhood education: A collaborative self-development approach to music mentoring for generalist teachers. *Music Educ. Res.* **2019**, *21*, 529–548. [CrossRef]
17. Bautista, A.; Toh, G.-Z.; Wong, J. Primary school music teachers' professional development motivations, needs, and preferences: Does specialization make a difference? *Music Sci.* **2018**, *22*, 196–223. [CrossRef]
18. Holden, H.; Button, S. The teaching of music in the primary school by the non-music specialist. *Br. J. Music Educ.* **2006**, *23*, 23–38. [CrossRef]
19. Julia, J.; Hakim, A.; Fadlilah, A. Shifting Primary School Teachers Understanding of Songs Teaching Methods: An Action Research Study in Indonesia. *Int. J. Educ. Pract.* **2019**, *7*, 158–167. [CrossRef]
20. Russell-Bowie, D. What me? Teach music to my primary class? Challenges to teaching music in primary schools in five countries. *Music Educ. Res.* **2009**, *11*, 23–36. [CrossRef]
21. Hennessy, S. Approaches to increasing the competence and confidence of student teachers to teach music in primary schools. *Education* **2017**, *45*, 689–700. [CrossRef]
22. Poulter, V.; Cook, T. Teaching music in the early years in schools in challenging circumstances: Developing student teacher competence and confidence through cycles of enactment. *Educ. Action Res.* **2020**, 1–17. [CrossRef]
23. Siebenaler, D. Training Teachers with Little or No Music Background: Too Little, Too Late? *Update Appl. Res. Music Educ.* **2006**, *24*, 14–22. [CrossRef]
24. See, B.H.; Morris, R.; Gorard, S.; Kokotsaki, D.; Abdi, S. Teacher Recruitment and Retention: A Critical Review of International Evidence of Most Promising Interventions. *Educ. Sci.* **2020**, *10*, 262. [CrossRef]
25. Allen, R.; Sims, S. *Improving Science Teacher Retention: Do National STEM Learning Network Professional Development Courses Keep Science Teachers in the Classroom*; Wellcome Trust: London, UK, 2017.
26. DeAngelis, K.J.; Wall, A.F.; Che, J. The Impact of Preservice Preparation and Early Career Support on Novice Teachers' Career Intentions and Decisions. *J. Teach. Educ.* **2013**, *64*, 338–355. [CrossRef]
27. Glazerman, S.; Seifullah, A. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years*; Final Report; Mathematica Policy Research, Inc.: Washington, DC, USA, 2012.
28. Gold, M.J.; City University of New York. Institute for Research and Development in Occupational Education; American Association of State Colleges and Universities. In *Retired Teachers as Consultants to New Teachers A New Inservice Teacher Training Model*; Final Report. Case 09–87ERIC Clearinghouse: Washington, DC, USA, 1987. Available online: https://eric.ed.gov/?id=ED306928 (accessed on 28 April 2021).
29. Ronfeldt, M.; McQueen, K. Does New Teacher Induction Really Improve Retention? *J. Teach. Educ.* **2017**, *68*, 394–410. [CrossRef]
30. Gluschankof, C. Research and practice in early childhood music education: Do they run parallel and have no chance to meet? The case of preschool singing repertoire. In *Proceedings of European Network for Music Educators and Researchers of Young Children*; Young, S., Nicolau, N., Eds.; University of Cyprus: Nicosia, Cyprus, 2007; pp. 27–31.
31. Young, S. Early Childhood Music Education Research: Research Studies in Music Education. Available online: rsm.sagepub.com (accessed on 28 May 2021).
32. Bremmer, M.; Hermans, C. What the body knows about teaching music. In *Embodiment in Arts Education. Teaching and Learning with the Body in the Arts*; Lectoraat Kunst-en Cultuureducatie: Amsterdam, The Netherlands, 2015; pp. 14–27.
33. Landis, B. The Eclectic Curriculum in American Music Education: Contributions of Dalcroze, Kodaly, and Orff. Washington: Music Educators National Conference. 1972. Dewey Decimal Class-780/.72973; Library of Congress-MT3.U5 L33: ID Numbers-Open Library-OL5341967M; LC Control Number-72195923. Available online: https://www.semanticscholar.org/paper/The-eclectic-curriculum-in-American-music-of-and-Landis-Carder/c7e3fa53862c677c295d30c762c5f5b6bcd9cd78 (accessed on 28 April 2021).
34. Caldwell, H.L.; Choksy, L. The Kodály Context: Creating an Environment for Musical Learning. *Music Educ. J.* **1981**, *68*, 58–59. [CrossRef]
35. Houlahan, M.; Tacka, P. *Kodály Today: A Cognitive Approach to Elementary Music Education Inspired by the Kodály Concept*; Oxford University Press: Oxford, UK, 2015.
36. Bassanezi, R.C. Modelling as a teaching-learning strategy. *Learn. Math.* **1994**, *14*, 31–35.
37. Cruess, S.R.; Cruess, R.L.; Steinert, Y. Role modelling—making the most of a powerful teaching strategy. *BMJ* **2008**, *336*, 718–721. [CrossRef] [PubMed]
38. Muijs, D.; Reynolds, D. *Effective Teaching: Evidence and Practice*; Sage: London, UK, 2017.
39. Greenhalgh, Z. *Music and Singing in the Early Years*; Routledge: London, UK; New York, NY, USA, 2014.

40. Forrai, K. *Music in Preschool, Translated and Adapted by Jean Sinor*; Corvina: Budapest, Hungary, 1988.
41. See, B.H.; Ibbotson, L. A feasibility study of the impact of the Kodály-inspired music programme on the developmental outcomes of four to five year olds in England. *Int. J. Educ. Res.* **2018**, *89*, 10–21. [CrossRef]
42. See, B.H.; Kokotsaki, D. Impact of arts education on children's learning and wider outcomes. *Rev. Educ.* **2016**, *4*, 234–262. [CrossRef]
43. Scott, S. Integrating Inquiry-Based Constructivist Music Education with Kodaly-Inspired Learning. *Aust. Kodaly J.* **2019**. Available online: https://search.informit.org/doi/epdf/10.3316/informit.104672790564030 (accessed on 23 June 2021).
44. Bryant, L.; Gibbs, L. *A Director's Manual: Managing an Early Education and Care Service in NSW*; Community Child Care Co-Operative, Ltd.: Marrickville, Australia, 2013.
45. Cochran, K.H. The Effects of Singing and Chanting on the Reading Achievement and Attitudes of First Graders. Ph.D. Thesis, Clemson University, Clemson, SC, USA, 2009.
46. Lyons, L.M. The Integration of Music with Reading Concepts to Improve Academic Scores of Elementary Students. Ph.D. Thesis, Colorado State University, Fort Collins, CO, USA, 2009.
47. Peters, B. A Formative Study of Rhythm and Pattern: Semiotic Potential of Multimodal Experiences for Early Years Readers. Ph.D. Thesis, University of Manitoba, Winnipeg, MB, Canada, 2011.
48. Register, D.; Darrow, A.-A.; Swedberg, O.; Standley, J. The Use of Music to Enhance Reading Skills of Second Grade Students and Students with Reading Disabilities. *J. Music Ther.* **2007**, *44*, 23–37. [CrossRef]
49. Walton, J.P. The Effect of Music on the Reading Achievement of Grade 1 Students. Ph.D. Thesis, Jones International University, Centennial, CO, USA, 2013.
50. An, S. The Effects of Music-Mathematics Integrated Curriculum and Instruction on Elementary Students' Mathematics Achievement and Dispositions. Ph.D. Thesis, Texas A&M University, College Station, TX, USA, 2013.
51. Courey, S.J.; Balogh, E.; Siker, J.; Paik, J. Academic music: Music instruction to engage third-grade students in learning basic fraction concepts. *Educ. Stud. Math.* **2012**, *81*, 251–278. [CrossRef]
52. Duncan, D.J. The Relationship between Creativity and the Kindermusik Experience. Master's Thesis, University of Central Missouri, Warrensburg, MO, USA, 2007.
53. Gromko, J.E.; Poorman, A.S. The Effect of Music Training on Preschoolers' Spatial-Temporal Task Performance. *J. Res. Music Educ.* **1998**, *46*, 173–181. [CrossRef]
54. Kaviani, H.; Mirbaha, H.; Pournaseh, M.; Sagan, O. Can music lessons increase the performance of preschool children in IQ tests? *Cogn. Process.* **2013**, *15*, 77–84. [CrossRef]
55. Nering, M.E. The Effect of Piano and Music Instruction on Intelligence of Monozygotic Twins. Ph.D. Thesis, University of Hawaii, Honolulu, HI, USA, 2002.
56. Harris, D.J. Shake, rattle and roll—Can music be used by parents and practitioners to support communication, language and literacy within a pre-school setting? *Educ. Int. J. Prim. Elem. Early Years Educ.* **2011**, *39*, 139–151. [CrossRef]
57. François, C.; Chobert, J.; Besson, M.; Schön, D. Music Training for the Development of Speech Segmentation. *Cereb. Cortex* **2012**, *23*, 2038–2043. [CrossRef] [PubMed]
58. Moreno, S.; Marques, C.; Santos, A.; Santos, M.; Castro, S.L.; Besson, M. Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cereb. Cortex* **2009**, *19*, 712–723. [CrossRef]
59. Degé, F.; Wehrum, S.; Stark, R.; Schwarzer, G. Music lessons and academic self-concept in 12- to 14-year-old children. *Music Sci.* **2014**, *18*, 203–215. [CrossRef]
60. Olson, C.A. Music Training Causes Changes in the Brain. *Teach. Music* **2010**, *17*, 22.
61. Schlaug, G.; Norton, A.; Overy, K.; Winner, E. Effects of Music Training on the Child's Brain and Cognitive Development. In *The Neurosciences and Music II: From Perception to Performance*; Avanzini, G., Lopez, L., Koelsch, S., Manjno, M., Eds.; New York Academy of Sciences: New York, NY, USA, 2005; pp. 219–230.
62. Harris, M. The Effects of Music Instruction on Learning in the Montessori Classroom. *Montessori Life* **2008**, *20*, 24–31.
63. Piro, J.M.; Ortiz, C. The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychol. Music* **2009**, *37*, 325–347. [CrossRef]
64. Costa-Giomi, E. The effects of three years of piano instruction on children's cognitive development. *J. Res. Music Educ.* **1999**, *47*, 198–212. [CrossRef]
65. Roden, I.; Grube, D.; Bongard, S.; Kreutz, G. Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychol. Music* **2014**, *42*, 284–298. [CrossRef]
66. Hurwitz, I.; Wolff, P.H.; Bortnick, B.D.; Kokas, K. Nonmusical effects of the Kodaly music curriculum in primary grade children. *J. Learn. Disabil.* **1975**, *8*, 45–52. [CrossRef]

education sciences

MDPI

Article

Teachers' Perspectives on the Intertwining of Tangible and Digital Modes of Activity with a Drawing Robot for Geometry

Anna E. Baccaglini-Frank [1,*], George Santi [2], Agnese Del Zozzo [3] and Eric Frank [4]

[1] Mathematics Department, University of Pisa, 56127 Pisa, Italy
[2] Faculty of Education, Free University of Bolzano-Bozen, 39042 Bressanone, Italy; gesanti@unibz.it
[3] Mathematics Department, University of Trento, 38123 Trento, Italy; agnese.delzozzo@unitn.it
[4] Department of Physiology, Tufts School of Medicine, Boston, MA 02111, USA; eric.frank@tufts.edu
* Correspondence: anna.baccaglinifrank@unipi.it

Received: 25 November 2020; Accepted: 16 December 2020; Published: 18 December 2020

Abstract: The GeomBot is a drawing robot that combines the well-known strengths and opportunities offered by Scratch with those of Papert's original robotic drawing-turtle. In this study we look at the GeomBot as a physical programmable artifact around which action research with a group of teachers and a researcher was carried out with the aim of designing, implementing and discussing geometry activities for primary school classes. The aim of this article is to investigate teachers' positioning and perspectives with respect to the activities and the educational environment emerging around the Geombot in the action research. The action research meetings between September 2018 and June 2019 included nine primary school teachers from seven different Italian schools, who met regularly with the first author. After the design and experimentation sessions, in June 2019, the teachers shared their experiences during a final meeting and group discussion guided by the first author. The data collected from the teachers were analyzed using cultural categories from the Semiotic Systems of Cultural Signification, theorized by the Theory of Objectification, to identify the most significant features defining the teachers' perspectives and identity. The seven emergent features cover teachers' positioning with respect to: forms of rationality and language and languages: sensuous cognition and the use of ideal and material semiotic resources; accepted teaching practices, problems and situations; mathematical knowledge; the conception of the student; social interaction and forms of rationality; ethical issues; technology.

Keywords: action research; drawing robot; GeomBot; geometry teaching; scratch-based programming

1. Introduction: The GeomBot as an Artifact for the Teaching and Learning of Geometry

This study lies at the crossroads of using artifacts in mathematics teaching and learning, embodied cognition in mathematics and action research. At their meeting point lies the artifact in focus: a drawing robot called the GeomBot, that was specifically designed and constructed by two of the authors for the teaching and learning of geometry at primary school.

There is a long tradition in mathematics education research concerning the implementation of material artifacts for mathematics teaching-learning activities in the classroom (e.g., [1]). In our opinion, a turning point in the study of the role of artifacts in processes of teaching and learning mathematics has been marked by research concerning semiotics. Indeed, the link between the use of artifacts and mathematical teaching and learning was established when studies in semiotics acknowledged what Arzarello [2] calls *outer enlargement*. He refers to Vygotskian sociocultural theories in mathematics education that allowed for the broadening of the notion of sign beyond Fregean classical semiotics to include gestures, objects, tools, linguistic devices etc. Among the most important

examples of outer enlargement are Arzarello's *semiotic bundle* [2], Bartolini Bussi and Mariotti's Theory of Semiotic Mediation [1], and Radford's Theory of Objectification [3]. Using artifacts in mathematical activities entails a multimodal approach to cognition that intertwines senses, emotions, material objects, bodily movements and the conceptual realm. Therefore, in order to fully understand where to position these in students' learning, we need to connect semiotic approaches with theories of embodied cognition in mathematics [4].

The GeomBot, the specific artifact introduced to the teachers in the action research activities we analyze in this study, together with the programming language used to give it commands from a computer, intertwines physical and digital modes of activity. On the one hand, the unique set of characteristics of the GeomBot, explicitly designed for the teaching of geometry, are new to the literature, as we will discuss. On the other hand, such characteristics make designing and implementing didactical activities with the GeomBot a potentially rich and novel experience even for teachers who have previously used digital artifacts in their classrooms, even other robots (we note that according to the definition proposed by Calmet, Hirtzig and Wilgenbus [5], a robot should contain sensors; the current prototype of the GeomBot does not have sensors, although they could easily be added. We choose to still refer to the GeomBot as a robot similarly to what is done in educational research.).

As we will point out in the next section, while there has been quite a bit of research on learning mathematics by programming physical or digital artifacts that respond to commands, much less attention has been paid to the teachers' perspective and to how their sensitivity towards specific design features of an artifact like the GeomBot may develop through many months of action research. Therefore, we chose to focus this study on how the GeomBot's design may shape teachers' interactions, identifying the most significant features that they noticed during an action research project that lasted 10 months long and involved nine primary school teachers and one researcher.

We will present our study as follows. Section 2 has two subsections: in Section 2.1 the empirical background is described, in Section 2.2 we outline the theoretical framework, and we present our research question; in Section 3 we describe materials and methods of our research; in Section 4 we present our results, organized in seven features, in order to answer to our research question; finally, Section 5 concludes the work reviewing the key points in the Discussion, and in Section 6 we present limitations and future perspectives of our work.

2. Empirical Background and Theoretical Foundation

To properly contextualize the GeomBot, we first present the empirical context related to programmable physical artifacts along with a brief presentation of the literature in mathematics education related to their study, so as to identify our potential contribution. Next, we present the theoretical foundations that we will refer to from Section 3 onwards, both for the description of the GeomBot's design and for the analysis of the data collected during the project. The research question, formulated with respect to these foundations, concludes Section 2.

2.1. Empirical Background

GeomBot's ancestor is Papert's drawing turtle shown in Figure 1. Papert saw his drawing turtle and the Logo environment used to program it as a microworld, an informal environment in which knowledge, applied to a specific discipline (in our case geometry), is incorporated. In this way, the objects and the modes of mediation of the microworld offer the user a direct encounter with the underlying mathematical objects. In many cases the interaction between the student and the microworld occurs in the school setting, and the role of the teacher is key, especially in the design of the activities.

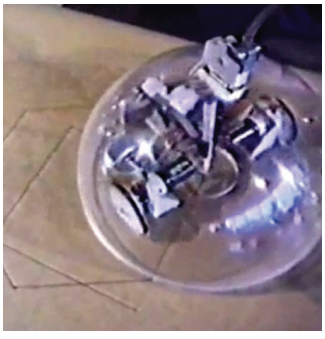


Figure 1. Papert's drawing turtle that was programmed using Logo; a screenshot from the YouTube video available at: https://youtu.be/BTd3N5Oj2jk.

Numerous studies in the 80s and 90s inspired by Papert's pioneering work [6,7] suggest educational benefits of learning geometry through Logo and Logo-based computer software [8–14]. This literature includes case studies about how the programming of Papert's robotic turtle can enhance the learning of mathematics, especially in the context of special education—e.g., [15,16]. More recently, studies have been carried out with young learners (ages 4–7) programming digital toys such as the Bee-bot, a very simple version of the original Logo robot turtle—e.g., [17–20]. These studies suggest that the experience of programming a robot that leaves a visible trace (we note that the Bee-bot does not have a marker-holder, but the teacher can "follow it" with a marker, to leave the trace of its path on paper, as was done in [21]) as it moves can be very powerful in fostering young students' productions of narratives and a rich variety of written and verbal signs—e.g., [21,22]. These experiences can be exploited by the teacher to foster and enhance mathematical learning. Indeed, in previous studies that focused specifically on achieving the transfer of learning from the Logo environment to Euclidean Geometry the role of the teacher has been recognized as fundamental—e.g., [14]. Moreover, being able to see the code in the form of blocks, as offered in Scratch or ScratchMaths, affords new representations (along with the drawings) of the geometrical objects and fosters further learning opportunities [23]. Benton, Hoyles, Kalas and Noss [23] identified a strong relationship between learning to program and learning to express mathematical ideas through programming in older students (ages 9–11).

In spite of Papert's original idea and in spite of the line of research outlined above, most of the activities with robots proposed in classrooms currently aim at introducing "programming" or "coding" per se, and the robots end up being "little more than toys made to drive to and fro or side to side without really providing an opportunity for any scientific exploration and learning" [24]. We believe that this is in part due to the fact that these robotic toys are not designed with an explicit intention for them to become tools for mathematical learning, in which coding is not introduced per se (we note that an acritical use of the word "coding", can hide some too simplistic ideas of what computational thinking and computer science education are; to deepen this delicate issue, from the perspective of computer science education, we recommend reading [25]), but as a canvas for expressing mathematical ideas, such as algorithms, equations, as well as geometrical properties of a figure. Furthermore, in spite of the acknowledgement of the key role of the teacher, to the best of our knowledge, there seem to be many more studies related to the students' learning than to teachers' perspectives on didactical practices in mathematics enhanced by physical programmable artifacts. Though studies like that of Reich-Stiebert, Eyssel and Khanlari [26,27] provide a focus on the teachers' point of view in using robots in teaching and learning, these are not contextualized in the field of mathematics education.

To sum up, our study should provide contributions in two directions. First, the design of the GeomBot itself should stand out as an example of a physical programmable artifact explicitly for the teaching and learning of geometry that combines the well-known potential of Papert's turtle with the

benefits of visual block programming languages. Second, the study contributes to bridging the gap in the literature on teachers' perspectives in approaching a new artifact such as the GeomBot.

2.2. Theoretical Foundation and Research Question

This study intertwines different dimensions embedded in a specific theoretical perspective. We will follow a sociocultural and historical approach drawing upon the Theory of Objectification [3] to look at the implementation of processes of mathematical thinking and learning fostered by using a drawing robot and the teachers' noticing of specific features of their work with this drawing robot.

2.2.1. Theory of Objectification and Sensuous Cognition

The Theory of Objectification is one of the most solid and acknowledged theoretical stances in mathematics education. Based on Vygotsky's and Leont'ev Cultural Historical Activity Theory [28,29], the Theory of Objectification goes beyond rationalist or individualist views of cognition reconciling the subjective and the objective, the sensual and the conceptual, the ideal and material. Cognition, revisited from a non-mentalist standpoint, is conceived as a sociocultural and historical practice, namely a praxis cogitans. Radford points out that "thinking is considered to be a mediated reflection in accordance with the *form or mode of the activity of individuals*" [3] (p. 218, emphasis in original). Conceptual objects, thinking, learning and meaning in mathematics are intertwined in reflexive mediated activity.

- Activity refers to the individual and social agency towards shared goals, significant problems, operations, labor etc.
- Reflexivity refers to the dialectical movement of the individual consciousness between his personal thinking, interpretations, emotions and feelings, perceptions and cultural-historical, marked by symbolic superstructures—in the terminology of the Theory of Objectification, *Semiotic Systems of Cultural Signification* [3]—that include systems of beliefs, systems of truth, forms of rationality, methods of inquiry, acceptable forms of knowledge, social rules, epistemology and ontology of mathematics etc. [3].
- Mediation refers to the role of artifacts in carrying out the praxis cogitans. In a Vygotskian sense, they are constitutive and consubstantial to thinking since they allow us to carry out activity. They are not mere aids to thinking but they shape the socio-cultural space, the *Territory of Artifactual Thought* [3] in which cognition occurs. They include a wide arsenal of semiotic resources ranging from material objects, gestures and bodily movements to icons, diagrams, natural language and symbolic language. The territory of Artifactual Thought molds the way we act according to the cultural and historical meanings they condense and are endowed with.

Learning is a specific praxis cogitans that Radford [3] terms a process of objectification. In its etymological meaning it refers to the process that allows the student to notice, find and encounter the cultural object. The artifacts that accomplish the objectification processes are called semiotic means of objectification [30] and they cover the whole range of ideal and material resources mentioned above. In Radford's [30] (p. 41, emphasis in original) words:

"These objects, tools, linguistic devices, and signs that individuals intentionally use in social meaning-making processes to achieve a stable form of awareness, to make apparent their intentions, and to carry out their actions to attain the goal of their activities, I call *semiotic means of objectification*."

The Theory of Objectification can be set into the strand of embodied cognition in mathematics—for an overview we refer the reader to [4]. Radford [4,31,32], resorting to a dialectic materialistic stance, conceives embodiment as a *sensuous cognition*, that is, a *multimodal sentient* form of responding to the world sprouting from cultural and historical activity. Cultural and historical activity intertwines, in sensuous cognition, senses, feelings, materiality, and the conceptual realm. The multimodality of the individual's response intertwines the manifold possibilities of human perceptions (sight, touch, sensorimotor activity, feelings) with the modes of activity realized by the variety of semiotic means of

objectification that cover the whole spectrum of human experiences, ideal and material, sensual and conceptual, subjective and objective.

The materiality of cognition is not something subsumed in the mind to acquire the nature of a concept, but the material is consubstantial to the conceptual. Senses, feelings, materiality and the conceptual realm culturally and socially develop into what [31] terms "highly sensitive cultural objects—*theoreticians*" [31] (p. 353, emphasis in original), in which the material and the ideal are tuned into the objectification of mathematical generality. The multimodal nature of cognition allows us to outline *levels of generality* [30] at which the student objectifies the mathematical concept. The level of generality specifies the blending of ideal and material in the process of objectification, according to the artifacts that realize the process of objectification:

- *Factual generalization*—characterized by perception, feelings, movement, spatial and temporal elements of the students' physical environment—is accounted for mainly by gestures, bodily movements, material objects and deictic and generative use of natural language.
- *Contextual generalization* intertwines material perception, movement and feelings with a new perceptual field in which emergent objects are detached from mediated sensory perception. Students start introducing more ideal semiotic means of objectification, such as new linguistic terms, natural language and the first elements of symbolic language.
- In *symbolic generalizations*, perception is no longer embedded or related to a concrete space-time context but in a new abstract and relational "space" where mathematical activity is carried out mainly by symbolic language.

2.2.2. Teachers' Action Research as a Process of Subjectification

The Theory of Objectification offers an effective lens to interpret teachers' collaboration [33]; it allows for the broadening of the notion of Community of Practice [34] that is a cornerstone in the action research experience. In particular, we want to analyze the teachers' positionings and perspectives with respect to technology and geometrical teaching and learning as they engage in collaborative activities with the drawing robot. We look at such a positioning through the lens of subjectification [35], the counterpart of objectification, related to the production of subjectivities as they engage in the reflexive mediated activity. If objectification pertains to the process of *knowing*, subjectification pertains to the process of *becoming*, that is, the changes and development of the individual. The Theory of Objectification outlines a dialectical co-production between individuals and their cultural and historical reality. Radford [35] (p. 43) conceives the individual as:

"[. . .] an entity in flux, in perpetual becoming—an entity who, through practical activity (like play) is continuously inscribing herself in the social world and, in doing so, she is continuously produced and co-producing herself within the limits and possibilities of her culture."

We would like to point out two fundamental features mentioned in the above quotation that define both the production of subjectivities (subjectification) and its entangled counterpart, that is, the production of knowledge (objectification): *agency* and the *cultural-historical context*. Agency and the cultural-historical context also live in a dialectical relationship where they continuously co-produce themselves. The cultural-historical context defines who we are, how we conceive of ourselves and how others conceive of us, the "fabric of our subjectivity" [35] (p. 46), the space, the constraints and the forms of our agentic maneuvering. This generative capacity of culture is termed by Radford [35] as *being*. If, on the one hand, we are existentially determined and generated by our cultural-historical context, on the other hand we hold a reflexive relationship with our cultural context, which means that "we *react agentically* to such a context" [35] (p. 4). *Being* materialized by the individual's deeds is termed by Radford [35] a process of *becoming*.

As we mentioned in the previous section, the cultural-historical context unfolds as a *symbolic superstructure*, the Semiotic Systems of Cultural Signification: a network of distinctive traits that makes up the fabric of a culture and its society. They emerge from individuals' reflective activity and they amount to dynamic systems, which are at the basis of a society's modes of meaning production.

Radford [3,35] singles out some of the fundamental Semiotic Systems of Cultural Signification that characterize mathematical thinking and learning. For the scope of our study we have focused on:

A. mathematical knowledge: geometry
B. language and languages: multimodality and semiotic resources
C. accepted teaching practices, problems and situations
D. conception of the student: affection, modes of learning
E. forms of rationality: sensuous cognition
F. social interaction
G. ethics
H. the role of technology in mathematical thinking and learning:

We are interested in analyzing subjectification as the critical positioning of the teachers involved in the action research as they discover new possibilities of moving, acting, feeling, perceiving, and thinking in the educational socio-cultural space that develops around the use of the drawing robot.

2.2.3. Research Question

In this study we address the following research question:

How do teachers position themselves with respect to the educational space emerging from designing and implementing classroom activities with the GeomBot?

In particular, what is the teachers' positioning with respect to the role of the GeomBot in fostering students' affective engagement, social interaction, cognition, geometrical knowledge and teaching practices?

3. Material and Methods

This section is comprised of several parts: in the first we present the participants; in the second a presentation of the study design is given; in the third we describe the design of the physical programmable artifact we consider, the GeomBot, outlining its potential to realize mathematics generalization as a form of sensuous cognition; in the fourth we present the implementations details of the action research around the GeomBot; and in the fifth we introduce the data collected.

3.1. Partecipants

Nine primary school teachers from the Group of Research and Experimentation in Mathematics Education of the Department of Mathematics (GRSDM) of the University of Pisa participated in this action research project, together with the first author of the paper, a mathematics education researcher. All the teachers had been members of the GRSDM for at least two years and had signed informed consent forms for all research carried out within their collaboration with the University. Moreover, many of the schools in which these teachers work have specific informed consent forms approved by their administration, which are signed each year by the parents of the students in their classes. In this paper we only include images from classrooms where such consent forms had been signed. The nine participants, all female, were from 7 different schools; and two pairs of them, Monica and Lynn, and Corinna and Lisa, worked at the same primary school building but at different grade levels (all teachers' names used in this paper are pseudonyms). All the participants were experienced teachers: they had been teaching in primary schools for between 10 and 35 years, though not all of them had been teaching mathematics for their whole career. Finally, five of the teachers were used to using interactive white boards regularly (Monica, Lynn, Fiona, Mary and Rebecca) and three of them had in previous cycles (so not with the children involved in this project) worked with robots such as the Bee-bot or the Pro-bot (Lynn, Fiona and Mary). The teaching habits of the nine teachers in this study were well-known to the researcher involved in the meetings with the teachers, because she had conducted other mathematics education projects with these teachers over the past

three years. The teachers' teaching habits in geometry involved the use of paper-and-pencil approaches, frequently accompanied by other physical experiences like paper folding or painting, and sometimes (though rarely) an interactive dynamic geometry applet.

3.2. Design of the Study

The data collected for this paper come from an action research project [36] that took place over an entire academic year between September 2018 and June 2019 through the different phases that are detailed in Section 3.4. The activities were carried out in the form of teachers' collaboration within a community of practice characterized by: a common domain of interest; social interaction between the members of the community; and a common practice. We conceived practice as a mediated reflexive activity realized with GeomBot. The purpose of such action research was to collaborate on the design, implementation and discussion of activities for teaching geometry with the GeomBot in primary school grades. All the teachers had been introduced to the GeomBot during the previous academic year and some of them had even used it in their classes for some pilot activities. The group used an online moodle platform to share information and materials. The experimental classes where the teachers implemented the activities were chosen based on teachers' availability and desire to participate, and with a preference towards those with cases of students with mathematics learning difficulties. Indeed, in each classroom in which the designed activities were implemented, there was at least one student with some form of mathematical learning difficulty—either reported by the teachers or certified in a clinical diagnosis of a specific learning disorder. These criteria also let us gather information on inclusive aspects of teaching and learning with the GeomBot.

3.3. Materials

The GeomBot uses an Arduino chip to activate two independent stepper motors connected to its wheels, which can be activated at the same or different speeds and in either direction, a buzzer, and LED lights that signal the different commands as they are executed. A marker-holder allows a marker to be placed between the wheels so that a trace mark is left on a sheet of paper (or whiteboard) on the floor, as shown in Figure 2. Most physical parts of the GeomBot are designed using 3D modelling software (SketchUp and OpenSCAD) and printed using a 3D printer. The wireless communicator (lower left-hand side of Figure 2a) uses a serial port connection on stable frequencies that do not interfere with one another when more than one robot is used in the same classroom. Finally, unlike other rechargeable robots that stop in the middle of an activity and need to be recharged, the GeomBot uses batteries that can be changed on the spot whenever they run down.

The following physical properties of the GeomBot's design make evident how the robot physically draws—a feature that can be exploited in multimodal activities entangling various geometrical meanings with physical sensorial experiences. Rubber rings around its wheels touch the ground in a small area (evoking a point); the drawing marker is inserted into a prism with a square base and the drawing tip of the marker comes out exactly at the square's center, touching the ground at the midpoint between the points of contact between the wheels and the ground, which is also the pivot point around which the GeomBot rotates in the basic "turn" commands (evoking the vertex of an angle) (Figure 2b). These features are designed so that the GeomBot's multimodality can be exploited to foster the generation of signs in relation with the geometrical meanings of point, distance, and angle. Moreover, the GeomBot can also be told to turn around one of its wheels (see the PIVOT commands in Table 2), in which case the marker rotates around the wheel that does not move (fixed point) drawing a circle or an arc (evoking the locus of points with a certain distance from a given fixed one).

So on the one hand, the physical properties mentioned above play a key role in materializing thought. Firstly, we note that the term "material" introduced in the Theory of Objectification has to be understood in the sense of dialectic materialism. Objects are not just lifeless "things", instead they are culturally related to the subject's praxis cogitans that endows them always with new life. On the other hand, we highlighted how the material properties of the GeomBot—intertwined with the ideal

ones—are designed to foster cultural development of the student's senses into theoreticians that allow them to notice and become aware of geometric concepts. These are not given per se but they are a cultural and historical way of perceiving, moving, feeling and thinking. Therefore, the design of the physical properties of GeomBot accomplishes the unity of the sensual and conceptual, of the material and ideal, of the subject and object.

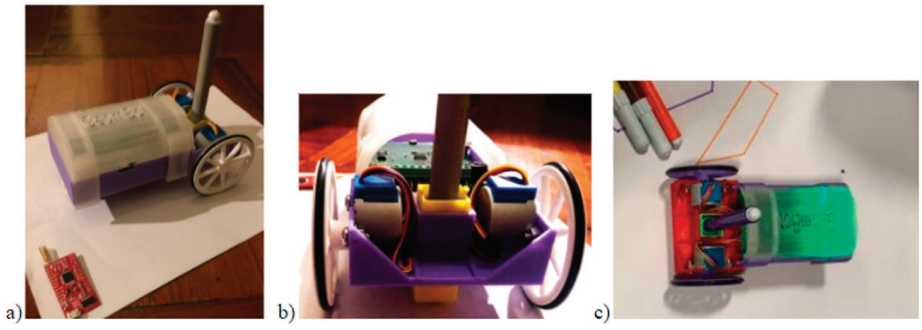


Figure 2. (**a**) The GeomBot and its wifi transmitter to be connected to a computer; (**b**) back view of the GeomBot showing the marker-holder with a marker touching the paper; (**c**) top view of the GeomBot in action.

To program the GeomBot and make it move and draw, we designed command blocks in a Scratch-based language (Tables 1 and 2). These command blocks are available in addition to Scratch's default set. The first set of blocks (Table 1) is the simplest, and is analogous to the commands that can be given to simple robots like the Bee-bot through push buttons. This analogy allows for meanings developed by many Italian students in preschool and kindergarten to be transferred to the GeomBot context in first grade. Moreover, the fact that when *talking* to the GeomBot, the commands, which are virtual blocks that can be manipulated on the screen, can foster *contextual generalization*: now sequences of commands are objects that can be manipulated and talked about, allowing their relationships with the movements and the drawn out figures to become more explicit and detached from immediate sensory perception.

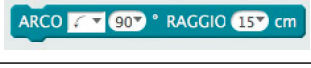
Table 1. First set of blocks designed for Grade 1 students.

Blocks	Command	What the GeomBot Does
PASSO ↑	step forward	moves 15 cm forward, flashes both LEDs green
PASSO ↓	step backward	moves 15 cm backward, flashes both LEDs red
GIRO S↶	90° turn counterclockwise	rotates 90° counterclockwise around the pen, flashes left LED blue
GIRO D↷	90° turn clockwise	rotates 90° clockwise around the pen, flashes right LED blue

Activities based on geometry can be designed to create the opportunity for students to gradually discover and make use of increasingly complex commands (such as those included in Table 2) with parameters that lead to new forms of movement and drawing. Indeed, these new commands can be used to explore new and more general ways—up to pre-symbolic generalizations—to describe,

notice and characterize the geometrical notions on which their teachers choose to focus during the proposed activities.

Table 2. Additional blocks we have designed in the Scratch-based environment.

Blocks	What the GeomBot Does
	moves forward any number of cm (input from dropdown menu or manual)
	moves backward any number of cm
	rotates counterclockwise around the pen any number of degrees
	rotates clockwise around the pen any number of degrees
	moves forward or backward (1st parameter) any number of cm (2nd parameter)
	rotates clockwise or counterclockwise (1st parameter) around the pen any number of degrees (2nd parameter)
	rotates counterclockwise around left wheel any number of degrees (there are 3 other PIVOT commands: the 4 commands allow rotations in both directions and around either wheel)
	moves along an arc clockwise or counterclockwise (1st parameter) of any number of degrees (2nd parameter) and any radius (3rd parameter)

We now provide an example on how the GeomBot has the potential to realize multimodal reflexive activities, thus including, in addition to the material properties of the artifact, ideal semiotic means of objectification. This potentially allows students to gradually move to higher levels of generalizations, towards what, in the light of the Theory of Objectification, we can call *sensuous generalization*. Let us consider the sequence in Figure 3. The same sequence allows the drawing of *any* rectangle with one fixed dimension (of 15 cm-"PASSO"), including a square (when the parameters of "PASSO with a parameter" are set to 15 as shown). Focusing on semiotic characteristics of an activity with the GeomBot leading, for example, to the execution of the programmed sequence in Figure 3a. The sequence contains 4 "PASSO" (step) commands and 4 "GIRO" (turn) commands, that correspond to the 4 sides and angles of the drawn square (Figure 4b). While the turn commands are all the same—and not generalizable because the commands with no parameters were chosen—the step commands are of two different kinds and they are used in an alternating pattern. This corresponds to generalizability: what now had been drawn out as a square, could become *any* rectangle with two sides of 15 cm and two sides of different lengths, as long as the two parameters in the sequence are set to the same value. There is an intricate interconnection between the meanings of the digital signs in the programmed sequence, the physical signs produced by the GeomBot's movement on the paper, and mathematical meanings, which constitute a rich multimodal activity. *Sensuous generalization* is realized by moving back and forth from factual to contextual generalizations, from tangible and digital; always pursuing the unity of senses, feelings, the material and the conceptual. This potentially allows the student to live a unity of senses, movement, feelings, the material and the conceptual.

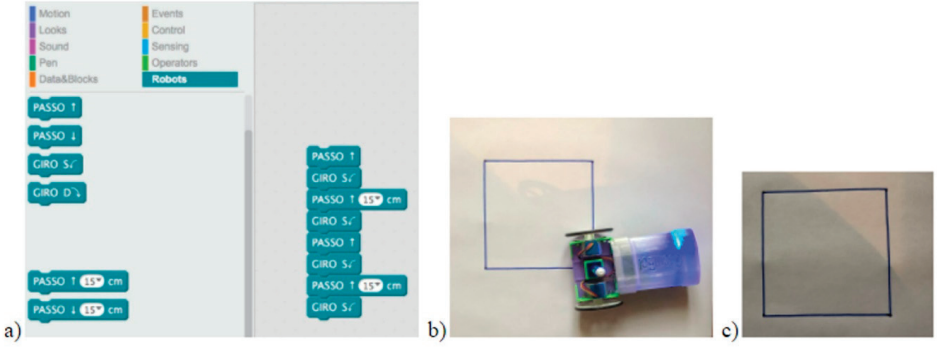


Figure 3. (**a**) GeomBot programming blocks (**left**) and a sequence for drawing any rectangle with one set dimension, including a square; (**b**) the GeomBot executing the sequence on paper; (**c**) drawing made.

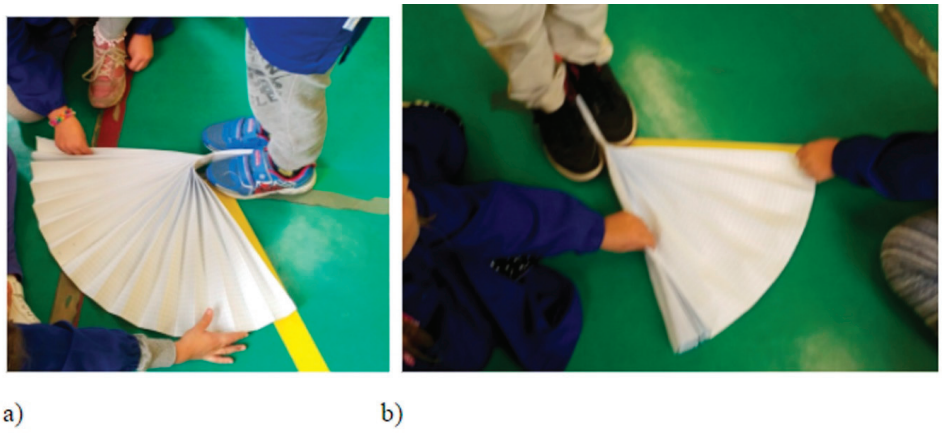


Figure 4. The "fan" artifact to highlight the "amount to turn" in two cases: (**a**) an external angle greater than a right angle; (**b**) an external angle that is acute. This photo comes from a presentation that a teacher Mary did during one of the monthly group sessions. Such choice suggests that the teacher felt it was important to document and share this new "fan" artifact with colleagues.

3.4. Implementation of the Action Research around the GeomBot

The action research comprised three different situations:

(1) a whole-group session that was carried out once a month at the University;
(2) the implementations of the activities, that teachers conducted autonomously with their selected classes between two consecutive monthly sessions;
(3) a conclusive whole-group session that was conducted by the first author. We now describe these three situations.

3.4.1. The Monthly Group Sessions

The first author of this paper and the nine teachers met each month for three hours at the University. During these sessions, the group worked both through collaborative design of the activities and through discussions triggered by the teachers' reports about the outcomes of their implementations. Specifically, once a certain designed activity had been implemented by one of the teachers in the classroom, this teacher would have 10 min at the next monthly session to share her results and observations with the others. The end of each presentation was always followed by a period of free discussion. At the

end of the whole action research experience, the group had produced 10 activities: some of them had been implemented in the classrooms, while other activities, in line with the cyclic nature of action research, were designed as variations or additions to activities that had been implemented.

3.4.2. The Experimental Phase between Two Consecutive Monthly Sessions

Between two consecutive meeting sessions at the University, some teachers implemented several of the designed activities in their classrooms. During this intermediate phase, several discussions and collaborative interactions occurred, always involving a pair of people—the researcher and one or two teachers, unlike what happened during the monthly group sessions. Discussions and exchanges of ideas between the researcher and the teachers occurred via email and focused on reaching an agreement on certain details of the activities. The researcher focused on these communications, trying to outline patterns and recurring themes among different teachers. In addition to the conversations involving teachers and the researcher, interactions between teachers occurred as well, based on the affinity of interests. These teachers' interactions were assiduous and intensive. As an example, the two teachers who teach in the same building, who both admitted to having a shared fear of technology, worked on the activities themselves, meeting at school, as if they were students, before implementing them in the classrooms.

3.4.3. The Conclusive Whole-Group Session

The last meeting of the action research experience took place in June 2019 at the University; the nine teachers and the researcher were present. Unlike the other monthly sessions, the discussion at this meeting was triggered by the researcher who had noticed some recurring themes raised in her private sessions with teachers between the group sessions. More precisely, the discussion was initiated by the researcher like this: "Most of you spoke about how the experience with the GeomBot led you to change your ideas about the notion of angle, both as a mathematical concept and in terms of how to teach it. Could we talk about this a bit more all together?" In the second part of the last session, the teachers were invited to meet with those who had worked with the researcher on a different project and share their experiences, using PowerPoint presentations if they wished.

3.5. Data Collected

In order to identify the teacher's positioning with respect to the educational space reified by the GeomBot, we collected various information. We include here the list of data on which we based our analysis.

- Materials prepared and used by the teachers to present their experience to the rest of the group (slides, photos and short videos). These materials provide information both on what happened in the classes and on what the teachers considered meaningful to report and share with the rest of the group.
- The notes and the recordings collected during the discussions in the monthly sessions. These documents provide information on thoughts, questions and proposals triggered by the reports of colleagues about the implemented activities.
- The emails exchanged between teachers and the researcher. These communications provide information on the teachers' questions, doubts and fears when they were preparing the implementation of a certain activity in the classroom.
- The email descriptions and mutual requests for support of the activities, exchanged among the teachers during the experimental intermediate phases. These descriptions provide information on the collaboration modalities carried out between the teachers with common interests when they were preparing to face the implementation of certain activities in their classrooms.
- The notes and the recordings collected during the conclusive session. These documents allow us to explore in depth and to better characterize the teachers' perspective regarding the notion of

"angle"—which was revealed to be a recurring theme among the whole action research experience for each of the nine teachers involved.

Since teachers typically report on something new to them and to their way of teaching, we reckon that the teachers' subjectification understood as their positioning in the educational space is effective to understand the impact of this distinct drawing robot on the teaching and learning of geometry in primary school. We did not consider any of the data collected to be more relevant than any other: all of it contributed to helping outline the teachers' perspective. To analyze the data, we used a thematic coding criterion based on the study's theoretical framework. We matched the data collected from the teachers with the network of symbolic superstructures that, according to Radford [3,35], characterize mathematical thinking and learning. We present the results of this analysis in the next section.

4. Results

In the following sections we describe the most significant positionings of the teachers that emerged from the data analysis, categorized according to the networking of the Semiotic Systems of Cultural Signification listed above (Section 2.2.2, list from A to H).

4.1. Feature 1 (Aspects B, E). Positioning with Respect to Forms of Rationality and Language and Languages: Sensuous Cognition and the Use of Ideal and Material Semiotic Resources

The teachers were familiar with using the body in various mathematical learning activities involving paper-and-pencil and, possibly, other physical artifacts; however, they commented on how in the activities with the GeomBot using one's body to imitate the robot's movements or to pretend to be the robot given a program was a key step in linking concrete signs (e.g., the drawn paths) to more abstract signs (e.g., the sequences programmed in the Scratch-based environment). The nature of the activities with the GeomBot fostered this in a way that was more natural than with other activities. Monica and Lynn explicitly compared the activities with the GeomBot, which they described as a "sort-of-digital tool", with those they had experienced in previous years with dynamic geometry applets (for a description of these activities see [37]): while use of the body could be fostered and exploited by the teacher in the latter context, Lynn said that with the GeomBot "the body becomes a natural and unavoidable link between the concrete and the abstract". Moreover, the teachers took advantage of the natural involvement of the body to enhance students' learning by highlighting specific gestures (like the action of opening arms and holding them parallel to the ground, with one pointing in the original direction and one opening to the direction reached at the end of a programmed rotation) or by designing new artifacts like an opening "fan" to place between students' feet to highlight the "amount to turn" (see Figure 4).

The teachers' recognition of the role of the body when using GeomBot is a key feature in order to consider it an artifact that realizes sensuous mathematical generalizations. A further strength that emerges is that the Geombot truly combines the material and the ideal, because students are able to move back and forth from the sensorimotor to the conceptual. This constitutes evidence that the levels of generalization are not compartmentalized, as unfortunately can happen when students reach higher levels of generality. Typically in algebra, we testify an "evaporation of meaning" [38] and pseudostructural approaches [39] that lead students to use symbols unrelated to any mathematical meaning. The use of other artifacts such as the opening of the arms and the "fan" (Figure 4) to objectify the angles shows Geombot's potential to attract other semiotic means of objectification.

4.2. Feature 2 (Aspect C). Positioning with Respect to Accepted Teaching Practices, Problems and Situations: Solving Real-Life Problems

Teachers used the words "real-life problems" to describe the kinds of activities they had experienced in their classrooms with the GeomBot, and they referred to "mathematical competence" (we note that this terminology is coherent with that used in the documents by the Italian National Institute

for Assessment of the Educational and Professional Development System (INVALSI) which writes the Italian national assessments, inspired by the OECD competency framework; see, for example, https://www.oecd.org/careers/competency_framework_en.pdf) and "problem solving skills" to describe students' learning that they thought went beyond the construction of the mathematical knowledge described in the design of each activity. Two teachers Mary and Rebecca insisted that the GeomBot activities were "more real-life" than other activities, including those they had experienced previously with dynamic geometry. When encouraged to provide examples, the teachers noticed that the GeomBot produces "real" drawings "with few imprecisions if any at all", and that students are faced with situations in which they know where they need to get but they have not been taught a procedure to get there, which is "true problem solving" (we note that many teachers in our research group have actively participated in years of professional development at the University of Pisa under the guidance of prof. Rosetta Zan and Pietro Di Martino, who frequently quote [40] in defining a problem for a subject as a situation in which the goal is clear to the subject, but not how to achieve it). As they worked, students had to react to the feedback they received from the robot, plan again or debug, and compare what they got with what they wanted. A recurring example in the teachers' comments was related to a group activity in which students in one group had to communicate with another group to achieve the same size of the figures drawn by each other. Indeed, the teachers highlighted that such a need for communication led students to figure out useful ways of communicating.

Moreover, some teachers noticed their students' different approaches to problem solving. Although most teachers reported "trial-and-error" as the most common approach, which involved students working on one piece of a drawing at a time and responding and adapting to the GeomBot's feedback, Lynn, Mary and Fiona, with classes of Grade 4 or Grade 5 students, mentioned how students who already knew Scratch (and how to use the "repeat" command) and quickly became acquainted with the GeomBot would approach problems differently. These students would plan out the commands for the entire drawing and, if necessary, make modifications that were "more general". For example, Lynn reported to the group that a pair of students in her class wrote a program for drawing a square with sides of three steps but also with "different size squares" as shown in Figure 5a.

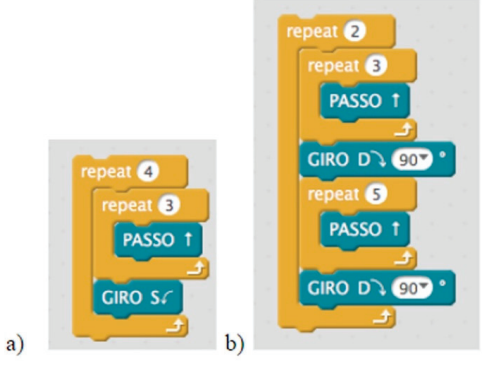


Figure 5. (**a**) program produced by a student of Lynn's for drawing a square with side of three steps but also "different size squares"; (**b**) program produced by a student of Fiona for drawing a rectangle with sides of three and five or "any rectangle", in the students' words. These screenshots come from the presentations of the teachers' reports during one of the monthly group sessions. Such choices suggest that the teachers felt it was important to document and share the code that their students wrote.

In Fiona's class, one student said she could make the GeomBot draw "any rectangle" with a program shown in Figure 5b and she modified the lengths of the sides as necessary to help students in other groups. The teachers admitted to being quite pleasantly surprised by these students' ideas that exceeded their expectations.

What the teachers call real-life problems in the frame of Theory of Objectification can be interpreted as the social and cultural dimension of reflexive mediated activity comprising thinking, teaching and learning. This is an important element to allow students to live mathematics as belonging to their personal experience, cognitively, socially and emotionally. The acquaintance of some students with effective and general solving strategies confirms that they are experiencing sensuous cognition. They are therefore able to maintain contact with meaningful mathematical activity and to move effectively across different levels of generality, in order to "geometrically" see and feel the best path towards the solution.

4.3. Feature 3 (Aspect A). Positioning with Respect to Mathematical Knowledge: Ease in Engaging with the Specific Geometry at Stake

The teachers' experience with the GeomBot led them to discover new ways of teaching specific geometrical notions, and, sometimes, to learn about a completely new notion. For example, for the notion of "angle", Corinna, Lisa, Lynn, Rebecca and Rita confessed to having previously defined it as "the intersection of two half-planes", or as "the part of the plane contained between two rays with a common origin". These definitions are mathematically difficult for students (and many teachers!) to understand—e.g., [41]. Moreover, all nine teachers noticed that they had previously only emphasized the "internal" angles of polygons. Instead, the activities with the GeomBot naturally led to defining angles as rotations and to seeing and talking about "internal" and "external" (meaning "supplementary", that is the angle that together with the first gives an angle of 180 degrees) angles: in a task where the drawing of a polygon is required, to get an "internal" angle of n degrees the GeomBot has to be programmed to rotate by an angle of $(180 - n)$ degrees. Three teachers agreed that the GeomBot provided a "friendly excuse" to go back and study or refresh important geometrical concepts.

The teachers also commented on how programming the GeomBot to draw geometrical figures exposes different geometrical properties and relationships between them with respect to those exposed by a software for dynamic geometry. This observation is perfectly in line with the analysis of the GeomBot in Section 3.3: it reinforces the crucial importance of a deep reflection upon the semiotic aspects involved in each activity. For example, Lynn reported that in her class students programmed a square by giving the GeomBot the commands "repeat 4 times: 3 steps forward, turn right" (as shown in Figure 5a), which highlights the four congruent sides and four right angles. However, the teachers noticed that in a dynamic geometry environment it is impossible to explicitly construct all four congruent sides and right angles of a square: if three consecutive congruent and perpendicular sides are constructed, the fourth side is completely defined and, when constructed, it "magically" is congruent to the other three sides and perpendicular to its two consecutive sides. These properties are derived from the other construction properties. This behavior of dynamic geometry software, that can also be exploited with older students, is too advanced for primary school children who are only beginning to discover and discuss geometrical properties of 2D figures.

In general, the teachers agreed that watching the GeomBot move (relatively slowly) and draw on paper provided greater ease for both themselves and their students to engage with the mathematics being studied. The teachers' recognition of the ease to engage with the mathematics at stake when using the GeomBot is an important finding. It confirms how, within the frame of the Theory of Objectification, artifacts are not merely inert "things", but they are bearers of cultural and historical mathematical activity that allow for the multimodal materialization of thinking in mathematics. Apparently, the teachers objectified new forms of mathematical meaning through the multimodality of the GeomBot.

4.4. Feature 4 (Aspect D). Positioning with Respect to the Conception of the Student: Affective Dimension for Teachers and Students

Teachers repeatedly mentioned how their students had strong positive emotional reactions to the GeomBot. These became visible on many occasions: for example, they became worried and wanted to "save GeomBot" if it was about to wander off the drawing paper or, worse, off the top of a table

(indeed, the robot was likely to break if it fell off a table). They also wanted to design clothing for it, especially in Monica and Lynn's classes. Children used diminutive and affective forms, such as "robottino" (translated "little robot"), to refer to it; they always gave it a gender (male in seven classes and female in two); and in some cases the children insisted on giving it a name (GeomBot was the name given to it by the children in one of the classrooms). Some children interpreted GeomBot's writing as "messages" to them, especially after an introductory activity conducted in some classes that ended with the GeomBot drawing a heart to "say goodbye" (see Figure 6). Children also talked about needing to "teach him to write letters" or to "learn to draw precise figures".

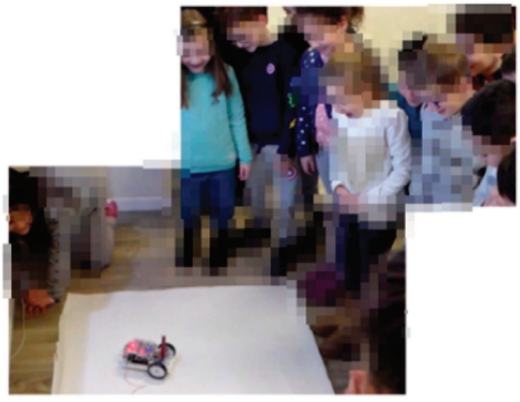


Figure 6. Children in a Grade 1 classroom extremely excited as the GeomBot said "goodbye" with a heart.

Although the teachers did not explicitly mention this, we noticed that they themselves would frequently treat the GeomBot as if it were alive, using their students' ways of talking about or to the GeomBot. They would say "bravo!" to it when it succeeded, shout "be careful!" if it was about to move off the sheet of paper, use diminutive forms like "robottino" or "poverino" (translated "little robot" and "poor little thing"), and assign it a gender (usually male, probably because "robottino" is masculine in Italian). These behaviors indicate a positive emotional connection with the GeomBot on behalf of the teachers as well.

We believe that the affective dimension related to the use of the GeomBot cannot be taken for granted. The material nature per se cannot account for triggering an affective involvement on the part of the students (and the teachers). Instead, this result can be drawn from the material nature of the artifact whose cultural, historical and social features establish in reflexive activity also an affective relationship between the individual and the object. This lack of the affective dimension would amount neither to objectification nor to sensuous cognition nor to cognition in general [41].

4.5. Feature 5 (Aspects E, F). Positioning with Respect to Social Interaction and Forms of Rationality: Shared Reflexive Activity and Developing the Geometrical Eye

The activities with the GeomBot in this study were designed during the action research project to involve work in small groups and whole class discussions about predicting what a programmed sequence will make the GeomBot draw and how to "fix" or debug a sequence that does not create the expected drawing (Figure 7). Students can discover that the same figure can be drawn in different ways, and thus "solved" through different sequences, which promotes problem solving skills and creativity. Students may feel empowered because their solutions do not need to be validated by the teacher; instead they can do this on their own by interacting with the GeomBot and interpreting its movements and the drawings it creates. In this sense, the use of the GeomBot in this kind of activity is extremely significant in developing the senses into *geometrical theoreticians*. The multimodality of the GeomBot constantly molds the students' "raw" sensuous perception into a geometric *theoretical* one. For example,

predicting what a programmed sequence will make the GeomBot draw requires the student to see a figure or a path that is not present in his/her space-time perception but can be accounted for in the geometrical one, a geometrical perception that s/he has built in previous activities or that s/he will build by comparing what s/he predicted with what the robot actually draws. When a programmed sequence of commands (a set of blocks on the computer screen) is clicked, the students can see the robot move as it executes the programs and traces them out on large sheets of paper or on a whiteboard on the floor (Figure 7b). Finally, the teachers noticed how the GeomBot would enhance social interaction, a constitutive key feature of reflexive mediated activity. Amongst the several possibilities of joint action most teachers asked students to work on sequences of commands in small groups and then, one at a time, to come program their sequences on a screen visible to the whole class, to collectively generate meanings, compare and share Scratch sequences, and discuss the resulting physical drawing on paper.

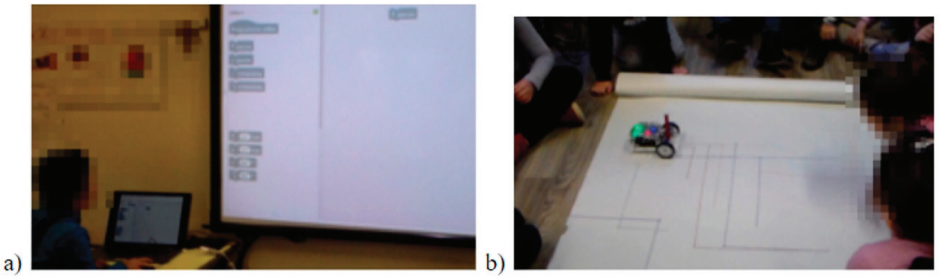


Figure 7. (**a**) A first grader programming of the GeomBot in the Scratch-based environment; (**b**) the GeomBot executing a programmed sequence.

The effective presence of the social dimension was highlighted by all the teachers of the group as they reported the experimental phase in their classes.

4.6. Feature 6 (Aspect G). Positioning with Respect to Ethical Issues: The GeomBot's Inclusive Nature

Mary, Sarah, Rita and Fiona mentioned an enhanced inclusiveness fostered in their classrooms by the GeomBot. This was an issue that arose during the final collective discussion as well. All classes of the nine teachers contained at least one student with special educational needs. Although we were not able to conduct case studies to investigate the learning processes of these students, the teachers commented on how the activities were easy to carry out in an inclusive setting and on how engaging they were generally for all students. A problem some teachers encountered was the lower attention levels of some students during the whole-class parts of some of the activities. Indeed, having only one GeomBot available in each classroom, with the students gathered on the floor around the GeomBot and in front of the computer screen projection where the programming occurred, led to a decrease in concentration of some students. This problem was solved by five of the teachers by having the class work in smaller groups at different learning stations and taking turns programming the GeomBot.

4.7. Feature 7 (Aspect H). Positioning with Respect to the Role of Technology

Teachers reported on a "cost" that comes with all the opportunities discussed: the time and energy that they must invest in becoming acquainted and at ease with the physical and digital aspects of the GeomBot. Almost all the teachers, including those who commonly use digital artifacts in their classrooms, mentioned feeling slightly frightened (unlike their students!) by the robot and by the Scratch-based programming environment: something might not work immediately and precious time that should be spent with the students might be wasted working through these issues; the robot might not perform as expected (initially some teachers did not recognize its behavior when the batteries were low); the surface on which the paper and the GeomBot are placed needs to be very flat, otherwise the drawings become less precise. Nevertheless, we note that this shared fear itself triggered some powerful

forms of collaboration such as the ones we described in Section 3.4 Moreover, teachers quickly learned to deal with the practical and digital issues that arose, and they repeated during the concluding session how useful it was to collaborate with a teacher from the same school and "rehearse" before going to class, when this was possible.

4.8. Answer to the Research Question

Let us recall the research question guiding this study: how do teachers position themselves with respect to the educational space emerging from designing and implementing classroom activities with the GeomBot? In particular, what is the teachers' positioning with respect to the role of the GeomBot in fostering students' affective engagement, social interaction, cognition, geometrical knowledge and teaching practices?

The data we have collected shows that the Geombot intertwines digital and tangible modes of activity fostering the emergence of a rich socio-cultural environment where objectification and subjectification processes may occur. Based on the eight symbolic superstructures (Section 2.2.2, list from A to H) that define the space where the teaching and learning of geometry take place, our analyses led to seven features that characterize the teachers' positioning with respect to the educational space involving mathematical practices with the GeomBot. The seven features cover the basic elements that amount to a significant and robust learning of mathematics: mathematical contents, language and languages, accepted teaching practices, problems and situations, conception of the student, forms of rationality, social interaction, ethics, technology. The teachers bring new possibilities of thinking and action that testify a change of perspective towards the teaching and learning of geometry.

5. Discussion

In this final section, we discuss the identified features to pinpoint how the GeomBot provides a unique combination of mathematics learning opportunities for students and teachers. In particular, we describe how the change in teachers' perspective allows for a fruitful implementation of the Geombot in students' objectification of geometry, testifying a dual nature of this physical programmable artifact in realizing a learning environment and allowing a critical positioning of the teachers.

A common denominator across the seven features that we identified is the multimodal nature of the drawing robot: indeed, the characteristics of the educational space and the subsequent teachers' positionings all pivot around this. The GeomBot intertwines tangible and digital modes of activities, thus realizing the entanglement of the ideal and the material that defines sensuous cognition. We believe that the intertwining of the tangible and the digital opened the way for the unfolding of the variety of features that allowed for a significant change in the teachers' identity during their action research experience with Geombot.

One aspect that was somewhat unexpected, but extremely important from our point of view, is the issue that emerged about how working with the GeomBot led to new mathematical considerations for the teachers, and therefore, very likely, to different mathematical learning of their students. For example, when discussing the activities that involved the GeomBot moving on a grid, the teachers noticed how the choice of making the GeomBot trace out its path starting from inside a square and moving from inside of one square to the next, was problematic: the 15 cm segments corresponding to one step almost never started or finished in the centers of the squares on the grid. Moreover, when counting the number of steps it took the GeomBot to finish its walk (and thus measure the length of the path) many students would get confused about whether they were counting "squares" (so "half squares" were necessary) or "steps". These considerations led the group to prefer making the GeomBot draw on the lines of the grid instead of inside the squares, and to modify the activities accordingly for future implementations. We find this choice to be coherent with higher level mathematics that involves, for example, coordinate systems and the Cartesian plane. Never before had the teachers thought of similar considerations, and they had never questioned positioning robots (like the Bee-bot) inside squares, as suggested by many common preschool and early primary school activities. Therefore,

the GeomBot contributed to pushing the senses towards new forms of geometrical perception, developing them into theoreticians in the sense implied by the Theory of Objectification.

With respect to the notion of angle as rotation, we realized that perhaps the Geombot's design could be ameliorated, so that teachers would not feel the need to introduce additional artifacts such as the "fan" to guide students' learning. After carefully analyzing the difficulties encountered by the teachers in approaching this geometrical notion, the first author has changed the design of the GeomBot so that it holds a second marker in front of the caster. This allows for a double trace (potentially with markers of different colors) to be left by the GeomBot, which becomes particularly significant during rotations centered between the wheels: the "back" marker stays in one point, while the front marker draws an arc sweeping the rotation. This was the most significant feedback (together with the students' proposal of "GeomBot" as the name of the robot) leading to modifications to the robot's design. Future didactical experimentations will reveal the value of this change in the GeomBot's design.

A second mathematical consideration that emerged was the teachers' noticing and giving value to students' different ways of thinking; in particular they seemed to re-valuate "trial-and-error" strategies, which in other mathematical contexts are not valued by teachers. This reminded us of Turkle and Papert's [42] thoughts on different ways of thinking and their re-evaluation of what they referred to as the concrete "bricoleur's approach" that involves conversations with the machine, as opposed to the "planner's approach" which corresponds to a more abstract way of thinking that traditionally was considered more advanced. The teachers, however, also quickly noticed and praised the ways of thinking that they considered to be "more general", as we noted in "feature 2". These changes in what the teachers considered to be mathematically relevant (which also led them to engage in deep mathematical discussions during the meetings) and valuable in terms of mathematical thinking may have been fostered by the physicality of the GeomBot, which made the situations very "approachable" and less threatening both for them and for the children. Indeed, the forms of sensuous cognition fostered by the GeomBot seem to facilitate this phenomenon. This leads us to a second issue that uniquely characterizes learning with a digital tool that also has physical modes of mediation like the GeomBot. The emergence of the new mathematical learning opportunities mentioned above may also have been initiated by the strong and positive emotional response of the students and the teachers to the GeomBot, as described in "Feature 4". This affective dimension fostered by body syntonic and ego syntonic [7] characteristics of the GeomBot appears to be much stronger when compared to other digital tools such as dynamic geometry applets (see, for example, [37]). The GeomBot was referred to by the teachers, who picked up on their students' words, as something "alive" with human or animal-like features to whom students assign a gender (in most cases male, in some female). It would be interesting to see the students' behaviors in languages where there is a neutral form—like "it" in English—to figure out whether there are differences.

A third issue, somewhat marginal in this paper, but extremely important from an educational perspective, is that of inclusiveness. Learning activities that involve friendly, moving robots are known in the literature for their great potential in providing more learning opportunities for students with different learning preferences or needs—e.g., [15,16,43,44]. Although this issue was mentioned in the "feature 5" and noticed by the teachers, more research and dedicated studies are definitely necessary.

6. Limitations and Directions for Future Research

The small number of participants is surely a limitation to the generalizability of our findings; however, for this kind of qualitative research and the amount of data collected, it can be considered a satisfying sample number. We believe that the major limitation to generalizability of this study is the composition of the group of participants to the action research project. Indeed, the participants in our sample were relatively experienced both as teachers and as participants in action research projects offered within the GRSDM at the University of Pisa. An important difference that this might have made is the quality of the documental data collected within this project. Moreover, the teachers' experience

in designing and experimenting with learning activities is likely to have positively influenced the quality of the activities themselves and of the identification of parts that needed to be ameliorated.

One direction to explore in future research is the extent to which the study can be empirically replicated. A second direction, that we are quite keen on, involves opening a dialogue between communities of educators who conduct research in mathematics education, like ourselves, and research in computer science education, respectively. At the moment there seems to be a very limited (if existent at all) dialogue between the communities and this seems to negatively impact most of the classroom activities that involve programming physical or digital robots.

7. Patents

Currently there is a patent pending on the GeomBot, request number 102019000018254, deposited 8th October 2019.

Author Contributions: Conceptualization, all authors; Methodology, A.E.B.-F., A.D.Z., G.S.; Software, E.F.; Design of the GeomBot, E.F., A.E.B.-F.; Investigation and Resources, A.E.B.-F.; Writing—Original Draft Preparation, A.E.B.-F., A.D.Z., G.S.; Writing—Review & Editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We wish to acknowledge the active collaboration on the development of the motherboard and wireless communication between the laptop and the GeomBot of David Ferster, without whom this study would not have been possible. We also wish to thank the teachers of the GRSDM of the University of Pisa, who took part in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bartolini Bussi, M.G.; Mariotti, M.A. Semiotic mediation in the mathematics classroom. In *Handbook of International Research in Mathematics Education*, 2nd ed.; English, L., Kirshner, D., Eds.; Routledge: New York, NY, USA, 2008; pp. 746–783.
2. Arzarello, F. Semiosis as a multimodal process. *Rev. Lat. Am. Investig. Mat. Educ.* **2006**, *9*, 267–299.
3. Radford, L. The ethics of being and knowing: Towards a cultural theory of learning. In *Semiotics in Mathematics Education: Epistemology, History, Classroom, and Culture*; Radford, L., Schubring, G., Seeger, F., Eds.; Sense Publishers: Rotterdam, The Netherlands, 2008; pp. 215–234.
4. Radford, L.; Arzarello, F.; Edwards, L.; Sabena, C. The multimodal material mind. In *First Compendium for Research in Mathematics Education*; Cai, J., Ed.; NCTM: Reston, VA, USA, 2017; pp. 700–721.
5. Calmet, C.; Hirtzig, M.; Wilgenbus, D. *1,2,3...Codez! Enseigner L'informatique à L'école et au Collège*; Editions Le Pommier: Paris, France, 2016.
6. Papert, S. Teaching Children Thinking. *J. Struct. Learn.* **1975**, *4*, 219–230. [CrossRef]
7. Papert, S. *Mindstorms: Children, Computers, and Powerful Ideas*; Basic Books: New York, NY, USA, 1980.
8. Clements, D.H. Longitudinal Study of the Effects of Logo Programming on Cognitive Abilities and Achievement. *J. Educ. Comput. Res.* **1987**, *3*, 73–94. [CrossRef]
9. Hoyles, C.; Sutherland, R. *Logo Mathematics in the Classroom*; Routledge: London, UK, 1989.
10. Hoyles, C.; Healy, L.; Sutherland, R. Patterns of discussion between pupil pairs in computer and non-computer environments. *J. Comput. Assist. Learn.* **1991**, *7*, 210–228. [CrossRef]
11. Hoyles, C.; Noss, R. *Learning Mathematics and Logo*; MIT Press: Cambridge, MA, USA, 1992.
12. Clements, D.H.; Meredith, J.S. *Turtle Math*; Logo Computer Systems: Montreal, QC, Canada, 1994.
13. Noss, R.; Hoyles, C. *Windows on Mathematical Meanings*; Kluwer Academic Publishers: Amsterdam, The Netherlands, 1996.
14. Clements, D.H.; Battista, M.T.; Sarama, J. Logo and Geometry. *J. Res. Math. Educ. Monogr. Ser.* **2001**, *10*. [CrossRef]
15. Atkinson, B. Learning disabled students and Logo. *J. Learn. Disabil.* **1984**, *17*, 500–501. [CrossRef] [PubMed]
16. Michayluk, J.O.; Saklofske, D.H. Logo and special education. *Can. J. Spec. Educ.* **1988**, *4*, 43–48.

17. Highfield, K.; Mulligan, J.T. Young children's embodied action in problem-solving tasks using robotic toys. In Proceedings of the 33rd conference of the International Group for the Psychology of Mathematics Education, Thessaloniki, Greece, 19–24 July 2009; Tzekaki, M., Kaldrimidou, M., Sakonidis, H., Eds.; PME: Thessaloniki, Greece, 2009; Volume 2, pp. 273–280.
18. Highfield, K. Robotic toys as a catalyst for mathematical problem solving. *Aust. Prim. Math. Classr.* **2010**, *15*, 22–27.
19. Sinclair, N.; Baccaglini-Frank, A. Digital technologies in the early primary school classroom. In *Handbook of International Research in Mathematics Education*, 3rd ed.; English, L., Kirshner, D., Eds.; Taylor Francis; Routledge: New York, NY, USA, 2016; pp. 662–686.
20. Ronsivalle, G.B.; Boldi, A.; Gusella, V.; Inama, C.; Carta, S. How to implement educational robotics' programs in Italian schools: A brief guideline according to an instructional design point of view. *Technol. Knowl. Learn.* **2019**, *24*, 227–245. [CrossRef]
21. Bartolini Bussi, M.G.; Baccaglini-Frank, A. Geometry in early years: Sowing the seeds towards a mathematical definition of squares and rectangles. *ZDM Math. Educ.* **2015**, *47*, 391–405. [CrossRef]
22. Bartolini Bussi, M.G.; Baccaglini-Frank, A. Using pivot signs to reach an inclusive definition of squares and rectangles. In *Proceedings of the 9th Congress of European Research in Mathematics Education (CERME 9)*; Krainer, K., Vondrová, N., Eds.; Charles University in Prague, Faculty of Education: Prague, Czech Republic, 2016; pp. 1891–1897.
23. Benton, L.; Hoyles, C.; Kalas, I.; Noss, R. Bridging Primary Programming and Mathematics: Some Findings of Design Research in England. *Dig. Exp. Math. Educ.* **2017**, *3*, 115–138. [CrossRef]
24. Cunningham, C.H. No, Engineering Isn't Just about Job Creation. *Educ. Week Spotlight* **2019**, *7*. Available online: https://www.edweek.org/ew/articles/2019/02/20/no-engineering-isnt-just-about-job-creation.html (accessed on 20 May 2019).
25. Corradini, I.; Lodi, M.; Nardelli, E. An investigation of Italian primary school teachers' view on coding and programming. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*; Pozdniakov, S.N., Dagienė, V., Eds.; Springer: Cham, Switzerland, 2018; pp. 228–243.
26. Reich-Stiebert, N.; Eyssel, F. Robots in the classroom: What teachers think about teaching and learning with education robots. In *International Conference on Social Robotics*; Agah, A., Cabibihan, J.-J., Howard, A.M., Salichs, M.A., He, H., Eds.; Springer: Cham, Switzerland, 2016; pp. 671–680.
27. Khanlari, A. Teachers' perceptions of the benefits and the challenges of integrating educational robots into primary/elementary curricula. *Eur. J. Eng. Educ.* **2016**, *41*, 320–330. [CrossRef]
28. Leont'ev, A.N. *Activity, Consciousness and Personality*; Prentice Hall: Englewood Cliffs, NJ, USA, 1978.
29. Leont'ev, A.N. *Problems of the Development of the Mind*; Progress: Moscow, Russia, 1981.
30. Radford, L. Gestures, Speech and the Sprouting of Signs: A Semiotic-Cultural Approach to students' Types of Generalization. *Math. Think. Learn.* **2003**, *5*, 37–70. [CrossRef]
31. Radford, L. Towards an embodied, cultural, and material conception of mathematics cognition. *ZDM* **2014**, *46*, 349–361. [CrossRef]
32. Radford, L. Sensed objects, sensing subjects: Embodiment from a dialectical materialist perspective. In *The Body in Mathematics*; Edwards, L., Krause, C., Eds.; Sense/Brill: Rotterdam, The Netherlands, in press.
33. Ferretti, F.; Gambini, A.; Santi, G. The Gestinv Database: A Tool for Enhancing Teachers Professional Development within a Community of Inquiry. In *Proceedings of the Twenty-fifth ICMI Study School Teachers of Mathematics Working and Learning in Collaborative Groups*; Borko, H., Potari, D., Eds.; University of Lisbon: Lisbon, Portugal, 2020; pp. 621–628.
34. Wenger, E. *Communities of Practice: Learning, Meaning, and Identity*; Cambridge University Press: Cambridge, UK, 1998.
35. Radford, L. Play and the production of subjectivities in preschool. In *Mathematics education in the early years. Results from the POEM4 conference 2018*; Carlsen, M., Erfjord, I., Hundeland, P.S., Eds.; Springer: Cham, Switzerland, 2020; pp. 43–60.
36. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*; Routledge: London, UK, 2007.
37. Baccaglini-Frank, A.; Di Martino, P.; Sinclair, N. Elementary school teachers' implementation of dynamic geometry using model lesson videos. In *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education*; Bergqvist, E., Österholm, M., Granberg, C., Sumpter, L., Eds.; PME: Umeå, Sweden, 2018; Volume 2, pp. 99–106.

38. Arzarello, A.; Bazzini, L.; Chiappini, G. A model for analyzing algebraic process of thinking. In *Perspectives on School Algebra*; Sutherland, R., Rojano, T., Bell, A., Eds.; Kluwer Academic Publisher: Dordrecht, The Netherlands, 2001; pp. 61–82.
39. Sfard, A.; Linchevski, L. The Gains and the Pitfalls of Reification: The Case of Algebra. *Educ. Stud. Math.* **1994**, *26*, 191–228. [CrossRef]
40. Duncker, K.; Lees, L. On problem solving. *Psychol. Monogr.* **1945**, *58*, 1–113. [CrossRef]
41. Smith, C.P.; King, B.; Hoyte, J. Learning angles through movement: Critical actions for developing understanding in an embodied activity. *J. Math. Behav.* **2014**, *36*, 95–108. [CrossRef]
42. Turkle, S.; Papert, S. Different ways people think: Epistemological pluralism and the revaluation of the concrete. *J. Math. Behav.* **1991**, *11*, 3–33.
43. Baccaglini-Frank, A.; Antonini, S.; Robotti, E.; Santi, G. Juggling reference frames in the microworld Mak-Trace: The case of a student with MLD. In *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*; Nicol, C., Liljedahl, P., Oesterle, S., Allan, D., Eds.; PME: Vancouver, BC, Canada, 2014; Volume 2, pp. 81–88.
44. Daniela, L.; Lytras, M.D. Educational robotics for inclusive education. *Technol. Knowl. Learn.* **2019**, *24*, 219–225. [CrossRef]

education sciences

MDPI

Article

Students' Perceived Mathematics Teacher Competence: Longitudinal Associations with Learning Outcomes and Choice of College Major

Dongjo Shin [1] and Jaekwoun Shim [2,*]

[1] Center for Curriculum and Instruction Studies, Korea University, Seoul 02841, Korea; dongjo@korea.ac.kr
[2] Center for Gifted Education, Korea University, Seoul 02841, Korea
* Correspondence: silent99@korea.ac.kr

Abstract: The quality of teaching and teacher plays a significant role in enhancing students' cognitive and motivational development. The purpose of this study was to identify longitudinal trajectories of student perceptions of mathematics teachers' professional competence, and to examine how these trajectories were related to students' engagement and achievement in mathematics and their choice of college major. The data used for analysis were obtained from the Seoul Education Longitudinal Study, a study of 2714 secondary students who were followed up for five successive years. Results showed four heterogeneous trajectories of students' perceptions of teacher competence, and these trajectories were associated with students' engagement and achievement in mathematics. In addition, those who consistently perceived their mathematics teachers to be highly competent and whose perceptions of their teachers' competence became more positive over time were more likely than other students to choose math-intensive majors. These findings imply longitudinal links between mathematics teachers' professional competence as perceived by their students and students' choice of college major as well as their engagement and achievement in mathematics.

Keywords: teacher competence; student perceptions; learning mathematics; college major choice; K-means; Seoul Education Longitudinal Study

Citation: Shin, D.; Shim, J. Students' Perceived Mathematics Teacher Competence: Longitudinal Associations with Learning Outcomes and Choice of College Major. *Educ. Sci.* **2021**, *11*, 18. https://doi.org/10.3390/educsci11010018

Received: 16 November 2020
Accepted: 30 December 2020
Published: 4 January 2021

1. Introduction

Teachers play a significant role in enhancing students' cognitive and motivational development [1,2]. Consequently, training and assisting high-quality teachers is considered one of the top priorities for educational research and policy making around the world [3]. The definition of a "good" teacher, however, varies across cultures, student ages, and subjects concerned [4,5]. Traditionally, teacher quality has been assessed by teachers' qualifications, including professional credentials, college degrees, and teacher test scores [6], because it was believed that a teacher's qualifications were positively correlated with students' academic success [7]. In contrast, recent studies have provided conflicting evidence regarding the relationship between teachers' test scores and their students' academic achievement [8–10]. Researchers have sought to conceptualize essential aspects of good teaching in terms of teachers' personal characteristics (e.g., instructional knowledge and motivation [11]) and teaching quality (e.g., classroom management and learning support [12]) and found positive effects of these aspects on student learning of mathematics [13–15].

The quality of teaching and teacher competence is frequently measured by external observers and teachers' self-reported teaching practices [16,17]. In addition to these measures, there is an extensive body of literature on students' ratings of their teachers. This body of research has demonstrated that student-rated teaching quality and teacher competence are potential predictors of successful student learning [18–20], although some researchers argue that teaching effectiveness should not be measured only by students'

perceptions [21]. However, little is known about how students' perceptions are associated with their academic outcomes in mathematics. Moreover, most of the previous studies were limited by cross-sectional designs; thus, we know little about how students' perceptions of their teachers are longitudinally related to their cognitive and motivational outcomes. More importantly, although students' motivation and achievement in mathematics significantly influence their selections of math-intensive courses and their college major [22,23], there have been few studies on how student-perceived teacher practice is associated with students' choice of college major. In the current study, we build on and extend the existing literature by identifying distinct longitudinal trajectories of student perceptions of teacher competence and by associating these trajectories with student outcomes and college major choice.

2. Theoretical Background

2.1. Teachers' Knowledge and Enthusiasm for Teaching

High-quality classroom instruction depends on teacher competence and teaching quality [11,15]. Distinguishing teacher competence from teaching quality, Fauth et al. [11] conceptualized teachers' professional competence as "teachers' personal characteristics" (p. 2), which consisted of two basic dimensions: cognitive and motivational. The cognitive dimension of teacher competence comprises teachers' profession-specific knowledge and beliefs. The motivational dimension of teacher competence includes teachers' self-efficacy and enthusiasm for teaching. Researchers have provided empirical evidence that these dimensions of teacher competence affect the teaching and learning of mathematics [11,15,17]. Among these subcategories of the construct of teacher professional competence, however, it is difficult to measure students' perceptions of teachers' self-efficacy because the construct of self-efficacy, in its nature, is almost invisible by students. Because the purpose of this study is to examine students' perceptions of mathematics teachers' competence (SP-MTC), the current study focuses on mathematics teachers' knowledge, beliefs, and enthusiasm for teaching as aspects of teachers' professional competence (Figure 1).

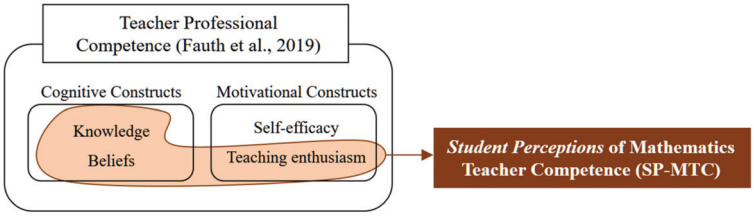


Figure 1. Conceptual framework for student perceptions of mathematics teacher competence.

2.1.1. Teachers' Mathematics Knowledge and Beliefs

Knowledge is undoubtedly one of the important indicators of successful mathematics teachers [24–26]. Researchers in mathematics education have found that not only teachers' mathematics content knowledge but also mathematical knowledge for teaching (MKT) and pedagogical content knowledge (PCK) are key factors affecting student outcomes in mathematics. Using two panel datasets, Hill et al. [24] showed that teachers' MKT positively predicted student achievement in both first and third grades. Baumert et al. [13] also revealed that secondary mathematics teachers' PCK positively affected students' learning gains. Teachers' strong knowledge also led to instruction based on students' mathematical thinking [27,28]. Dick [29], for example, demonstrated that prospective teachers' specialized content knowledge led to greater levels of noticing of students' mathematical thinking. As another cognitive construct of teacher professional competence, teacher beliefs directly influence teachers' instructional decisions [30]. Bobis, Way, Anderson, and Martin [31] state that mathematics teacher beliefs need to be considered when examining the teaching and learning of mathematics because teachers' teaching practices are shaped by what they

believe about the teaching and learning of mathematics. Research has shown that mathematics teachers who believe their students have the ability to acquire new knowledge and to solve challenging mathematics problems tended to build their instructions on students' mathematical thinking [32]. However, it was noted that teachers' knowledge and beliefs are not sufficient for successful mathematics teaching [33]. The ability to motivate students (e.g., teacher enthusiasm for teaching) is another area that attracted many researchers' attention due to its significant influences on student learning [11].

2.1.2. Mathematics Teachers' Enthusiasm

Past researchers have developed two main conceptualizations of teacher enthusiasm. Some have considered teacher enthusiasm to be the demonstrated behaviors of teachers in a classroom, which Keller, Hoy, Goetz, and Frenzel [34] called displayed enthusiasm. According to this definition of teacher enthusiasm, enthusiastic teachers are dynamic and energetic (e.g., humorous) and use various non-verbal expressions (e.g., gestures and facial expressions) in order to create a learning environment that encourages students to express their ideas [35]. Others have regarded teacher enthusiasm as a dispositional characteristic or a positive affective experience of teachers while teaching, which Keller et al. called it experienced enthusiasm. In this line of research, Kunter et al. [15] conceptualized teacher enthusiasm as "the degree of enjoyment, excitement, and pleasure that teachers typically experience in their professional activities" and differentiated teacher enthusiasm for teaching from teacher enthusiasm for the subject that they teach (e.g., mathematics, science, etc.). Research showed that teachers' enthusiasm for teaching mathematics, in contrast to their enthusiasm for mathematics as a subject, was likely to lead to higher instructional quality [11,15,36].

Frenzel et al. [14], for example, found that mathematics teachers' enjoyment in mathematics classrooms was significantly related to student enjoyment and that teachers' enthusiasm for teaching mediated the effect of teachers' enjoyment on student enjoyment. The findings indicated that teachers' enthusiasm for teaching is important for student motivation in mathematics lessons. Keller et al. [17] revealed that dispositional teacher enthusiasm, measured by teachers' self-reported positive affect and emotional expressivity, was significantly related to students' interest. Further analysis showed that the effect of teacher enthusiasm on student interest was mediated by student ratings of teacher enthusiasm, indicating that teacher enthusiasm was only effective in stimulating students' interest in learning when the students actually perceived it [36].

2.2. Student Ratings of Teacher Quality

In the research on teaching quality, there is much discussion on the evaluation of teaching practices. Fraser and Walberg [37] stated that students' joint ratings of teaching practices, when based on long experience with the teacher, are more predictive of students' achievement than external observers' evaluations. Student ratings of instruction are, therefore, usually considered a useful measure of teaching quality, especially in higher education [19]. Patrick et al. [20], for example, revealed that student-rated instructor enthusiasm was a significant predictor of undergraduate students' intrinsic motivation. In contrast, ratings of secondary school students [18], as well as elementary school students [38], have been neglected in evaluating teachers' instructional quality due to bias in student ratings based on teacher popularity [15].

However, researchers argue that secondary students' perceptions of teaching quality can provide valid information on the different dimensions of instructional practices [39]. Furthermore, in a longitudinal study, Fauth et al. [38] found that elementary students' ratings of teachers' classroom management skills predicted their academic achievement and that their ratings of teachers' cognitive activation and supportive climate predicted their interest in subjects even after the teacher popularity was controlled. These findings indicate that, even at the elementary level, student-rated teaching practices can be used as

a useful measure of teacher quality, and they should therefore be taken into account when evaluating teacher effectiveness [18].

2.3. Course and Major Choice

Students' choices of courses and majors are important steps toward their future career. Researchers in the field of education have identified significant predictors affecting students' course and college major choices [22,23,40,41]. Based on the expectancy-value theory, Simpkins et al. [22] found that elementary school students' beliefs, measured by their assessment of their ability, interest, and perceived importance of mathematics, predicted the number of mathematics courses that they took in high school. Musu-Gillette et al. [40] found that students who presented consistently high levels of confidence in their ability in mathematics were more likely to choose mathematics-related college majors than others who had lower confidence in their ability. Similar patterns were observed in terms of students' interest and the perceived importance of mathematics [40].

Using a longitudinal non-parametric analysis, Ahmed [23] showed that students who presented consistently low levels of mathematics anxiety were much more likely to be employed in Science, Technology, Engineering, and Mathematics (STEM) careers than those presenting consistently high or increasing levels of anxiety over time. This finding indicated that students' experiences with mathematics anxiety may lead them to avoid mathematics-related careers and activities in their future. Fewer studies, however, have addressed how students' perceptions of teacher practices affected their academic choice. Shih et al. [41], for example, found that secondary students' perceptions of being invited or encouraged to participate in learning by their mathematics teachers indirectly affected their desire to engage in mathematics in the future (i.e., math-related course taking and future career), which was mediated by students' self-concept related to their mathematics ability.

In summary, the prior research has shown that teachers' professional competence significantly affects students' learning of mathematics [11,15,17]. However, secondary school students' perceptions of their teachers' practices, compared to college students' perceptions, have been neglected [11,15,18]. In addition, the literature on students' academic choices has tended to focus mostly on students' concepts of their own ability based on the expectancy-value theory [22,23,40]. To our knowledge, no research has addressed how SP-MTC is longitudinally related to their learning engagement and achievement and their choice of college major. The current study, therefore, addresses three research questions: (a) What are the distinct longitudinal trajectories of SP-MTC? (b) How are the longitudinal trajectories of SP-MTC related to students' mathematics achievement and engagement in mathematics? (c) Are longitudinal trajectories of SP-MTC predictive of students' choice of college major?

3. Methods

3.1. Participants and Setting

The current study used data drawn from the Seoul Education Longitudinal Study (SELS) in Korea. The SELS study was initiated to systematically investigate various factors of education, to deal with educational issues, and to help establish effective education policies and long-term plans. Because the purpose of this study was to identify longitudinal trajectories of SP-MTC, we collected five-year longitudinal data from the beginning of eighth grade to the end of twelfth grade (from 2014 to 2018). For each year, students completed a set of SELS questionnaires and took achievement tests in mathematics. Those who missed the questionnaires for more than three years were excluded from our analysis. Data from 2714 students (about 47.6% females and 52.4% males) were finally included in this study.

3.2. Measures

3.2.1. Teacher Knowledge, Beliefs, and Enthusiasm for Teaching

In the SELS questionnaire, students were asked to rate their mathematics teachers' professional competence related to teachers' knowledge, beliefs, and enthusiasm for teaching. Mathematics teachers' professional competence perceived by students was measured with seven items (sample items: "My math teacher is enthusiastic about teaching mathematics", "My teacher has a high level of knowledge in mathematics", and "My teacher believes that students will achieve high achievement in mathematics"). All items were rated on a five-point Likert-type scale (See Appendix A). Inter-reliabilities among the items for each grade were excellent ($\alpha = 0.93 - 0.94$ from 8th to 12th grade). Therefore, the average of the seven items was used to indicate students' ratings of their mathematics teachers' competence.

3.2.2. Student Engagement in Mathematics Classes

Students' engagement in mathematics classes was measured using five items (sample items: "I concentrate in math class" and "I actively participate in math class"). Students' perceptions of their levels of engagement in mathematics classes were also rated on a five-point Likert-type scale (See Appendix B). Inter-reliabilities among the five items were good ($\alpha = 0.87 - 0.94$ from 8th to 12th grade). Missing values were replaced using the expectation maximization (EM) technique.

3.2.3. Student Mathematics Achievement

Students' mathematics achievement was measured by their scores on a standardized mathematics test (approximately 30 multiple-choice items) conducted by the SELS every year. In order to examine the extent to which a student has achieved the educational goals of school mathematics, the standardized test asks students if they know basic knowledge of school mathematics. The test is also used to identify and diagnose at-risk students in mathematics. In the current study, the EM algorithm was used to impute missing mathematics test scores. The SELS used vertical scaling, which placed scores from five different tests on the same scale in order to control for any differences in test difficulty. Vertical scaling, therefore, can facilitate the estimation of a student's or a panel's growth over time (see Table 1).

Table 1. Descriptive statistics for students' mathematics achievement.

	Min	Max	Mean	SD
Grade 8	162	505	345.22	59.16
Grade 9	176	515	363.86	64.13
Grade 10	190	533	373.68	63.64
Grade 11	216	558	370.58	42.91
Grade 12	220	559	377.04	53.03

3.2.4. College Major Choice

Grade 12 students in 2018 were asked, "Which major do you want to study in college?" The SELS coded college majors into nine categories (1 = Humanities, 2 = Social Sciences, 3 = Math and Science, 4 = Technology and Engineering, 5 = Medicine and Pharmacy, 6 = Education, 7 = Art and Athletics, 8 = Police and Military, and 9 = Others). We collapsed three categories, Math/Science, Technology/Engineering, and Medicine/Pharmacy, into a single category, *Math Intensive Major*, and the other six categories into *Non Math Intensive Major* (i.e., Humanities, Social Science, Education, Art/Athletics, Police/Military, and Others). Note that the category *others* includes students whose college majors did not fall into the eight categories or who did not plan to go to college.

3.3. Data Analysis

We used a longitudinal, non-parametric approach, *K-means for longitudinal data* (KmL), to identify clusters with a similar longitudinal pattern regarding student perceptions of teacher competence. KmL, suggested by Genolini and Falissard [42], is a variant of the K-means algorithm, designed to take into account longitudinal data. Researchers in education have used KmL analysis to identify distinct longitudinal trajectories of students' course-taking [43], math anxiety [23], and forum participation in a Massive Open Online Course (MOOC) [44].

The KmL approach is an iterative process: (a) once the number of clusters is predefined by researchers, each observation (In the KML algorithm, the *i*th observation's trajectory regarding the y variable is denoted by $y_i = (y_{i1}, y_{i2}, \cdots, y_{it})$, where t indicates time.) is initially assigned to a cluster; (b) the centers of each cluster (called centroids) are automatically computed; (c) each observation is reassigned to its nearest cluster based on distances (In the present study, we used the Euclidean distance, denoted by $Dist(y_i, y_j) = \sqrt{\frac{1}{t}\sum_{k=1}^{t}(y_{it} - y_{jt})^2}$.) between the observation and different centroids; and (d) the procedure is repeated until no further changes happen [42,45]. R package KmL was used to perform the KmL analysis. To select the optimal number of clusters, we compared three non-parametric criteria (Calinski Harabatz (CH), Ray and Turi (RT), and Davies and Boudin (DB)) that were provided by the package, because there is no single criterion that helps to choose the best number of clusters [42,43,45]. Because the KmL is able to deal with intermittent missing values in longitudinal data [42], cases with missing values at more than three time points were excluded from the analysis. After the longitudinal clusters regarding SP-MTC were identified, we performed linear mixed model analyses with student engagement and achievement in mathematics as outcome variables in order to examine how the longitudinal clusters were related to the outcome variables. The R package *lmerTest* was used to conduct the linear mixed model analysis. Finally, binary logistic regression was conducted to explore the contribution of the longitudinal clusters in predicting math-intensive major choice.

4. Results

4.1. Trajectories of SP-MTC

We first used the KmL approach to test different models with three to six clusters. The CH criterion favored three clusters followed by four clusters, while the RT criterion favored four clusters followed by three clusters. Finally, the DB criterion favored four clusters followed by five clusters. Based on these criteria, we selected four clusters as the optimal number of clusters. The four trajectories of SP-MTC are illustrated in Figure 2. These four clusters are labeled: *high* (34.6%), *decreasing* (28.7%), *increasing* (19.0%), and *low* (17.7%), based on their patterns over time. Students in the *high* and *low* clusters were likely to rate their mathematics teachers' professional competence as consistently higher and lower than did those in the *increasing* and *decreasing* clusters, respectively. Students' ratings in both the *decreasing* and the *increasing* clusters significantly changed their perceptions of teacher competence over time. Overall, students in the *increasing* cluster presented the opposite trajectory to those in the *decreasing* cluster.

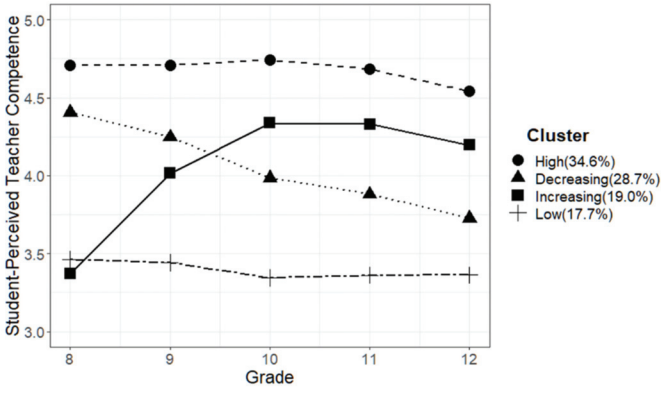


Figure 2. Four distinct trajectories of students' perceptions of mathematics teachers' competence (SP-MTC).

The means and standard deviations (in parenthesis) of SP-MTC are shown in Table 2.

Table 2. Descriptive statistics for distinct trajectories.

Grade	High	Decreasing	Increasing	Low
8	4.71 (0.39)	4.41 (0.46)	3.37 (0.61)	3.46 (0.70)
9	4.71 (0.43)	4.25 (0.54)	4.01 (0.65)	3.44 (0.66)
10	4.74 (0.38)	3.98 (0.54)	4.33 (0.56)	3.34 (0.66)
11	4.68 (0.42)	3.88 (0.54)	4.33 (0.51)	3.36 (0.59)
12	4.54 (0.49)	3.72 (0.57)	4.20 (0.51)	3.36 (0.55)

4.2. Relationship between SP-MTC and Student Learning

Using the four clusters identified, we examined if SP-MTC was associated with students' engagement and achievement in mathematics. First, we drew lines to connect annual mean scores in students' levels of engagement and achievement in mathematics and analyzed the differences among the four clusters over five years (see Figure 3).

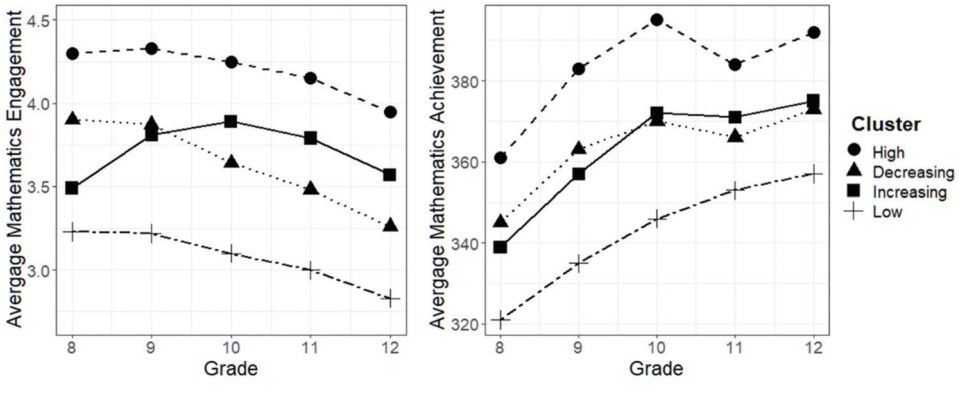


Figure 3. Students' levels of engagement (**left**) and mathematical achievement (**right**).

As reported in the previous research, students' engagement in mathematics tended to decrease [46], whereas their achievement in mathematics was likely to increase over time. However, trends in students' engagement and achievement in mathematics among the four clusters showed a similar tendency to the changes in their perceptions of teacher

competence during the five years. In general, the differences in engagement and achievement among the four clusters were bigger, but there was a minor difference in mathematics achievement between students in the *decreasing* and *increasing* clusters. To examine whether the differences are statistically significant, we conducted linear mixed model analyses. The equation of the model is as follows:

$$Y_{it} = \beta_0 + \beta_1(grade) + \beta_2(high) + \beta_3(increasing) + \beta_4(low) + \beta_5(high \times grade) + \beta_6(increasing \times grade) + \beta_7(low \times grade) + student_i + e_{it}$$

where Y_{it} is either mathematics engagement or mathematics achievement for a student i at grade t that is centered around 0, and the intercept refers to the *decreasing* cluster at the start of this study. We entered individual students as a random factor to account for within-student dependence and incorporated cluster, grade, and interaction into the models to understand patterns in different clusters' mathematics engagement and achievement during the five years. The results of the linear mixed model analyses are detailed in Table 3 and Figure 3.

Table 3. Summary of linear mixed-effect models.

Fixed	Estimate	SE	*t*-Value	Random	Variance	*SD*
Model 1: Mathematics Engagement						
(intercept)	3.96 ***	0.03	144.77	(Intercept)	0.33	0.57
high	0.41 ***	0.04	11.00	Residual	0.42	0.65
increasing	−0.28 ***	0.04	−6.47			
low	−0.69 ***	0.04	−15.55			
grade	−0.17 ***	0.01	−22.71			
high × grade	0.08 ***	0.01	7.93			
increasing × grade	0.18 ***	0.01	15.42			
low × grade	0.08 ***	0.01	5.56			
Model 2: Mathematics Achievement						
(intercept)	352.02 ***	2.40	192.90	(Intercept)	1840	42.90
high	18.45 ***	2.47	7.48	Residual	1251	35.38
increasing	−6.76 *	2.89	−2.34			
low	−27.77 ***	2.95	−9.41			
grade	5.64 ***	0.40	14.07			
high grade	0.56	0.54	1.03			
increasing × grade	3.16 ***	0.64	4.97			
low × grade	3.41 ***	0.65	5.25			

Note: * $p < 0.05$, *** $p < 0.001$; SE indicates standard error; Reference group is the *decreasing* cluster.

Model 1 in Table 3 suggests that, when students' initial level of engagement in mathematics (i.e., Grade 8) was compared using the *decreasing* cluster as a reference group, a significantly higher engagement level was observed for the *high* cluster, while significantly lower engagement levels were observed for the *increasing* and the *low* clusters. The engagement of students in the *increasing* cluster slightly increased over time, while that of students in the *decreasing* cluster was significantly decreased over time ($\beta = -0.17$, $p < 0.001$). Interestingly, it was found that the engagement in mathematics of students in the *high* cluster decreased as they advanced from one year to the next. There were significant interactions between cluster and grade such that students' engagement significantly decreased for the *decreasing* cluster in comparison with the *high*, *increasing*, and *low* clusters over time. As seen in Figure 4, the engagement of students in the *increasing* cluster has been higher than that of students in the *decreasing* cluster since Grade 10. The statistical significances of the linear mixed model analyses for achievement in mathematics (i.e., Model 2 in Table 3) were generally similar to those for mathematics engagement (i.e., Model 1 in Table 3). The initial level of mathematics achievement in the *decreasing* cluster was higher than that in the *increasing* cluster ($\beta = -6.76$, $p < 0.05$), but a significant increase was observed in the *increasing* cluster in comparison with the *decreasing* cluster even though

students' mathematics achievement in the two clusters increased with advancement to the next grade.

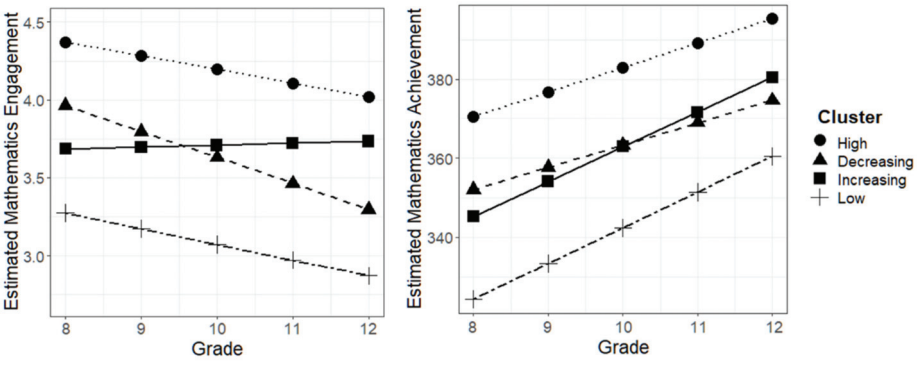


Figure 4. Results of linear mixed effect models for mathematics engagement and achievement.

4.3. Relationship between SP-MTC and Math-intensive Major Choice

To examine the contribution of the longitudinal clusters in predicting students' math-intensive college major choice, we performed binary logistic regression. As indicated in the unadjusted model (see Model 1 in Table 4), the *high*, *increasing*, and *decreasing* clusters were significantly different from the *low* cluster (reference group) in predicting students' math-intensive major choice. Students in the *high*, *increasing*, and *decreasing* clusters were about 2.58, 1.94, and 1.50 times as likely to choose math-intensive majors as those in the *low* cluster, respectively. The results of the adjusted model in which *gender* was taken into account (see Model 2 in Table 4) were almost same as those of the unadjusted model. With regard to gender, male students were more likely than female students to choose math-intensive majors in college. We further fitted a logistic regression model on college major choice using the *decreasing* cluster as a reference group and found significant difference in math-intensive major choice between the *decreasing* and the *increasing* clusters. The odds of choosing a math-intensive major were about 1.3 times higher in the *increasing* cluster than in the *decreasing* cluster.

Table 4. Summary of logistic regression predicting math-intensive career choice.

Variable	Model 1: Unadjusted				Variable	Model 2: Adjusted			
	B	SE	OR	95% CI		B	SE	OR	95% CI
low			1		*low*			1	
high ***	0.95	0.14	2.58	1.98–3.36	high ***	0.93	0.14	2.54	1.95–3.31
increasing ***	0.67	0.15	1.94	1.45–2.61	increasing ***	0.68	0.15	1.97	1.46–2.64
decreasing ***	0.41	0.14	1.50	1.14–1.99	decreasing ***	0.43	0.14	1.53	1.16–2.03
					female			1	
					male ***	0.44	0.09	1.55	1.31–1.83

Note. *** $p < 0.001$; B, SE, OR, and CI indicate regression coefficient, standard error, odds ratio, and confidence interval, respectively; the *low* cluster and *female* are reference groups.

5. Discussion

Considering the importance of teachers to student development, the present study explored changes in SP-MTC over five years. We identified four distinct trajectories of SP-MTC and found that the trajectories of change were associated with students' academic outcomes (i.e., engagement and achievement in mathematics) as well as their choice of a math-intensive college major. These findings suggest the construct of SP-MTC, consisting

of student-perceived mathematics teacher knowledge, beliefs, and enthusiasm for teaching is a critical factor for students' learning and future careers.

5.1. Mathematics Teachers Matter for Students' Learning and College Major Choice

The current study found a significant relationship between SP-MTC and students' engagement in mathematics classes. Moreover, significant differences in students' engagement among the four clusters were found, which resonates with previous research that identified distinct profiles of students' engagement based on teaching practices [47]. This finding indicates that students might be engaged in mathematics classes because they actually perceive their mathematics teachers to be highly enthusiastic, which extends a one-year cross-sectional study [14,17,36] by providing longitudinal links between perceived teacher competence and students' engagement in learning mathematics.

Our study also found a positive relationship between SP-MTC and students' mathematics achievement, which may indicate that the higher students' ratings of teacher knowledge, beliefs, and enthusiasm are, the more likely it is that they will perform well on mathematics tests. This suggests that SP-MTC can be considered one of the core factors predicting students' achievement in mathematics [18,38,39]. However, this finding contrasts with prior research showing that teachers' professional competence predicts students' interest but not their achievement [11]. These inconsistent findings could be explained by differences in measurement methods between the studies. Fauth et al. [11] measured teacher knowledge with a standardized test and teacher beliefs and enthusiasm with teachers' self-reports, whereas the current study measured them with students' reports of teacher knowledge, beliefs, and enthusiasm. Consistent with the findings of Keller et al. [17], this indicates that teachers' knowledge, beliefs, and enthusiasm are important components of teachers' professional competence, but, more importantly, these components must be perceived by students to have an impact on student achievement in a subject.

We also investigated the contributions of developmental trajectories of SP-MTC to students' choice of a math-intensive major in college. Perhaps unsurprisingly, students in the *high* cluster wanted to study a math-intensive major, whereas students in the *low* cluster were more likely to avoid such a major. This finding is consistent with previous studies [22,41,48,49] and extends prior research [40] by showing that distinct trajectories of change in SP-MTC were differentially associated with students' choice of a math-related major. There was also a significant difference in college major choice between the *increasing* and *decreasing* cluster; that is, students in the *increasing* cluster were more likely than those in the *decreasing* cluster to choose math-intensive majors. As discussed earlier, engagement and achievement of students in the *decreasing* cluster were higher than those in the *increasing* cluster at Grade 8 and then became similar at Grade 9. Significant changes in students' engagement and achievement were observed between Grade 9 and Grade 10, which is the transition year from middle school to high school in South Korea; engagement and achievement of students in the *increasing* cluster were higher than those in the *decreasing* cluster from Grade 10 (see Figure 3). This finding suggests that, for students in either the *increasing* or *decreasing* cluster, the transition to high school was a critical period for selecting their college majors, which is in line with prior studies that found that 9th grade students' perceptions of mathematics predicted their desire for math-related future careers [50].

On the contrary, several studies demonstrated different results. Li, Alfeld, Kennedy, and Putallaz [51], for example, revealed that students' interest in mathematics in middle school has a long-term impact on their college major choice. Researchers have also argued that the early student years (before 7th grade) are important for students' academic plans and choices in their future [22,48,52]. In this regard, the findings in the current study provide new insight into students' career choices by helping to predict which students are likely to choose a math-intensive major. For half of the students who consistently rated their mathematics teachers' professional competence as high (about 34.6%) or low (about 17.7%), it is likely that their early student years were a critical time to determine their math-intensive college majors and future careers because patterns of change in their perceptions

(and also their engagement and achievement in mathematics) were consistent from the beginning of middle school (i.e., Grade 8). However, for the other half of the students who showed substantial changes in their perceptions as they advanced from one year to the next (i.e., the *increasing* and *decreasing* cluster), it is likely that the transition to high school was an important year for their future academic choice because significant changes occurred in their engagement and achievement in mathematics at that time. The findings of this study extend previous studies that reported an important time point for determining students' future career and college major [22,48,50–52] by showing that there is no single critical time point for students' math-intensive college major choice; rather, it depends on their perceptions of teacher competence and their learning.

Taken together, our results indicate that mathematics teachers matter, not only for their students' learning, but also for their future career. Previous studies have consistently provided evidence that mathematics teachers' knowledge is one of the most important factors for student learning [53]. In the current study, the SELS did not measure teachers' knowledge, beliefs, and enthusiasm by using a standardized test, nor by observing their teaching practices, and thus we do not know whether or not the teachers were actually competent. Regardless of the teachers' actual knowledge, beliefs, and enthusiasm for teaching, however, students' learning could be maximized if they feel their teachers are professionally competent. As seen previously, SP-MTC among different clusters has hardly changed since the students became high school students (i.e., Grade 10). In conclusion, mathematics teachers need to identify how students perceive their teaching practices early in middle school, and additional attention (e.g., changing their teaching methods, instructional materials, or interaction with students) should be given, especially to students who were likely to rate their mathematics teachers' competence as consistently or gradually low (i.e., students in the *low* and *decreasing* cluster), in order to prevent them from giving up learning mathematics in high school.

5.2. Limitations and Recommendations for Future Research

There were several limitations to this study, mostly due to the nature of longitudinal studies. First, the participants in the current study were not a nationally representative sample of students but were a sample from a metropolitan city, Seoul, where there is a highly competitive educational culture. Second, some data were initially excluded from our analysis, especially because they were not followed up during high school (i.e., Grade 10 to 12). This might lead to a potential underestimation of the influence of SP-MTC on students' academic outcomes and choice of major. Third, although previous studies demonstrated that students' individual characteristics and their previous attainment significantly affect their math-intensive major choices as well as their engagement and achievement in mathematics [36,54,55], the present study did not take those characteristics into account, except for gender. Because of these missing variables, it will be difficult to attribute students' mathematics achievement and engagement to SP-MTC alone. Future research is needed to examine how these variables are associated with SP-MTC and students' learning outcomes. Lastly, the findings in the study do not denote a causal relationships between SP-MTC and students' learning and future careers. For example, high-achieving students in mathematics may tend to have positive perceptions of their teachers. Future studies, therefore, may examine how SP-MTC affects students' learning outcomes and their choice of college major. Despite these limitations, the current study contributes to the field of student development by identifying distinct trajectories of SP-MTC and by showing that the heterogeneous trajectories are associated, not only with students' academic success, but also with the direction of their future studies.

Author Contributions: Conceptualization, D.S.; methodology, D.S.; software, J.S.; validation, J.S. and D.S.; formal analysis, J.S. and D.S.; investigation, D.S.; data curation, D.S.; writing—original draft preparation, D.S.; writing—review and editing, J.S. and D.S.; visualization, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to the data came from Seoul Education Longitudinal Study (SELS) and are publicly available for educational research. This means that the SELS collected the data and obtained informed content and then provided researchers to use them publicly.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.serii.re.kr/.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. What Do You Think about Your Math Teacher?

	Item	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	My teacher is enthusiastic about teaching math.					
2	My teacher has a high level of knowledge in math.					
3	My teacher teaches math well so that we can understand it easily.					
4	My teacher expects us to study hard.					
5	My teacher believes that we will achieve high achievement in math.					
6	My teacher conscientiously checks our math assignments.					
7	My teacher monitors how well students understand math in class.					

Appendix B

Table A2. How Much Do You Engage in Math Class?

	Item	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I concentrate in math class.					
2	I actively participate in math class.					
3	I do my math homework conscientiously.					
4	I review what I have learned in math class.					
5	I prepare what I will learn in math class.					

References

1. Hanushek, E. Valuing teachers: How much is a good teacher worth? *Educ. Next* **2011**, *11*, 40–45.
2. Rivkin, S.G.; Hanushek, E.A.; Kain, J.F. Teachers, schools, and academic achievement. *Econometrica* **2005**, *73*, 417–458. [CrossRef]
3. Darling-Hammond, L. Teacher education around the world: What can we learn from international practice? *Eur. J. Teach. Educ.* **2017**, *40*, 291–309. [CrossRef]
4. Lee, H.H.; Kim, G.M.L.; Chan, L.L. Good teaching: What matters to university students. *Asia-Pacific J. Educ.* **2015**, *35*, 98–110. [CrossRef]
5. Burnett, P.C.; Meacham, D. Measuring the quality of teaching in elementary school classrooms. *Asia Pac. J. Teach. Educ.* **2002**, *30*, 141–153. [CrossRef]
6. Goe, L. The Link between Teacher Quality and Student Outcomes: A Research Synthesis. Available online: https://files.eric.ed.gov/fulltext/ED521219.pdf (accessed on 9 November 2020).
7. Clotfelter, C.T.; Ladd, H.F.; Vigdor, J.L. Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. Available online: https://www.nber.org/system/files/working_papers/w13617/w13617.pdf (accessed on 9 November 2020).
8. Aloe, A.M.; Becker, B.J. Teacher verbal ability and school outcomes. *Educ. Res.* **2009**, *38*, 612–624. [CrossRef]
9. Kunter, M.; Klusmann, U.; Baumert, J.; Richter, D.; Voss, T.; Hachfeld, A. Professional competence of teachers: Effects on instructional quality and student development. *J. Educ. Psychol.* **2013**, *105*, 805–820. [CrossRef]

10. Rockoff, J.; Jacob, B.; Kane, T.; Staiger, D. Can you recognize an effective teacher when you recruit one? *Educ. Finance Policy* **2011**, *6*, 43–74. [CrossRef]
11. Fauth, B.; Decristan, J.; Decker, A.T.; Büttner, G.; Hardy, I.; Klieme, E.; Kunter, M. The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teach. Teach. Educ.* **2019**, *86*, 102882. [CrossRef]
12. Rimm-Kaufman, S.; Hamre, B. The role of psychological and developmental science in efforts to improve teacher quality. *Teach. Coll. Rec.* **2010**, *112*, 2988–3023.
13. Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M.; Tsai, Y.M. Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *Am. Educ. Res. J.* **2010**, *47*, 133–180. [CrossRef]
14. Frenzel, A.C.; Goetz, T.; Lüdtke, O.; Pekrun, R.; Sutton, R. Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *J. Educ. Psychol.* **2009**, *101*, 705–716. [CrossRef]
15. Kunter, M.; Tsai, Y.M.; Klusmann, U.; Brunner, M.; Krauss, S.; Baumert, J. Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learn. Instr.* **2008**, *18*, 468–482. [CrossRef]
16. De Jong, R.; Westerhof, K.J. The quality of student ratings of teacher behaviour. *Learn. Environ. Res.* **2001**, *4*, 51–85. [CrossRef]
17. Keller, M.M.; Goetz, T.; Becker, E.S.; Morger, V.; Hensley, L. Feeling and showing: A new conceptualization of dispositional teacher enthusiasm and its relation to students' interest. *Learn. Instr.* **2014**, *33*, 29–38. [CrossRef]
18. Ampadu, E. Students' perceptions of their teachers' teaching of mathematics: The case of Ghana. *Int. Online. J. Educ. Sci.* **2012**, *4*, 351–358.
19. Feldman, K.A. Identifying exemplary teachers and teaching: Evidence from student ratings. In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*; Perry, R., Smart, J., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp. 93–129. ISBN 978-1-4020-4944-6.
20. Patrick, B.C.; Hisley, J.; Kempler, T. "What's everybody so excited about?": The effects of teacher enthusiasm on student intrinsic motivation and vitality. *J. Exp. Educ.* **2000**, *68*, 217–236. [CrossRef]
21. Crumbley, L.; Henry, B.K.; Kratchman, S.H. Students' perceptions of the evaluation of college teaching. Quality assurance in Education. *Qual. Assur. Educ.* **2001**, *9*, 197–207. [CrossRef]
22. Simpkins, S.D.; Davis-Kean, P.E.; Eccles, J.S. Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Dev. Psychol.* **2006**, *42*, 70–83. [CrossRef]
23. Ahmed, W. Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. *J. Adolesc.* **2018**, *67*, 158–166. [CrossRef]
24. Hill, H.C.; Blunk, M.L.; Charalambous, C.Y.; Lewis, J.M.; Phelps, G.C.; Sleep, L.; Ball, D.L. Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cogn. Instr.* **2008**, *26*, 430–511. [CrossRef]
25. Tchoshanov, M.A. Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educ. Stud. Math.* **2011**, *76*, 141–164. [CrossRef]
26. Walshaw, M. Teacher knowledge as fundamental to effective teaching practice. *J. Math. Teach. Educ.* **2012**, *15*, 181–185. [CrossRef]
27. Dreher, A.; Kuntze, S. Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educ. Stud. Math.* **2015**, *88*, 89–114. [CrossRef]
28. König, J.; Blömeke, S.; Klein, P.; Suhl, U.; Busse, A.; Kaiser, G. Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teach. Teach. Educ.* **2014**, *38*, 76–88. [CrossRef]
29. Dick, L.K. Investigating the relationship between professional noticing and specialized content knowledge. In *Teacher Noticing: Bridging and Broadening Perspectives, Contexts, and Frameworks*; Schack, E.O., Fisher, M.H., Wilhelm, J.A., Eds.; Springer International Publishing: New York, NY, USA, 2017; pp. 339–358. ISBN 978-3-319-46752-8.
30. Goldin, G.; Rösken, B.; Törner, G. Beliefs—No longer a hidden variable in mathematical teaching and learning processes. In *Beliefs and Attitudes in Mathematics Education: New Research Results*; Maass, J., Schloeglmann, W., Eds.; Sense: Rotterdam, The Netherlands, 2009; pp. 1–18. ISBN 978-90-8790-722-8.
31. Bobis, J.; Way, J.; Anderson, J.; Martin, A.J. Challenging teacher beliefs about student engagement in mathematics. *J. Math. Teach. Educ.* **2016**, *19*, 33–55. [CrossRef]
32. Fennema, E.; Carpenter, T.P.; Franke, M.L.; Levi, L.; Jacobs, V.R.; Empson, S.B. A longitudinal study of learning to use children's thinking in mathematics instruction. *J. Res. Math. Educ.* **1996**, *27*, 403–434. [CrossRef]
33. Ainley, J.; Luntley, M. The role of attention in expert classroom practice. *J. Math. Teach. Educ.* **2008**, *10*, 3–22. [CrossRef]
34. Keller, M.M.; Hoy, A.W.; Goetz, T.; Frenzel, A.C. Teacher enthusiasm: Reviewing and redefining a complex construct. *Educ. Psychol. Rev.* **2016**, *28*, 743–769. [CrossRef]
35. Collins, M.L. Effects of enthusiasm training on preservice elementary teachers. *J. Teach. Educ.* **1978**, *29*, 53–57. [CrossRef]
36. Carmichael, C.; Callingham, R.; Watt, H.M. Classroom motivational environment influences on emotional and cognitive dimensions of student interest in mathematics. *ZDM Math. Educ.* **2017**, *49*, 449–460. [CrossRef]
37. Fraser, B.; Walberg, H. Psychosocial learning environment in science classrooms: A review of research. *Stud. Sci. Educ.* **1981**, *8*, 67–92. [CrossRef]
38. Fauth, B.; Decristan, J.; Rieser, S.; Klieme, E.; Büttner, G. Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learn. Instr.* **2014**, *29*, 1–9. [CrossRef]

39. Wagner, W.; Göllner, R.; Helmke, A.; Trautwein, U.; Lüdtke, O. Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learn. Instr.* **2013**, *28*, 1–11. [CrossRef]
40. Musu-Gillette, L.E.; Wigfield, A.; Harring, J.R.; Eccles, J.S. Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educ. Res. Eval.* **2015**, *21*, 343–370. [CrossRef]
41. Shih, J.; Ing, M.; Phelan, J.; Brown, R.; Maiorca, C. The influence of students' self-perceptions and mathematics experiences on learning more mathematics in the future. *Investig. Math. Learn.* **2019**, *11*, 220–229. [CrossRef]
42. Genolini, C.; Falissard, B. KmL: K-means for longitudinal data. *Comput. Stat.* **2010**, *25*, 317–328. [CrossRef]
43. Wang, X.; Lee, Y.; Wickersham, K. Exploring the relationship between longitudinal course-taking patterns and in-state transfer into stem fields of study. *J. High. Educ.* **2019**, *90*, 272–297. [CrossRef]
44. Tang, H.; Xing, W.; Pei, B. Exploring the temporal dimension of forum participation in MOOCs. *Distance Educ.* **2018**, *39*, 353–372. [CrossRef]
45. Genolini, C.; Alacoque, X.; Sentenac, M.; Arnaud, C. kml and kml3d: R packages to cluster longitudinal data. *J. Stat. Softw.* **2015**, *65*, 1–34. [CrossRef]
46. Plenty, S.; Heubeck, B.G. A multidimensional analysis of changes in mathematics motivation and engagement during high school. *Educ. Psycho.* **2013**, *33*, 14–30. [CrossRef]
47. Watt, H.M.; Carmichael, C.; Callingham, R. Students' engagement profiles in mathematics according to learning environment dimensions: Developing an evidence base for best practice in mathematics education. *Sch. Psychol. Int.* **2017**, *38*, 166–183. [CrossRef]
48. Durik, A.M.; Vida, M.; Eccles, J.S. Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *J. Educ. Psychol.* **2006**, *98*, 382–393. [CrossRef]
49. Wang, M.T.; Willett, J.B.; Eccles, J.S. The assessment of school engagement: Examining dimensionality and measurement invariance across gender and race/ethnicity. *J. Sch. Psychol.* **2011**, *49*, 465–480. [CrossRef]
50. Watt, H.M. The role of motivation in gendered educational and occupational trajectories related to maths. *Educ. Res. Eval.* **2006**, *12*, 305–322. [CrossRef]
51. Li, Y.; Alfeld, C.; Kennedy, R.P.; Putallaz, M. Effects of summer academic programs in middle school on high school test scores, course-taking, and college major. *J. Adv. Acad.* **2009**, *20*, 404–436. [CrossRef]
52. Eccles, J.S.; Vida, M.N.; Barber, B. The relation of early adolescents' college plans and both academic ability and task-value beliefs to subsequent college enrollment. *J. Early Adolesc.* **2004**, *24*, 63–77. [CrossRef]
53. Hill, H.C.; Rowan, B.; Ball, D. Effects of teachers' mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* **2005**, *42*, 371–406. [CrossRef]
54. Nagy, G.; Trautwein, U.; Baumert, J.; Köller, O.; Garrett, J. Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educ. Res. Eval.* **2006**, *12*, 323–345. [CrossRef]
55. Wang, M.T.; Eccles, J.S.; Kenny, S. Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychol. Sci.* **2013**, *24*, 770–775. [CrossRef]

education sciences

MDPI

Article

Teachers and Teacher Education: Limitations and Possibilities of Attaining SDG 4 in South Africa

Toyer Nakidien [1,*], Marcina Singh [1] and Yusuf Sayed [1,2]

[1] Centre for International Teacher Education, Cape Peninsula University of Technology, Mowbray 7700, South Africa; SINGHM@cput.ac.za (M.S.); y.sayed@sussex.ac.uk (Y.S.)
[2] Centre for International Education, University of Sussex, Brighton BN1 9QR, UK
* Correspondence: tnakidien@gmail.com

Abstract: As we enter the last ten years leading to the realisation of the Sustainable Development Goals in 2030, African countries are still plagued with poverty and underdevelopment. For most children in Africa, the attainment of Sustainable Development Goal (SDG) 4 will remain elusive. Drawing from two interrelated empirical studies, one focusing on citizenship and social cohesion at high school level and the other on the implementation of assessment for learning at primary school level, it was found that not only are schools not equipped to provide the quality education as set out in SDG 4, but teachers need additional training to give expression to the ideals of SDG 4. In order for this to be adequately addressed, all interested stakeholders—government, business, and NGOs—need to be involved.

Keywords: assessment for learning; social cohesion; citizenship; sustainable development goals; teachers

Citation: Nakidien, T.; Singh, M.; Sayed, Y. Teachers and Teacher Education: Limitations and Possibilities of Attaining SDG 4 in South Africa. *Educ. Sci.* **2021**, *11*, 66. https://doi.org/10.3390/educsci11020066

Academic Editor: Jim Albright

Received: 21 December 2020
Accepted: 3 February 2021
Published: 9 February 2021

1. Introduction

More than twenty-five years into democracy and entering the final decade of the realisation of the SDGs, this paper reflects on the possibilities and limitations of attaining SDG 4, focusing on teachers in the context of South Africa.

In order to realise this laudable goal of equitable and quality education, a range of criteria need to be in place, most important of all being well-trained and suitably qualified teachers. However, to attain SDG 4 requires conducive teaching and learning environments characterised by, among other things, safe school spaces, running water and sanitation, and teaching and learning support material.

Drawing on data from two qualitative empirical studies, the findings illustrate the limitations and potentialities of attaining quality education—looking at both the affective and pedagogic dimensions of teaching—in countries wrestling to rid themselves of the legacies of past colonial and oppressive regimes. The article presents empirical data sourced from high school teachers in Quintile 1 and Quintile 5 urban and rural schools, and from foundation phase teachers in Quintile 1 and Quintile 2 rural schools. Quintiles represent the school poverty index, with Quintile 1 being the poorest and Quintile 5 being the least poor. The data highlight the kinds of challenges that need to be addressed and overcome in order to realise SDG 4.

2. Teachers and the Global Education Agenda: Framing Teachers and Their Work

This section presents a critical discussion of SDG 4 in the context of inequality, equity, and systemic challenges in education in the Global South, with a particular emphasis on the South African case. This section also discusses the notion of quality education as it relates to social cohesion, citizenship, and pedagogy.

2.1. Unpacking SDG 4

The focus on equitable and quality education and the role of teachers has a long policy trajectory in the development of the SDG for education. In particular, debates about the role of teachers in providing equitable and quality education can be traced to the UNESCO Position Paper on Education post-2015 and the Muscat Global Education Meeting (GEM) Agreement, which rightly emphasises a concern with teachers, teaching, and teacher education (see Sayed and Ahmed, 2015). Specifically, the policy recommendations in the UNESCO Position Paper [1] suggest a range of key aspects regarding teachers including "(a) recruiting and retaining well-trained and motivated teachers who use inclusive, gender-responsive, and participatory pedagogical approaches to ensure effective learning outcomes, (b) providing content that is relevant to all learners and to the context in which they live, (c) establishing learning environments that are safe, gender-responsive, inclusive and conducive to learning, and encompass mother tongue-based multilingual education, (d) ensuring that learners reach sufficient levels of knowledge and competencies according to national standards at each level, (e) strengthening capacities for learners to be innovative and creative, and to assimilate change in their society and the workplace and over their lifespans, and (f) strengthening the ways education contributes to peace, responsible citizenship, sustainable development and intercultural dialogue" [1] (p. 8). These ideals are carried over in the final SDG 4 goal and its ten associated targets, including target 4c, which commits to "by 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing states" [2]. The framework for action for the implementation of SDG 4 notes that " ... teachers and educators should be empowered, adequately recruited and remunerated, motivated, professionally qualified, and supported within well-resourced, efficient and effectively governed systems" [3].

While the inclusion of teachers and teacher education in the final SDG 4 reflects a victory for the policy advocacy, it is instructive to consider how the target regarding teachers is constructed. In particular, as the empirical findings below show, the notion of equity is underplayed in the conceptualisation of the target. In particular, our findings highlight the need for the provision of quality education by deploying qualified, competent teachers in hard-to-reach areas, such as rural contexts and contexts riddled with crime, violence, and abject poverty. Further, as we discuss below, the target fails to address the teacher competencies teachers need to teach in difficult and hard to reach schools and in societies emerging from protracted histories of conflict, segregation, division, and inequities, specifically the competencies teachers require to meaningfully engage marginalised and disadvantaged learners, providing them with epistemic access and with the skills and understandings to live peacefully together, exercising their citizenship rights. This is most pressing in an unequal education context such as South Africa, which is marked by the existence of two systems of education—privileged and marginalised [4,5]. This can partly be attributed to the fact that good teachers working in an enabling learning environment are clustered in the wealthier school sector, which, when added to the cultural capital of learners, create a double privilege [5].

Furthermore, the absence of the reference to continuing professional development (CPD) of teachers in all the global documents and SDG 4 goals and targets is striking. Recruiting "well-trained" teachers places the focus on initial teacher education, neglecting the significance that CPD can have in tackling the issue of, for example, unqualified and underqualified teachers that in addition present a potential teaching force, as this paper argues.

In the context of the global education teacher goals and targets, this paper turns its attention to how the inclusion of teachers post-2015 will be realised in highly unequal education contexts marked by long legacies of colonisation and racism. In particular, this paper speaks to the issue of social cohesion and citizenship that teachers enact in diverse spaces, and teacher pedagogy for meaningful epistemic access for learners in rural contexts in South Africa.

2.2. Teachers' Beliefs, Social Cohesion, and Citizenship

A number of policy directives have emphasised the importance of realising quality education systems, including Education for All, the Millennium Development Goals, the Sustainable Development Goals, as well as localised South African education policies, such as the South African Schools Act, No. 84 of 1996 [6]. Quality education as enunciated in these policies is directly related to improved economic conditions, reduction in inequalities, as well as social cohesion. However, while many interventions to achieve quality education have been directed at improving the teaching and learning of cognitive elements such as numeracy and literacy, not enough emphasis is placed on the affective dimensions that may contribute equally to quality education.

Quality education is multidimensional and can be regarded as "an outcome or a property, or even a process" [7] (p. 2). It is also regarded as the worth of education [8] and is used interchangeably with concepts such as equity, effectiveness, and efficiency [9]. The Sustainable Development Goal Target 4c notes that "teachers are a fundamental condition for guaranteeing quality education" [2]. Positioning all teachers as providers of quality education and equitable learning experiences undermines how teachers' beliefs, teacher positionality, and teacher agency, individual or collective, impact their teaching. More importantly, what the SDG 4 fails to acknowledge is the impact of teacher beliefs on classroom practices.

Beliefs, defined as "an explanatory principle for practice", are both real and profound [10] (p. 16). The beliefs that teachers hold influence their perceptions and judgements and affect their behaviour in the school environment [11–13]. Thus, mechanisms to change or improve teachers' classroom practices will require a change in teacher beliefs [14,15]. This has implications for how teachers are professionally developed throughout their careers. The importance of acknowledging teachers' beliefs and putting in place professional development programmes that consistently and effectively empower teachers to be more reflexive is crucial in highly unequal and post-conflict settings such as South Africa. A long history of separate development, social division, and legally enforced discrimination, particularly in education, requires robust and rigorous teacher professional development programmes to align teachers' beliefs with the values of citizenship and social cohesion, and in doing so, align their beliefs with global policy mandates.

Social cohesion and the values of citizenship should be at the core of teaching and learning in a democratic context [13,16]. This is because schools "are assigned the task of producing 'good citizens', ensuring that when children grow up and leave school . . . they are prepared to practise the civic virtues most valued in their respective societies" [17]. In a liberal mode of education, which is the prevailing mode of Western education systems, these values are underpinned by democracy, which argues for inclusivity, mutual respect, and equality. However, recent studies that investigated teachers' pedagogical practices in South African schools demonstrate that the legacies of apartheid in terms of how teachers were trained as well as the effects of Christian National Education and Bantu Education still linger in the current system, to the detriment of learners [13,16]. Furthermore, the sub-optimal teaching and learning contexts that characterise most of the public schools in South Africa make it difficult to fully realise the values of citizenship and do little to promote social cohesion.

The assumption of the SDGs that teachers will always do good, and that those who come into the profession have the best intentions that are aligned with the principles of democracy, social cohesion, and citizenship, undermines the effect of the political, social, and cultural contexts in which schools are situated. The work of teachers is both intellectual as well as affective. Thus, it is imperative that professional development efforts are cognisant of this duality and respond to the professional development needs of teachers accordingly. Not acknowledging the affective dimension of schooling promotes very narrow understandings of education and undermines the power teachers wield, through schools, to perpetuate social inequalities. It also hampers the realisation of quality education.

2.3. Teachers and Pedagogy: Assessment for Learning (AfL)

One of the keys to realising quality teaching and learning for all could be the implementation of AfL. Alexander [18] (p. 59) defines pedagogy as "the observable act of teaching together with its attendant discourse of educational theories, values, evidence, and justifications. It is what one needs to know, and the skills one needs to command, in order to make and justify the many different kinds of decisions of which teaching is constituted". Similarly, Westbrook et al. [19] (p. 8) argue that effective pedagogy refers to "those teaching and learning activities which make some observable change in students, leading to greater engagement and understanding and/or a measurable impact on student learning". AfL, which adopts a learner-centred approach, satisfies these criteria as Dylan Wiliam asserts, " . . . there is now a strong body of theoretical and empirical work that suggests that integrating assessment with instruction may well have unprecedented power to increase student engagement and to improve learning outcomes" [20] (p. 13).

However, in the South African context, teachers appear to have limited knowledge and experience in using assessment effectively for improving learning and teaching [21–26]. Notwithstanding several initiatives in South Africa to: (i) improve the national curriculum, (ii) support teachers improve their content knowledge and classroom practice, and (iii) develop relevant resources for learners and teachers, there has been limited focus on improving teachers' assessment practices [27–29]. Improving teacher competence for assessment for learning is key to realising education quality in South Africa and elsewhere in the Global South. This requires effective CPD for teachers to use assessment for improving teaching and learning in the classrooms effectively, which this paper considers below.

3. The South African Policy Context

In 1994, the new democratic government of South Africa abolished the nineteen education departments based on race and completely overhauled the education system. The old Christian National Education (CNE) system and syllabi were scrapped, and in 1996, the new education act was passed into law. The act prescribes both the content and manner of the teaching and learning relationship. Part of the preamble to the SA Schools Act, No. 84 of 1996, reads: " [T] his country requires a new national system for schools which will redress past injustices in educational provision, provide an education of progressively high quality for all learners and in so doing lay a strong foundation for the development of all our people's talents and capabilities, advance the democratic transformation of society, combat racism and sexism and all other forms of unfair discrimination and intolerance, contribute to the eradication of poverty and the economic well-being of society, protect and advance our diverse cultures and languages, uphold the rights of all learners, parents and educators . . . " [6] (p. 5).

Since then, the school curriculum has undergone a few revisions, the latest, implemented by the current minister of basic education, Angie Motshekga, is the Curriculum and Assessment Policy Statement (CAPS) introduced in 2012. In the foreword to the Curriculum and Assessment Policy Statement [30] Minister Angie Motshekga states, "The National Curriculum Statement Grades R–12 represents a policy statement for learning and teaching in South African schools . . . " Under the heading 'General aims of the South African curriculum', it is stated: "Inclusivity should become a central part of the organisation, planning and teaching at each school. This can only happen if all teachers have a sound understanding of how to recognise and address barriers to learning, and how to plan for diversity" [30] (p. 5).

CAPS, which is a content-based and time-bound (pace set) curriculum, also reflects an increased emphasis on formal assessment, with common testing being introduced at each of the exit stages of an education phase, namely, Grade 3—foundation phase (FP), Grade 6—intermediate phase (IP) and Grade 9—senior phase (SP). Throughout the revisions, though, the policies of "inclusivity" and the "rights of all learners" were recognised. In addition, assessment for formative purposes was still core to teaching and learning but arguably accorded lesser importance.

4. Methodology

In examining the viability of attaining SDG 4 in the South African context, data from two interrelated empirical studies on teacher professional development by a research team at the Centre for International Teacher Education (CITE) are drawn upon to illustrate the limitations and potential of attaining quality education, examining both the pedagogic and affective dimensions of teaching. These two interrelated studies address the key research questions of teachers engaging learners in the teaching and learning process using learner-centred approaches and how they enable and equip learners with the skills and knowledge to function in a democratic society.

In the one study, conducted in 2018, Grade 3 Foundation Phase teachers from five rural primary schools attended six AfL workshops over a six-month period. Approximately eight months after the workshops, a purposive sample of ten of the teachers who had attended the workshops were observed teaching a lesson, after which a semi-structured interview was conducted with each of them. The teachers were selected on the basis of the way in which they interacted with the workshop facilitators during the workshops and during subsequent classroom support visits. The interviews were related to AfL issues in general, rather than focusing on the techniques and strategies used in a particular lesson. The interview questions included the teachers' views about AfL, what they felt were the advantages and disadvantages of using AfL, and how they saw their roles as teachers in realising their agential space.

The second study that this paper draws on investigated teachers' understandings and experiences of citizenship in South African high schools, conducted in 2016. The study also included students; however, the data for this paper only presents the responses from teachers and a principal. Eight teachers from four schools participated in the study. Disa High is a Quintile 1 school situated in a rural context, Protea High is a Quintile 5 school situated in a rural context, Lily High is a Quintile 1 school situated in an urban context, and Strelitzia High is a Quintile 5 school situated in an urban context. The study included teachers and principals from the Further Education and Training Phase, who volunteered to participate in the study, thus resulting in the study sample. The semi-structured interview schedule used in this study elicited their views on their experiences of their school environment, including issues of safety. It also included their understanding of citizenship, how teachers practice the values of citizenship in their classrooms, teachers' views on citizenship in the curriculum, information about school resources, information about CPD provided to teachers that help teachers develop the skills to teach in a socially cohesive manner, and also their views on how schools may go about producing good citizens.

Ethical clearance for both studies was obtained from the Cape Peninsula University of Technology as well as the Western Cape Education Department. All participants were assured of their and their schools' anonymity and informed that the data obtained would only be used for research purposes. The names and all participants and institutions were changed to protect their anonymity. Participants were also informed that they could withdraw their contributions at any point until the findings were presented, without any fear of repercussions. Both studies were located in the Western Cape Province of South Africa and focused on the pedagogies and learning strategies used by teachers in public schools.

In both studies, qualitative data were analysed using thematic analysis. The analysis identified common themes based on the research focus and which are reported in this paper. In particular, the analysis sought to understand the limits and constraints of teacher agency in realizing meaningful learning in disadvantaged schools and classrooms as a conduit to realizing quality education.

Trustworthiness of the qualitative data was established through credibility, dependability, and confirmability. In both studies, the data were presented to the respondents for feedback and to ensure accurate reporting. The findings of both studies were also discussed with researchers and other research students who form part of a professional

learning community for research conducted on and about the professional development of teachers.

5. Findings

This section discusses the findings of the two empirical studies conducted in South Africa, one relating to the affective dimension of teaching and the other focusing on teacher pedagogy. What these studies demonstrate overall is that both dimensions of teaching need to be developed in order to achieve the imperative of quality education.

5.1. Teachers, Citizenship, and Social Cohesion in South Africa

This section draws on empirical research from a study conducted in South Africa that included an investigation of teachers' practices and experiences in classrooms as they relate to the values of citizenship and social cohesion.

5.1.1. Teachers Do Not Always Promote Inclusive Classrooms

One of the teachers in the study noted that she openly promotes her own religion, including sexual bias, in the classroom, as demonstrated below:

"I will call a spade by name, I will tell them . . . One or two children will say it, I don't think it's right, because the Bible says it, but there's never, for example, if there's a boy in the class that we can see is not [interviewee makes hand gestures] they won't refer to him or be nasty to him." (Protea High, Female, White, Life Orientation and History Teacher, Rural, Q5).

In this instance, the teacher did not see her behaviour as being problematic or that she may, through her actions and behaviour, make the learner feel isolated. This lack of awareness of her discriminatory behaviour could result in undemocratic attitudes towards learners with differing sexual orientations being perpetuated in school and in broader society, thus undermining the realisation of the values of citizenship and social cohesion.

5.1.2. Teachers Do Not Always Respect Learners

A teacher from Strelitzia High noted that teachers often mistreat students through name-calling and insults as demonstrated below:

"There are some teachers that they dislike terribly because they just don't trust them. It's terrible, three teachers in the school I have constant complaints about . . . they will say nasty things, they will name, call you a *vetkop* [fathead], a *domkop* [stupid head], and you're stupid and things like that . . . So, if you call a child a name like that, you're going to get reaction, so I think it's the name-calling, you get so frustrated, so you just call the child a name. Awful letters from children, awful letters of what teachers say to them." (Strelitzia High, Urban, Quintile 5, Male, Coloured, Life Orientation Teacher).

What the quotation above also demonstrates is how teachers abuse their power to bully learners at their school. This abuse of power suggests a lack of social cohesion at the schools and the absence of democratic values. Teachers in this instance demonstrate poor citizenship practices and are not mirroring the kind of schooling environment that enables the realisation of SDG 4 and quality education.

5.1.3. Lack of Resources Limits Teachers' Work in the Classroom

A teacher at Disa High reported that owing to a lack of textbooks, teachers are unable to teach effectively, causing anxieties to teachers as well as students.

"That's the whole thing, we can't give them homework, or make copies or whatever, because many times there is no paper, there's no ink at the school to make photocopies, so I think it actually has a negative impact on the learners themselves, because I can only do what I can do . . . So, we can't actually give homework to the child to learn during the exams. Everything has to happen in the class." (Disa High, Quintile 1, Rural, Life Orientation, Female Teacher).

The lack of textbooks and printing resources reported by teachers at Disa High suggests that teachers' teaching strategies, including how learners learn, are handicapped. This limitation has implications for quality teaching and learning, as both teachers and learners have their potential curtailed because of structural inefficiencies.

5.1.4. Unsafe School Contexts Negatively Impact the Realisation of Citizenship

The principal at Disa High, a remote rural school, noted, in the extract below, that teachers and students do not feel safe at school owing to a lack of school fencing.

"The students and teachers are not safe at school as a result of the fact that there is no fencing, which allows for free entrance of gangsters, and . . . the number of learners at the school that are involved in gangsterism and there is a number of them" (Disa High, Rural Quintile 1, Male, Principal, ID).

Optimal teaching and learning cannot occur when teachers and students are feeling unsafe and where there is a constant and imminent threat of violence. Teachers and students have a right to teach and learn in safe spaces. The lack of safety contradicts the values of citizenship and democracy [31,32].

The findings presented above relating to the empirical evidences of teachers' realities suggest a number of notable implications for the realization of SDG 4, that is, quality education, in South Africa.

First, the path to realising SDG 4 cannot be divorced from redressing the legacies of past political regimes. "Religion and education were synonymous during apartheid", promoting the agenda of "Christianity in all schooling institutions" [13] (p. 137). In a post-apartheid context, there have been policy mandates emphasising equality, inclusivity, and recognition of all religions and that CNE no longer forms part of the democratic dispensation. However, the practices of teachers in this study contradict the instructions and philosophy of these policy mandates. Thus, what this suggests is that teachers need to be professionally developed to teach in a post-apartheid context, particularly teachers who have received their initial teacher education prior to 1994. Teachers also need to be taught how to teach in a manner that is inclusive and that does not promote personal agendas that may be inconsistent with the values of citizenship and social cohesion.

Second, the quotation from the teacher at Strelitzia High suggests that teachers need to be trained to use their power effectively in the pursuit of democracy, citizenship, and social cohesion. Sayed et al. [16] argue that teachers are agents of social cohesion and have the ability to impart these values to students through their actions and interactions. Thus, professional development programmes aimed at addressing this are crucial to realising quality education.

Third, in South Africa, schools that were marginalised during apartheid have not been equally capacitated to the level of affluent public schools in the post-apartheid context [13,16]. Thus, these schools, particularly rural schools, remain under-resourced, which impacts the quality of teaching and learning experiences. In this instance, governmental stakeholders, who are primarily responsible for the provision of quality education in the country, need to be mobilised urgently to address this inefficiency as a conduit to realising SDG 4.

Fourth, quality education is as much about content as it is about form. The context in which teachers operate contributes to teaching experiences they create in the classroom. Cohen et al. [32] argue that when teachers and students feel safe, it contributes to a positive school climate and improves teaching and learning as a result. Many schools in South Africa, particularly in communities with high unemployment rates, high crime rates, and sub-optimal living conditions, operate in unsafe contexts [33], making the realisation of delivering quality education difficult. The challenging teaching and learning context of the majority of South African public schools disables the realisation of citizenship and suppresses efforts towards social cohesion.

Overall, this study suggests that the SDGs are hard to realise, owing to the political, cultural, and social context in which teachers operate. It also suggests that continuous pro-

fessional development is key to realising quality education. Awareness of the deficiencies that teachers possess gives direction to the kinds of interventions required to pursue the goal of quality education.

5.2. Teachers' Pedagogic Shift towards Assessment for Learning (AfL) in the Western Cape Province

This section draws on interview data derived from a purposive sample of ten foundation phase teachers who participated in a series of AfL training workshops. All the teachers in this research project expressed the view that employing AfL to improve teaching and learning was beneficial and thus had positive views about it. The following extracts capture the teachers' positive views of AfL as a pedagogic tool to enhance teaching and learning:

"With these [AfL] tools you will have a different perspective of teaching, you will want to teach again." (Ms Porter, 2018)

"It also helps us with teaching and learning because all the learners in the class get a chance to speak . . . " (Ms Randall, 2018)

"You do AfL throughout the day. Mostly, it is the observations that you do . . . A teacher's observations play a very important role, especially when you work one-on-one with your children, and especially in your group work." (Ms Moyo, 2018)

These quotations point to teachers perceiving AfL in a positive light, as it motivates the teachers (Porter) and facilitates effective teaching and learning in their classrooms, enabling them to reach all the learners in their classrooms (Randall, Moyo). In unpacking these positive views of AfL, three key themes are discussed, namely: (i) AfL facilitates classroom discipline; (ii) the factors that limited the implementation of AfL; and (iii) the need for a more inclusive AfL CPD programme. The three themes draw on teacher comments gleaned from individual teacher interviews.

5.2.1. AfL Facilitates Classroom Discipline

An unexpected outcome of the AfL programme, identified by almost all teachers, was the general impression that the learners in the classes of teachers who are familiar and comfortable with, and are implementing AfL, were more well behaved than their peers. The following extracts point to improved discipline when applying AfL:

"And also the discipline . . . I think for the school we can have the proper learning at the school and also the discipline . . . If ever [you are] using the strategies you will get a disciplined class . . . " (Randall, 2018)

"It [AfL] makes the learners very disciplined." (Nakedi, 2018)

The two quotations, both by Grade 3 teachers from two different quintile (1 and 2) schools in two different townships, underline the positive effect that AfL had in the classes, and on classroom management, in particular. The following example illustrates the positive effect AfL can have on discipline: During a lesson, the class teacher was called to the principal's office, and she asked the student teacher to oversee the class while she was gone. Within minutes, the noise level in the class had risen and the student teacher found it difficult to quieten the learners. She walked to the front of the class, picked up a tennis ball—an AfL tool—and called for silence as she had the ball and hence should be given a hearing. There was immediate silence, she was able to speak to the class without raising her voice and the class remained quiet until the class teacher returned. Significantly, the student had not attended the AfL workshops, but had observed the class teacher applying this technique.

5.2.2. Factors Limiting the AfL Pedagogy to Improve Teaching and Learning

While the section above notes the positive ways in which individual teachers viewed and experienced AfL approaches to enhance their pedagogy, the widespread diffusion of such an approach in schools was limited by several school-level factors.

Staff Deployment Affects Effective Implementation

One contextual factor that affected the diffusion and efficacy of AfL to improve teaching and learning was how schools organised their work and teaching, and in particular how staff were utilised and deployed in the schools to meet curriculum needs. In a particular case, which is illustrative of how some schools organise work, the three teachers at one of the schools—Kojack, Mabile, and Nakedi—were all moved at the start of the new academic year to teaching Grade 2 by school management, and replaced by different Grade 3 teachers who had not been exposed to AfL. The teachers noted that the internal organisation of teacher allocation by the school impacted their use of AfL. Although the three teachers were still applying some of the AfL techniques and strategies in their classes, since they were not given the support by management or colleagues—who were not au fait with AfL—and were not able to use the AfL lesson planning template for support, their enthusiasm would predictably start waning, ultimately leading to the teachers abandoning AfL. One of the three teachers noted:

"The lesson plan of AFLA (Assessment for Learning in Africa) is a little work. They [school management] want a lesson plan that consists of every information with more detail. That lesson plan of AFLA is a little work, that's why we are using the old lesson plan. The principal noticed that lesson plan of AFLA is brief, brief, brief now they say they want a lesson plan with more detail [i.e., the official, school lesson plan]." (Kojack, 2018)

Lack of Resources Impacts Effective Implementation

An associated factor regarding lesson plans, especially within poorer schools, is a lack of resources. As mundane as it may sound, in poorer schools, teachers have to consider the cost of printing. One of the teachers raised the issue that using the AfL lesson plan template would imply using more paper and ink:

"The challenge is more paper . . . you gonna have five lessons [i.e., lesson plans] per week [Foundation phase teachers at this school used one lesson plan per subject per week] . . . now most use the papers . . . the papers and the ink." (Nakedi, 2018)

At another school, Ms Porter's AfL equipment had been stolen at the end of the year, and she lamented that she would have to replace them from her own pocket, as the principal claimed that the school did not have the funds to finance their replacement.

What emerges from these two accounts are the effects lack of resources have on the effective implementation of AfL in poorer schools. On the one hand, teachers have to consider the amount of printing they are allowed to use in the normal carrying out of their function as teachers, and on the other, a lack of funds stymies the teacher in accessing the equipment she feels she requires to carry out her function as a teacher implementing AfL effectively.

In both cases, the lack of resources and the lack of support from school management led to a drop in the efficacy with which the teachers could apply AfL. In the first case, one finds school management being inflexible in the way in which teachers are expected to prepare lessons—and the perceived additional cost of alternative/additional lesson preparation sheets. In the second, a lack of funds/resources and management's reluctance to replace lost equipment led to a teacher's loss of enthusiasm in implementing AfL.

Lack of Continuity Hinders Effective Implementation

Third, an issue that was of concern regarding the effective implementation of AfL at the schools was the lack of continuity across the grades. The teachers who participated in the research expressed their concern that the AfL programme was not part of a systematic, whole-school improvement effort. Many of the teachers felt that it would have been more effective had the teachers of the whole school been exposed to AfL, thus allowing for continuity in the implementation of AfL. Some of the teachers' comments in this regard are:

"I have recommended it [AfL] for the whole school because now the Grade 3 child does it but when it comes to Grade 4 it stops so I have recommended it . . . so that the child continues with it" (Wilson, 2018).

"More special for the intermediate phase, you see, it's gonna be the challenge because of the . . . changing of the classes [from Grade 3 to Grade 4], you see . . . " (Randall, 2018).

The teachers felt that it would have been more effective for the learners had the programme been implemented in the whole phase—or even in the whole school—to allow for continuity. For the Grade 3 learners, this would be an even greater challenge, as they would also have to deal with the transition to intermediate phase, which is characterised as being more regimented; learners more often sit in rows rather than groups, lessons are more teacher-centred, and more emphasis is placed on individual learning.

The concern raised by the teachers is valid for a number of reasons, the most compelling being continuity both for learners and teachers. With AfL only being implemented in one of the grades, the learners would have to adapt to a different learning approach in the next grade, and the advantages of the AfL approach would be lost. Not only could this be detrimental to the learners, but it could serve as a demotivating factor for the teachers, especially since there is no support from colleagues who are not familiar with the approach.

From the findings presented above, it can be seen that the teachers who had been introduced to AfL felt positive about the approach. They recognised the benefits both for teaching and learning. As such, most of them were seen to be drawing their learners into the teaching and learning space.

An unexpected outcome of the introduction of AfL in the classes was an improvement in classroom management and discipline. This phenomenon of improved discipline in classes utilising AfL is significant in the context of South African schooling, where issues of classroom discipline are a constant concern.

What also emerged from the data was that, despite the teachers' positive views of AfL, this was not without its challenges, which has implications for the realisation of SDG 4. Owing to various contextual factors, the teachers experienced challenges in implementing AfL.

Firstly, the way in which schools utilise their staff from one year to the next has implications for any form of staff developmental programme. This suggests that whatever CPD programme is introduced, demands that the particularities of the school be considered and be adapted to suit the needs of the school.

Secondly, in the lower quintile schools (highly impoverished schools) the lack of resources and facilities could impact the efficacy of the programme and how it is implemented.

Thirdly, for any CPD programme to be implemented successfully, there has to be consistency and support. None of the schools were consistent in the application of AfL beyond the grade in which it was being implemented—except where the teachers progressed with the learners, but then only to Grade 3 (in two of the six schools). There was no extrinsic motivation for the teachers to continue with AfL practices as the learners would find themselves in "conventional" classes the following year. The teachers in the study expressed their concern that AfL was only applied in certain classes in the foundation phase, while they felt that its implementation would be more effective if it were to be done throughout the school. In this way, maximum benefit would be derived from the programme, as identified by Wiliam and Thompson [34].

6. Conclusions

The findings presented above relating to the empirical evidence of teachers' realities point to a number of notable implications for the realisation of SDG 4, that is, quality education, in South Africa.

First, the path to realising SDG 4, particularly the commitment to education systems, starts with qualified and competent teachers. However, these cannot be divorced from redressing the legacies of past political and colonial regimes in a country like South Africa. In the post-apartheid context, there have been numerous policy mandates emphasising equality, inclusivity, and recognition of all cultures and religions. The introduction of Curriculum 2005 and subsequent curriculum revisions have also emphasised the utilisation of assessment to facilitate learning. However, some of the practices of the teachers in the two

studies and their comments are contrary to the instructions and philosophy of the policy mandates. What this suggests is that teachers need to be professionally developed further to teach in a post-apartheid context. Teachers also need to learn how to teach in a manner that is inclusive and that does not promote personal agendas, which may be inconsistent with the values of citizenship and social cohesion. Teachers need to be trained to use their power effectively to give expression to the policy mandates. Thus, professional development programmes aimed at addressing this are crucial towards realising quality education.

Second, in South Africa, schools that were marginalised during apartheid have still not been equally capacitated to the level of affluent public schools in the post-apartheid context. Thus, these schools, particularly those in rural settings, remain under-resourced, which serves as a barrier to the effective delivery of quality teaching and learning experiences.

Third, the South African case illustrates that quality education is as much about content as it is about form. The context in which teachers operate contributes to the teaching experiences they create in the classroom. Many schools in South Africa, particularly in communities with high unemployment rates, high crime rates, and suboptimal living conditions, operate in unsafe contexts, making the realisation of delivering quality education difficult. This suggests that for the delivery of quality education to be effected, the socio-economic challenges of those communities should necessarily also be addressed.

Fourth, the way in which schools utilise their staff from one year to the next has implications for any form of staff developmental programme. This suggests that wherever a CPD programme is to be introduced, it is imperative that the particularities of the school be considered and the programme be adapted to suit the needs of the school. In addition, for any CPD programme to be implemented successfully, there has to be consistency and support on the part of all the role players, especially peers and management.

Fifth, South African empirical studies reveal that teachers' beliefs and values matter in the way they teach, the way in which they relate to and interact with learners, how they relate to their peers, and how they work towards realising the goals of equity in and through education. This suggests that any reform efforts in Africa, and possibly globally, need to take cognisance of the beliefs and values teachers hold.

Sixth, the two interrelated studies of teacher professional development in South Africa speak to how teachers are able to realise their agency for change in contexts where there is a broad range of expectations thrust on them, suggesting that the global education agenda runs the real risk of overstating the potential of schools and their teachers to effect broad social transformation. Teacher agency, as envisaged in the global education agenda, is not a realistic possibility, nor is agency possible when faced with multiple and conflicting demands in highly unequal societies scarred by legacies of oppression. Yet, the paper also points to the possibilities of progressive change in the classrooms when teachers are professionally supported, empowered, and working within structures and systems that affirm their agency.

What this paper reveals is both the limitations and possibilities of realising quality education in a still highly stratified and unequal society such as South Africa. As much as the South African education policy seems to resonate with SDG 4, the inequities with which teachers and learners in both privileged and marginalised contexts are confronted in South Africa point to the need for serious and drastic interventions from government, national, and international stakeholders in education to realise the goal of equitable and quality education for all.

Author Contributions: Conceptualisation: Y.S., M.S. and T.N.; methodology: T.N. and M.S.; software: M.S.; validation: T.N. and M.S.; formal analysis: T.N. and M.S.; investigation: T.N. and M.S.; resources: T.N., M.S., and Y.S.; data curation: T.N. and M.S.; writing—original draft preparation: T.N., M.S., Y.S.; writing—review, additions, and revisions: Y.S.; supervision: Y.S.; project administration: T.N. and M.S.; funding acquisition: Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the generous financial assistance of the National Research Foundation (NRF), which funds the South African Research Chair in Teacher Education. In addition, part of the work reported in this paper was part of a collaborative Assessment for Learning in Africa project led by Professor Therese N. Hopfenbeck. Opinions expressed in this paper and the conclusions resulting from these studies are those of the authors and are not necessarily to be attributed to the National Research Foundation. The NRF was not involved in the production of this paper.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Cape Peninsula University of Technology (EFEC 2-11/2017 dated 08 November, 2017, EFEC 4-3, /2016 dated 03 March 2016, and the South African Western Cape Education Department Ref: 20160224-8119).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Participants were assured that their responses would remain confidential and that publications resulting from the data will not reveal their identity.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. UNESCO. Position Paper on Education Post-2015. 2014. Available online: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/ED/ED_new/UNESCOPositionPaperOnEDpost-2015.pdf (accessed on 15 December 2020).
2. United Nations. Department of Economic and Social Affairs. Sustainable Development Goal 4c. Available online: https://sdgs.un.org/goals/goal4 (accessed on 15 December 2020).
3. UNESCO. *Education 2030. Incheon Declaration and Framework for Action for the Implementation of Sustainable Development Goal 4: Ensure Inclusive and Equitable Quality Education and Promote Lifelong Learning Opportunities for All*; UNESCO: Paris, France, 2016. Available online: http://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en_2.pdf (accessed on 29 January 2021).
4. Badat, S.; Sayed, Y. Post-1994 South African education: The challenge of social justice. *Ann. Am. Acad. Political Soc. Sci.* **2014**, *652*, 127–148. [CrossRef]
5. Sayed, Y.; Ahmed, R. Education quality, and teaching and learning in the post-2015 education agenda. *Int. J. Educ. Dev.* **2015**, *40*, 330–338. [CrossRef]
6. Department of Education. South African Schools Act 84 of 1996. *Gov. Gaz.* **1996**, *377*, 1–50. Available online: https://www.gov.za/sites/default/files/gcis_document/201409/act84of1996.pdf (accessed on 15 December 2020).
7. Henard, F.; Leprince-Ringuet, S. The path to quality teaching in higher education. In *Working Document Draft 18 November 2008, FH, SLR*; OECD: Paris, France, 2008; pp. 1–15. Available online: http://www.oecd.org/education/imhe/41692318.pdf (accessed on 15 December 2020).
8. Babalola, J.B. Quality assurance and child friendly strategies for improving public school effectiveness and teacher performance in a democratic Nigeria. In *Management of Primary and Secondary Education in Nigeria*; Fagbamiye, E.O., Babalola, J.B., Fabunmi, M., Ayeni, A.O., Eds.; NAEP: Ibadan, Nigeria, 2004; pp. 303–312.
9. Madani, R.A. Analysis of educational quality, a goal of Education for All Policy. *High. Educ. Stud.* **2019**, *9*, 100–1019. [CrossRef]
10. Skott, J. The promises, problems, and prospects of research on teacher's beliefs. In *International Handbook of Research on Teacher Beliefs*; Fives, H., Gill, M.G., Eds.; Routledge: New York, NY, USA, 2015; pp. 13–30.
11. Pajares, M.F. Teachers' beliefs and educational research: Cleaning up the messy construct. *Rev. Educ. Res.* **1992**, *62*, 307–332. [CrossRef]
12. Singh, M.; Wessels, H.; Kanjee, A. Beliefs about professional knowledge. In *Learning to Teach in Post-Apartheid South Africa: Student Encounters with Initial Teacher Education*; Sayed, Y., Badroodien, A., Carrim, Z., McDonald, Z., Singh, M., Eds.; SUN Media: Stellenbosch, South Africa, 2018; pp. 73–105.
13. Singh, M. Citizenship in South African Schools: A Study of Four High Schools in the Western Cape. Ph.D. Thesis, Cape Peninsula University of Technology, Cape Town, South Africa, 2020. Unpublished.
14. Stipek, J.; Givven, K.B.; Salmon, J.M.; MacGyvers, V.L. Teachers beliefs and practices related to mathematics instruction. *Teach. Teach. Educ.* **2001**, *17*, 213–226. [CrossRef]
15. Beswick, K. Influencing teachers' beliefs about teaching mathematics for numeracy for students with mathematical learning difficulties. *Math. Teach. Educ. Dev.* **2008**, *9*, 3–20.
16. Sayed, Y.; Badroodien, A.; Omar, Y.; Balie, L.; McDonald, Z.; de Kock, T.; Salmon, T.; Raanhuis, J.; Singh, M.; Robinson, N.; et al. Engaging teachers in peacebuilding in post-conflict context: Evaluating education interventions in South Africa. In *South Africa Country Report: ESRC/DFID Research Report*; University of Sussex: Brighton, UK, 2017.
17. Merry, M.S. Can schools teach citizenship? *Discourse: Stud. Cult. Politics Educ.* **2020**, *41*, 124–138. [CrossRef]

18. Alexander, R. *Education for All, the Quality Imperative and the Problem of Pedagogy*; Institute of Education, University of London: London, UK, 2008.
19. Westbrook, J.; Durrani, N.; Brown, R.; Orr, D.; Pryor, J.; Salvi, F. *'Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries' Final Report, Education Rigorous Literature Review*; Department for International Development: London, UK, 2013.
20. Wiliam, D. What is assessment for learning? *Stud. Educ. Eval.* **2011**, *37*, 3–14. [CrossRef]
21. Kanjee, A.; Mthembu, J. Assessment literacy of Foundation Phase teachers: An exploratory study. *S. Afr. J. Child. Educ.* **2015**, *5*, 142–168. [CrossRef]
22. Kanjee, A.; White, C.J. *Evaluation of a National Professional Development Programme to Improve Teacher Formative Assessment Practices*; Unpublished Report; Department of Educational Studies, Tshwane University of Technology: Pretoria, South Africa, 2014.
23. Kuze, M.W.; Shumba, A. An investigation into formative assessment practices of teachers in selected schools in Fort Beaufort in South Africa. *J. Soc. Sci.* **2011**, *29*, 159–170. [CrossRef]
24. Mkhwanazi, H.N.; Joubert, I.; Phatudi, N.C.; Fraser, W.J. Teachers' use of formative assessment for the teaching of reading comprehension in Grade 3. *Mediterr. J. Soc. Sci.* **2014**, *5*, 468–475. [CrossRef]
25. Schuld, N.; Kanjee, A.; White, T. Tinder or flint: Igniting Grade 2 teachers' understanding of learning, teaching and assessment. *J. Educ.* **2017**, 109–132. Available online: http://joe.ukzn.ac.za (accessed on 28 January 2021).
26. Kanjee, A. Exploring primary school teachers' use of formative assessment across fee and no-fee paying schools. *Outh. Afr. J. Child. Educ.* **2020**, *10*, a824. [CrossRef]
27. Pryor, J.; Lubisi, C. Reconceptualising educational assessment in South Africa—Testing times for teachers. *Int. J. Educ. Dev.* **2002**, *22*, 673–686. [CrossRef]
28. Killen, R.; Vandeyar, S. Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised National Curriculum Statement hold? *Perspect. Educ.* **2003**, *21*, 119–134.
29. Kanjee, A.; Sayed, Y. Assessment policy in post-apartheid South Africa: Challenges for improving education quality and learning. *Assess. Educ. Princ. Policy Pract.* **2013**, *20*, 442–469. [CrossRef]
30. Department of Basic Education. Curriculum and Assessment Policy Statement. Available online: https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements(CAPS).aspx (accessed on 15 December 2020).
31. Feu, J.; Serra, C.; Canima, J.; Lazaro, L.; Simó-Gil, N. Democracy and education: A theoretical proposal for the analysis of democratic practices in schools. *Stud. Philos. Educ.* **2017**, *36*, 647–661. [CrossRef]
32. Cohen, J.; McCabe, L.; Michelli, N.M.; Pickeral, T. School climate: Research, policy, practice, and teacher education. *Teach. Coll. Rec.* **2009**, *111*, 180–213. [CrossRef]
33. Burton, P.; Leoschut, L. School Violence in South Africa: Results of the 2012 National School Violence Study. In *Centre for Justice and Crime Prevention Monograph No. 12*; Centre for Justice and Crime Prevention: Cape Town, South Africa, 2013.
34. Wiliam, D.; Thompson, M. Integrating assessment with learning: What will make it work. In *The Future of Assessment: Shaping Teaching and Learning*; Dwyer, C.A., Ed.; Routledge: New York, NY, USA, 2008; pp. 53–82.

education sciences

MDPI

Article

Towards Acquiring Teachers' Professional Qualification Based on Professional Standards: Perceptions, Expectations and Needs on the Application Process

Tiiu Leibur [1,*], Katrin Saks [1] and Irene-Angelica Chounta [2]

[1] Institute of Education, University of Tartu, 50103 Tartu, Estonia; Katrin.Saks@ut.ee
[2] Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, 47057 Duisburg, Germany; irene-angelica.chounta@uni-due.de
* Correspondence: tiiu.leibur@ut.ee

Abstract: This paper presents the findings of a study into the application process of the professional standards (PS) qualification in Estonia, which is a standard used to identify the level of professionalism of a teacher. The study collected the views of active teachers who have applied for the PS and those of the assessors to understand the shortcomings in the application and the expectations for the application process. The main challenges identified are that teachers lack respective skills to adequately comprehend qualification requirements (unambiguous interpretation of PS) and the lack of evidence-based performance indicators (difficult to prove) and weak self-analysis skills; hence, the teacher needs assistance. The research found that teachers need collaborative support, motivation and constructive feedback on their professional development (PD), both from colleagues and at school and national levels. Based on this research, it can be said that teachers who have gone through the application process have gained greater self-confidence, value themselves as teachers and continue their in-service training.

Keywords: teacher professional development; professional standards; professional qualification

Citation: Leibur, T.; Saks, K.; Chounta, I.-A. Towards Acquiring Teachers' Professional Qualification Based on Professional Standards: Perceptions, Expectations and Needs on the Application Process. *Educ. Sci.* **2021**, *11*, 391. https://doi.org/10.3390/educsci11080391

Academic Editors: Beng Huat See and Rebecca Morris

Received: 29 May 2021
Accepted: 27 July 2021
Published: 30 July 2021

1. Introduction

The world around us is constantly changing, setting new challenges for members of society. Economic and technological changes in society add increasing responsibility to teachers' work, and teachers must be prepared to apply their pedagogical competencies as knowledge and skills in order to cope in a rapidly changing society. Economic change has led to the globalization of society, competition between schools and market orientation [1], which requires teachers to be responsible, rational and also competitive.

The so-called "new professionalism" of the era [2] presents new challenges for the teacher. The teacher's role has expanded, becoming a provider of knowledge, counsellor, supervisor, supporter, social worker, problem solver, etc. This has brought with it higher expectations of teacher professionalism. As teachers' work expands beyond the classroom, new emotional and intellectual demands are placed on them. Thus, the teacher must be collegial and collaborate with different target groups and focus on new forms of communication and collaboration [3] inside as well as outside the school [4].

There has been an increase in attention to the lifelong development of teachers throughout their careers [5]. As a result, there are many competing views and concepts on professionalism. The concept of professionalism has been approached differently over time.

In describing his stages, Hargreaves [6] was guided by the concept of professionalism, which is related to the status of the teaching profession in society. He [6] characterized professionalism through four stages: the pre-professional stage, the stage of the autonomous professional, the stage of the collegial professional and the stage of the post-professional or postmodern professional. In the period of pre-professionalism, the teaching was based

on traditions and standardized curricula. The next era of autonomous professionalism was characterized by teachers' individualism, isolation and privacy. The period of collegial cooperation is characterized by educational reforms, i.e., the diversity of teaching methods, the increased role of the social worker and the increase in the number of learners with special needs. In the postmodern (or post-professionalist) era, the school has been heavily influenced by global economic competition and has reduced teachers' autonomy. Hoyle [7], on the other hand, defined professionalism as restricted (for the teacher teaching was an intuitive activity in a classroom environment that did not engage much broader professional reading or activities and was based mainly on experience) and extended professionalism (interpreted as a teacher's collegial activity seeking to improve the practice through reading and continuous professional development, later referred to as PD). Pratte and Rury [8] defined professionalism as a set of educational professional values. Eraut [9], Goodson and Hargreaves [3] and Evans [10] defined the concept of *new professionalism*, which includes performance standards, public accountability, lifelong PD of professionals, collaboration with colleagues and stakeholders and professional innovation where three interrelated concepts—knowledge, autonomy and responsibility—are considered to form the core. Sachs [11], on the other hand, described teacher professionalism as a collegial relationship and collaboration, while organizational professionalism is part of management professionalism that focuses on performance and accountability.

Holroyd [12] summed it up as "professionalism is not some social-scientific absolute, but a historically changing and socially constructed concept-in-use". Thus, professionalism has been studied for a long time, and the definitions differ in both content and purpose. Professionalism is not static but in constant change, caused by the era and national interest. It is closely related to education policy and the needs of the national educational system.

The current study is grounded in the concept of professionalism of Snoek et al. [13,14], which is related to the output-based and competency requirements of the modern education system, an important part of which is the implementation of professional standards (PS) in improving the quality of education (see Section 1.1). We explore how to support a teacher to become a professional in their field.

The PD of teachers is a major educational issue in many countries. Changes in society pose new challenges for schools. Teachers have a key role in the process of the implementation of the changes, which requires extended professionalism from them [15] and should ensure the quality of the education provided [13,16]. Thus, the professionalism and PD of teachers play an important role to ensure the quality of education [15].

One strategy for assuring the increase in professionalism is using professional standards (PS) and competency frameworks that focus on teacher training, certification, continuing PD, career advancement and performance evaluation [17]. PS are the national basis for developing the knowledge, skills and professionalism required of teachers [18]. However, PS alone do not provide a solution to problems in education [19]. Based on previous research, it has been found that more qualified teachers ensure better learning outcomes for students [20] and are particularly productive at the time of applying for a qualification [21]. According to Darling-Hammond [19], "professional standards are not magic" formulae that can solve various educational problems (outdated curricula, unequal distribution of resources and lack of social support for children), but they support teachers' self-development and PD, academic knowledge and professional knowledge and the ability to apply this knowledge ethically and appropriately [22]. PS-based self-analysis helps teachers to identify their strengths and weaknesses and also highlights the need for further training [23] and development. Hayes [24] argued that teachers should have an opportunity to contribute to the development of PS, albeit in collaboration with stakeholders in the education community. PS as one of the key concepts are explored in more detail below (see Section 1.3).

The PS for teachers have been in use in Estonia since 2005. The aim of the PS is to support teachers' pre-service and in-service training, assess teachers' competencies and support them in their continuous professional development throughout their career [25].

Although the university may reward the graduates of master's studies with the certificate of professional qualification (if they wish), our study focuses on working teachers who want to upgrade their qualifications (for more details, see Section 1.4). Professional qualification (PQ) is raised and a professional exam is held on the basis of PS (Section 1.4). However, applying for a PQ is not a regular practice for all teachers, and hence, this study seeks to identify teachers' lack of interest in going through the process of applying for the qualification. Although the impact of the PS on teacher quality [26,27] and support for professional development [21,26] has been studied, little is known about the expectations and needs of teachers in both the PS and the application process.

Previous research conducted in Estonia [25] has shown that the PS have a strong impact on teacher training, but the PS were not found to influence teacher careers and their PD. However, it was shown that in-service training increases teachers' self-esteem and professional autonomy, although applying for a qualification level is their free choice, i.e., it depends on teachers' own views, attitudes and intrinsic motivation [28]. We assume that the level of qualification acquired and the accompanying moral and material recognition will certainly help to reduce the turnover and attrition of teachers.

The goal of this research is to understand the shortcomings of the application process and the needs and expectations of those who apply. The authors of the paper also try to understand the reasons for teachers' low interest in applying for the qualification and to explore how to support teachers in their professional development and in the process of applying for a PQ. The present research defines PD as a process of systematic self-analysis of teachers' competencies and planning of self-improvement (continuous professional development) based on PS that ensures the overall quality of teaching [25]. As research has shown, the content of PS is relatively similar across countries. As professional standards have been developed in Estonia in cooperation with representatives of different communities, the sharing of experiences on the shortcomings of applying for a profession and implementing professional standards, as well as suggestions for improvement, will certainly be of interest to international readers as well. It would give others the opportunity to compare the success and performance of their professional standards and application process with the Estonian experience.

For our research purposes, we analyse the different stages of applying for the teaching profession, based on the currently operating PS and the process of applying for the qualification in Estonia, and highlight the aspects and proposals that need to be developed. Proceeding from the goal of the research the following research questions (RQs) were posed:

Research Question 1 (RQ1): *What are the shortcomings perceived by the different target groups under the current criteria and the application process?*

Research Question 2 (RQ2): *What are the expectations and needs of different target groups for professional standards and professional application procedures?*

1.1. Theoretical Framework

The concept of professionalism used in the work focuses on the expectations of today's competitive and knowledge-based society. This research is based on the concept of professionalism by Snoek et al. [13,14]. Consequently, the requirement of professionalism, i.e., responsibility for one's professional activity and quality, is the core of the concept. Strengthening professional responsibility is closely linked to teaching and learning, curriculum renewal, peer support and cooperation and PD. Professional teachers are not "service" providers. They monitor their PD and training needs in order to keep abreast of modern educational innovations and thus constantly improve their knowledge (professional, academic and technological), i.e., they are lifelong learners. It is also important to work with different target groups (colleagues, students, parents) and with different communities, supporting innovation and the application of new knowledge and skills. The development and improvement of PS is also an area of competency for joint community action. An

important feature of professionalism is the ability to reflect and the ability to apply theoretical knowledge in practice. This concept forms a holistic approach to professionalism, which is based on lifelong professional learning and recognizes the development of an autonomous professional.

The definition of professionalism relies on professional autonomy, professional application and further PD, PS, academic knowledge consisting of formal or technical knowledge, lifelong learning, cooperation inside and outside the profession, responsibility for professional quality, use of an expanded knowledge base, developing research and supporting innovation and commitment. These criteria characterise the role of the teaching profession in knowledge-based competition, lifelong learning and the innovations that accompany it, as well as the responsibility for the results and quality of the work [13].

The research also draws on the European Qualifications Framework and the Estonian Qualifications Framework, derived from the teacher competency model [29] which allows the identification of a teacher's PD and the continuation of their career. The qualification framework is defined as a means to achieve a common understanding of quality and content [18]. In this research, standards are considered both as evaluation criteria for determining competencies and as criteria describing the recommended level of achievement [30].

1.2. Professional Standards for Teachers

The aim of PS is to support the quality of teaching and the quality of teachers and can be described in general as a broad set of descriptions that teachers can strive for but also specific, that is, clearer descriptions, allowing evaluators to evaluate teachers' work based on evaluation criteria [31]. The concept of the standards is to define the quality or the level of achievement, i.e., the measure of teachers' qualities against the requirements of the framework (comparative assessment). PS are also seen as a tool in competition and for ranking. At the same time, standards can serve as goals in striving for idealism, thus being a driving force in PD [32].

PS serve as a measurement tool for determining teachers' competency [30], makes teachers' work transparent, treats teachers' work as an agreed set of rules in teaching [33] (p. 13) and can be used to measure teachers' effectiveness [11]. PS can also be used to study teachers' professional behaviour, improve teachers' performance and develop alternatives, i.e., standards as a reference framework and basis for self-regulation [34]. PS also serve as a decision-making tool based on common understandings and values [35], i.e., it not only measures but also shows how evidence is gathered, the extent to which the standards are met and how PS are used to evaluate performance [36]. PS provide a structure that teachers can use to analyse their work. The standards create a common understanding of competencies, which in turn support collaboration between different target groups [37]. PS as a measurement tool in the professional assessment framework give an idea of the ideal teacher, which on the one hand, can be a driving force for the need for self-development. On the other hand, it can create uncertainty for teachers about the accuracy of their actions [38]. PS are also seen as a tool for improving teachers' reputation [39]. They play an important role in assessing teachers' PD, providing opportunities for certification and recognition [40]. It is very important to understand that PS alone do not determine the need for in-service teacher training [37] but rather display the combined effect of teacher self-analysis and PS. Self-assessment of the achievement or non-achievement of goals affects teachers' self-efficacy and decisions made. Self-assessment increases teachers' self-responsibility and develops their key skills in working life, i.e., being reflective, self-governing and independent [41]. This research looks at PS as an opportunity for PD for a teacher with the aim of improving their professional quality throughout their career [11] and as an opportunity to manage their own development.

In most countries, PS have been implemented with the aim of improving the quality of the education and school system [22]. Teacher qualification frameworks and PS provide an overview of expectations of a teacher's knowledge and skills at the national education

policy level at different stages of a teacher's career [18]. Although the governments have a constitutional responsibility for ensuring the quality of education, this can only be realised through cooperation between the ministries, schools and teachers, recognizing the role and responsibility of everyone [42]. When setting PS, it is agreed how good is good enough, i.e., setting an acceptable level of performance for the purpose. For example, whether the teacher has reached a higher standard level [43].

Ingvarson and Kleinhenz pointed out [23] the fact that PS in different countries have common structural features and a similar process of converging [44]. That is, the PS are based on a common understanding of the competencies required of teachers. For example, Belgium, Estonia, the Netherlands and the United Kingdom have implemented specific competency frameworks (in the form of knowledge, skills and attitudes and value lists by subject areas) that differentiate between beginner and advanced teacher levels [45].

1.3. Professional Standards and Their Implementation

Based on the needs of educational policy, the PS are constantly evolving and changing, which should take place in cooperation with teachers and stakeholders in educational communities [11,24]. This evolvement is based on the development of society and expectations for the quality and work of the teacher, which are closely interrelated. When compiling PS, it is important to rely on logic and monitor the use of language to avoid overlaps and repetitions. PS must also be regularly adjusted in the light of research and professional knowledge [43].

Research on the career of European teachers [46] shows that competency frameworks are mainly used for teacher learning, for defining curricula and learning outcomes and as a reference framework for in-service teacher training. They are rarely used to support teachers' careers (through the transition to new levels), and teachers in about half of European countries have a career structure with no prospects for advancement at all [47].

PS may differ from country to country in terms of structure and description of objectives. In Australia, PS have been developed on a subject basis and teachers are required to be accredited at the basic level of professional competency, with the aim of encouraging teachers' PD. The goal of the PS is to improve the quality of teaching and their professional status [26]. In the UK, PS are compulsory and also linked to the salary. They describe competencies at five levels (from a qualified teacher to an advanced teacher). PS have also been developed in the USA, but they are used differently [27]. In the USA, PS are used to improve the quality of teaching and teachers by supporting teacher preparation, remuneration, certification and professional development [48]. PS are applied in teacher training programs in the countries that have adopted a standard-based approach, such as Australia, Canada, Chile, the United Kingdom, Germany, New Zealand, the Philippines, Scotland, Singapore, Thailand, the United States, etc. [43]. In Estonia, PS support the teachers' professional training throughout the pre-service training and career [25]. For the teaching profession, an eight-level qualification framework was developed in 2005 and implemented in 2008 [29]. The Estonian Qualifications Framework (EstQR) combines national curricula for general education, vocational training standards, higher education standard and PS [49].

Teachers who have already received a professional certificate feel more valued professionally [28], which may increase teachers' motivation to assess their PD and apply for a higher qualification. Thus, the PS are not "elite" standards but standards that all qualified teachers should be able to achieve, given their continuous professional development. Passing the assessment fulfils an important psychological function—it provides the teacher with a professional certificate based on the assessment of the quality of the applicant's practice by respected, knowledgeable and trained colleagues. Professional certification confirms that the teacher is able to meet higher PS [30].

1.4. The Process of Applying for a Profession in Estonia

PS for teachers, launched in Estonia in 2005, emphasize the teacher's role as a reflective practitioner and a lifelong learner with the aim of supporting a teacher throughout their teaching career [25]. In Estonia, PS support the teachers' professional training throughout the pre-service training and career [25]. For the teaching profession, an eight-level qualification framework was developed in 2005 and implemented in 2008. The original PS were one-tier standards and described competencies for every teacher without specifying their level in three areas: planning and managing the learning process (including setting up the learning environment, mentoring, analysing and assessing student development) and learning (interpersonal competencies), communication and collaboration skills to support student motivation and PD and self-analysis. This formulation did not allow peers to distinguish between novice and experienced teachers [25].

In Estonia, PS have been in constant development. In 2013, the Education Professional Council developed PS, formulating the standards on different levels. In cooperation with various parties (teachers, educational experts, etc.), new PS for teachers were prepared and approved. As an innovation, a description of a teacher's professional activity was introduced at three different levels: teacher—level 7, senior teacher—level 7 (1) and master teacher—level 8 [50]. The new PS describe the competencies of teachers in the areas of planning learning and teaching activities; designing the learning environment; supporting learning and development; reflection and professional self-development; coaching and mentoring; and research and development, plus recurrent competencies and optional competencies: teacher level 7—supporting pupils with SEN, level 7 (1) senior teacher—leadership or teacher supervision and training; level 8 master teacher—either teacher supervision and training, development of methodology and learning materials or leadership (one option). The profession (higher qualification) can be applied for by all teachers who have a master's degree and work as a teacher.

The Vocational Committee (VC) is responsible for the development and updating of the PS, the procedure of awarding the qualification and resolving disputes related to it. Since 2014, it has been possible to apply for a qualification through the Teachers' Union. The Teachers' Union is a teachers' professional association with voluntary membership and institutional members and includes a number of teachers' subject associations. The focus of the Teachers' Union is on emphasizing the role of the teacher, representing and protecting the professional interests of teachers in Estonian society. For awarding certifications on PQ, the Teachers' Union convenes a Vocational Assessment Committee (VAC), where teacher-practitioners with the highest (master teacher) qualifications (29 teachers) act as voluntary assessors.

At present, the application procedure for PQ consists of three interrelated steps (Figure 1):

At present, the system of applying for a PQ is based on the principle that the Vocational Assessment Committee (VAC) gives the applicant its assessment on the basis of the documentation submitted by the applicant and the interview conducted. The committee assesses the competency of the applicant based on the PS criteria and compliance with the performance indicators. The Vocational Committee (VC), consisting of representatives of educational associations and universities (10 people), approves the reasoned proposal of the VAC to issue the qualification or not. It is a multi-step process (Figure 1) accomplished by two different committees (Vocational Assessment Committee and Vocational Committee), with the common goal of providing an objective and transparent assessment of teacher competency and supporting teacher PD while providing constructive professional feedback.

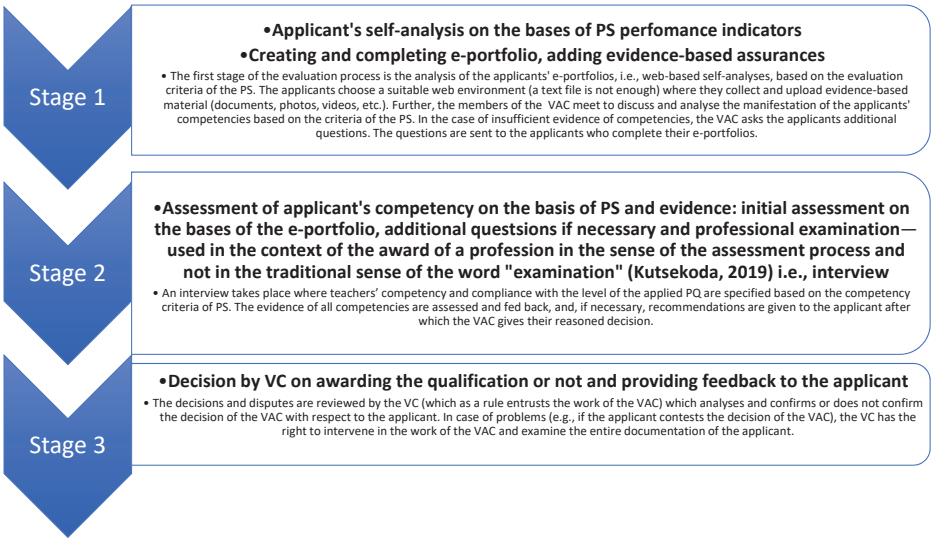


Figure 1. Process of professional application.

2. Methods

2.1. Context and Participants

In order to research the opinions and suggestions of the participants in the process (teachers who received the qualification, teachers who evaluated the applicants' suitability to the profession and one of the developers of PS), qualitative research was conducted. Qualitative analysis was found to be most suitable to investigate teachers' opinions and suggestions, providing the best opportunity to gather and understand their perceptions [51]. The phenomenological research method was used to gain an overview of the experience of applying for the qualification and to describe and interpret the universal nature of the interviewees' experience [52]. In order to obtain detailed, profound answers, semi-structured open-ended interviews were used [53]. Through open questions, teachers' attitudes and opinions were gathered in individual discussions while ensuring the coverage of shortcomings in PS and in the application process providing suggestions for their improvement.

Proceeding from the aim of the research to investigate the teacher applicants', assessors' and standards developer's perceptions of the PS and the application process, targeted sampling was used. The target group was selected based on their experience in applying for the qualification. The study focused on teachers who were employed and had experience in applying for the qualification. Applying for the professional qualification whilst in teacher training is a completely different process compared to that of already working teachers (not covered in this study). Thus, the target groups consisted of seven teachers, the inclusion criteria being:

- Working as a teacher;
- Being linked to the process of applying for qualifications (applicant, assessor, developer);
- Having acquired the teacher qualification.

The participants (Table 1) were all currently working teachers.

The PQ system for teachers is divided into four levels, i.e., level 6 (kindergarten teachers); level 7 (teacher)—a teaching profession that is lifelong and does not need renewal; and level 7 (1) (senior teacher) and the highest level 8 (master teacher)—also need to be renewed after every 5 years.

Table 1. Background data of the participants.

Participant	Gender	Professional Experience as a Teacher (in Years)	PQ Level
T1	F	35	Level 7 (1) *
T2	F	13	Level 7 (1)
T3	F	40	Level 8
A1	F	43	Level 8
A2	M	15	Level 8
A3	F	45	Level 8
D	F	15	Level 8

Notes: * (1)—senior teacher (level 7.1); T—teacher with qualification; A—assessor; D—developer.

The participants were reached through the Estonian Union of Teachers. They were the ones who were either in the process of applying for the qualification or the members of the assessment board. The sample size was small because since 2014 only 251 teachers out of 16,357 teachers, who work in schools in Estonia, have successfully obtained a PQ certificate (there are only about 160,000 general education school teachers in Estonia: about 14,000 women and 2000 men [54]). These figures are so small because of our small (teacher) population. Due to the small teacher population in Estonia, it was not possible to recruit more people who had recently been involved in the application process and had volunteered to participate in the study. The participants were first contacted via e-mail, and they were informed about the aim of the research and the background of the interviewer and asked for their consent to be interviewed. The individual interviews were conducted between May 2019 and July 2019.

2.2. Data Collection

For collecting data, individual semi-structured interviews were used. The focus of the research was on the criteria of PS. The interviews were conducted face-to-face at a time and place appropriate for the interviewees. The interview with one of the developers of PS took place via Skype because meeting in person was not possible due to the long distances and busy schedule. The interview questions were formulated proceeding from the research questions. The semi-structured design of the interview enabled us to go deeper and investigate the interviewees' perceptions in more detail. The interview questions were different for different target groups according to the role the person had in the process of applying for qualification (Table A1): 9 questions for the applicants, 9 for the assessors and 8 for the developer of PS. The questionnaire was compiled to investigate the interviewees' perceptions of the shortcomings in the current PS and the procedure of applying for them, i.e., RQ 1 (questions 1–5). The second part of questions, i.e., RQ 2 (questions 6–9), was to find out the expectations and needs of teachers with regard to PS and the procedure of applying for them. In order to improve the quality of the research tool and test the interview questions, a pilot interview was conducted with one of the teachers from the sample group. It became evident that the interviewee understood all questions and the interviewer managed to gather necessary information. As the prepared questions did not need changing, the data received from the pilot interview were included and analysed together with the data from other interviews. At the beginning of each interview, the interviewer introduced themselves and the aim of the research. The interviewees were informed of the confidentiality of the process and were asked for permission to audio-record the interview. The shortest interview lasted 20 min and the longest 1 h and 15 min.

2.3. Data Analysis

Qualitative inductive content analysis was used to analyse the data gathered. In the course of the data processing, we looked for similar patterns appearing in the collected data. This method enabled us to elicit thorough information and enabled the interviewees to substantiate their arguments. The interviews were fully transcribed, 37 pages in total. The shortest interview was 4 A4 pages long and the longest 13 A4 pages. Pseudonyms

(Table 1) were used in the transcripts. The fragments of the recordings which could have revealed the identity of the interviewee were not transcribed.

For data analysis, QCAmap was used. This enabled us to code, store and organise the data. It also allowed the codes to be used repeatedly. The coding was based on RQ. In order to increase the reliability of the work, the coding was done repeatedly and the codes were checked by co-authors. The coding process provided 10 codes for the first research question divided into 7 subcategories which later were grouped into 5 main categories; 9 codes for the second research question resulted in 2 main categories. In the following section, the research results are presented with extracts from the transcription attached.

3. Results

The results of the data analysis are presented according to the research questions and described regarding sub- and main categories. The findings are illustrated with excerpts from the interviews, accompanied by the pseudonym of the interviewee.

Research Question 1 (RQ1): *What are the shortcomings perceived by the different target groups under the current criteria and the application process?*

The analysis of data revealed that there are shortcomings in five categories: perceptions of deficiencies while interpreting the PS, shortcomings in the first stage of application (e-portfolio and self-analysis), evidence-related shortcomings and shortcomings in the second stage of application (discussion/interview) and shortcomings in the third stage of application for PQ (Vocational Committee).

3.1. Perceptions of Deficiencies While Interpreting the PS

The interviews revealed a common understanding between assessors and applicants that teachers are not actually able to read PS and do not understand their content. This is where the description of the performance criteria of the PS is either incomprehensible or not sufficiently understood:

> "And when I first applied, I realized that I still couldn't read the document. Look, here's the tricky bit that teachers stumble on. They cannot understand what the document actually says..." (T3)

> "To be honest, it doesn't make it easy for a teacher—the language contains too much jargon" (A2)

All three target groups found the content of PS difficult to understand due to the complex wording (A, C) and the large number of repetitions (T) of some of the criteria's performance indicators.

In addition, all target groups unanimously acknowledged that at the moment of applying for the qualification, the teachers are not able to explain the criteria to themselves, which leads to a significant deficiency in self-analysis. Teachers may be competent, based on PS criteria, but they are not aware of their own competency.

As a shortcoming, all target groups also noted the fact that although applicants have the right to choose the level within the profession they want to apply for, they lack the skills and abilities to assess themselves, even when relying on the performance criteria, thus often choosing the wrong level to aim at. Thus, it is difficult for the applicants to understand whether their competencies are sufficiently demonstrated against the requirements of PS.

Both the evaluators and the developer pointed out that all applicants have the freedom to apply for teacher level 7, which lasts until the end of their careers. This can lead to the situation where teachers are contented with level 7 as the process of applying for a higher PQ is labour-intensive and complex and only lasts 5 years.

> "There have been cases in the application rounds where teachers who could actually apply for the profession of Master Teacher in their evidence and professional experience apply for Teacher Level 7 simply because they are not familiar

enough with the awarding system, they apply for Teacher Level 7 because it is open-ended"... (D)

3.2. Shortcomings in the First Stage of Application (e-Portfolio and Self-Analysis)

The process of applying for the qualification at stage 1 consists of creating an e-portfolio and self-analysis based on the performance indicators of PS. It is the task of the applicants to give a thorough overview of their competencies in the e-portfolio by including evidence-based documents. The assessors acquire the first impression of the competency of the applicants by reading the e-portfolio and finding evidence of the expected competencies for the performance indicators described in the PS.

As a shortcoming, the assessors and the developer jointly pointed out that it is certainly not possible to draw definitive conclusions about teachers' competencies from e-portfolios only because e-portfolios do not provide a complete overview of the applicants' competencies ("some teachers are not good writers, some teachers are able to write loads about nothing", A2).

In addition, reading and evaluating e-portfolios can be difficult for the assessor objectively because of the applicants' self-analysis, structuring, technical and self-expression skills:

"...e-portfolios are different, some are more scientific; others in essay form—which make it difficult to find the important points in them." (A1)

The need for the development of self-analysis skills emerged as a generic theme for all target groups:

"...so that applying teachers would be more capable of analysing their work and not rely on their gut feeling" (A3)

All target groups stressed that the ultimate goal should not be overlooked and that it is important for teachers to understand that a complete compliance of all competencies with PS is the ideal and almost impossible to achieve.

"Realistically, no one can achieve the maximum, it is not possible" (A1, A2, A3)

There is no systematic approach to what a teacher should do over a period of time (in-service training, etc.). Teachers are free to decide on their own development needs, and there are no direct demands from the state. As a result of that, teachers' development can be chaotic.

"We do not have an institution at the national level that sets out the directions or priorities we have in teaching or being a teacher that would bring all teachers to a particular level. The big downside to me is that there is no mechanism that works for all teachers to have at least some training in every 3 or 5 years, to become familiar with both SEN and digital..." (D)

There is also no national incentive system for teachers with higher qualifications in terms of salary, i.e., even if the teacher has level 8, their salary may stay the same. The salary increase depends primarily on the head of the school, their values and the budget of the school. Thus, there is no state institution in Estonia that would prioritise teachers with higher levels.

3.3. Evidence-Related Shortcomings

Evidence-based self-analysis is one of the requirements for applying for the qualification. The current system of applying for qualification is built on trust:

"It would be a trust-based system for teachers to regain confidence by analysing their activities and receiving support from other teachers, and feedback on the necessity and importance of their work." (D)

Both applicants and assessors pointed to the lack of the requirement for school evaluation, which can also be interpreted as a lack of interest and support from the school. Also highlighted as a shortcoming is the lack of a video lesson or the requirement for observing a

lesson—teachers want to demonstrate their commitment in the teaching process. However, the following concern arises: filming can become a distraction, and this type of proof may not be suitable for students with special needs in the classroom. Teachers' poor IT skills—editing, publishing, etc.—are also hindering factors in creating video material. The eSafety aspect (i.e., netiquette—follow the rules of how to behave safely on the Internet to protect yourself and your devices) and the General Data Protection Regulation (GDPR), which requires parental permission to film their children and publish the film, also play a major role.

There was also confusion among applicants and assessors about guidance by and cooperation with universities, which is very difficult in practice. Research guidance the universities provide high school teachers with offers an outlet and an opportunity to collaborate with them, but it is a much bigger challenge for primary school teachers as universities offer fewer projects that would involve primary school students and teachers.

There may also be some problems with the evidence-based self-analysis in the e-portfolio (not sufficient evidence or the evidence too general), but these are mostly compensated for in the next step of applying for the PQ.

3.4. Shortcomings in the Second Stage of Application (Discussion/Interview)

The second stage of the application process is the interview based on pre-completed documentation (self-analysis and e-portfolio) with the applicant. One shortcoming the applicants highlighted was the undefined structure of the interview, i.e., the applicants do not know what the particular VAC wants to hear and whether they prefer the scientific approach or the applicant's own views:

> "Later, I realized that I was stuck in too much detail, I should have seen the bigger picture. I should have brought out scientifically sound or my own views, but I was just stuck in details..." (T3)

As with applicants, the content of the conversation is confusing for assessors as well; assessors expect the interview to complement and support the applicants' self-analysis:

> "...this conversation should complement the applicant's self-analysis, that would be the ideal" (A2)

The developer of the PS highlighted the shortcomings in the joint discussions and training of the assessors, which again clearly indicates a lack of time and, to some extent, irresponsibility of the assessors. In addition, some of the shortcomings in the work of assessors can be due to the busy work schedules, namely, because they work as practising teachers.

> "Everyone makes their own notes (A1), but they should be collected and reviewed together. Unfortunately, this is not always possible due to lack of time" (T3)

Hence there is a need for the assessors to find more time to cooperate and to write the unified opinion of the applicant's progress before sending off the feedback document and asking additional questions concerning the applicants' under-performance.

As a major shortcoming, the developer highlighted the quality of feedback provided by the assessors, which should support applicants' self-esteem and self-development needs, highlight the best performance and motivate applicants to develop further...

> "It would be necessary to highlight those strengths and the specific competences that need to be developed from the point of view of the applicants so that they not only receive the analysis but also receive a meaningful assessment." (D)

High-quality feedback should include an assessment of the teacher's performance according to PS and recommendations for setting goals for further development

3.5. Shortcomings in the Third Stage of Application for Professional Qualification (Vocational Committee)

The third stage of the assessment process is carried out by a Vocational Committee (VC) consisting of 10 people from different educational backgrounds. The committee shall approve or reject the decisions of the VAC. The work of the VC is based on the feedback and suggestions of the VAC, i.e., the VC trusts and accepts the decisions of the VAC. As a shortcoming, the developer once again pointed out the limited time resource for commission meetings.

Research Question 2 (RQ2): *What are the expectations and needs of different target groups for professional standards and professional application procedures?*

The analysis of the data revealed two main categories: expectations and needs for the first stage of the application process and expectations and needs for the second stage of the application process.

The application procedure for professional qualification consists of three interrelated steps: e-portfolio and self-analysis; professional examination—used in the context of the award of a profession in the sense of the assessment process and not in the traditional sense of the word "examination" [50], i.e., interview; and decision by Vocational Committee on awarding the qualification.

3.6. Expectations and Needs for the First Stage of the Application Process

All three target groups found that creating an e-portfolio is a long, time-consuming and complex process that requires the teachers to stretch themselves. Teachers are expected to determine the structure of their e-portfolio and compile their self-analysis based on PS and their work experience:

> "The teacher's own experience is what is most important, theory is to confirm this. Not that you have a lot of theory and two sentences about yourself." (A3)

The common understanding of all target groups was that teachers' self-analysis skills need to be developed. The target groups expected that a video lesson or lesson observation could be used in the qualification application process for assessors to familiarise themselves with the teacher's work and for the teacher to improve their self-analysis skills. A video lesson or lesson observation provides an insight into the teacher's work and is evidence-based.

> "... video lesson should be recorded or a member of the committee would come to watch your lesson or talk to you beforehand. I imagine that it would help a lot in terms of self-analysis." (T1)

All target groups expected that the teacher could upload the recorded video lesson in their e-portfolios in clips, followed by the analysis of the successful and unsuccessful parts of their work:

> "It doesn't have to be a 45-min video, it could be one chosen clip I might like and another where I feel like I've failed as a teacher. It could be inside in the e-portfolio because it makes you analyse yourself. Otherwise, this self-analysis would remain very superficial." (T3)

All target groups found that teaching is not only the individual work of the teacher but collaboration with colleagues and management, and thus it is important that the head of the school assesses the teacher.

> "...as the teacher's professionalism does not only come from teaching, it is also thanks to cooperation with..." (A1)

In conclusion, it can be said that the purpose of all the above-mentioned activities is to provide an objective overview of a teacher applying for a profession, that is, an overview of their professionalism. At the same time, they are expected to supplement this overview with evidence-based videos, pictures and various references (from the employer, community of practice, etc.).

3.7. Expectations and Needs for the Second Stage of the Application Process

The most important thing in an interview is the direct communication with the applicant. Before the interview, the assessors examine the applicant's written self-analysis, but the interview is the part of the process that ensures direct contact. Assessors expect the interview to reveal the applicant's personality traits and provide a more thorough overview of the applicant's professionalism and competencies:

> "...Face-to-face with the applicant to assure that everything in their digital portfolio and everything they have provided in answer to our supplementary questions is very good; that they are doing a good and necessary job and they do not have to underestimate themselves, even if they may not have the school feedback on their work, which can occasionally happen..." (A2)

The interview gives the assessors the opportunity to specify and talk about the shortcomings in their self-analysis and confirms the manifestation of competencies in their reflection:

> "... the need to indicate those competencies that are scarce or would like to be questioned and clarified..." (A3)

All target groups expected the interview to become a discussion between colleagues, which is a supportive system for applicants but also supports the PD of assessors. If some competencies are questionable, additional pre-interview questions would help to clarify the applicant's competencies.

Both assessors and developers believed that it is the task of the VAC to conduct the interview so that teachers feel that the committee has worked through the documents they have submitted and that the committee emphasizes the more outstanding aspects of the candidates and recommends the development of less evident competencies.

> "So that this were not just a statement of what this teacher is doing, but this conversation would really highlight the strengths of the teachers and the points that need improvement. So, it would not be just another self-analysis but also an assessment of the activities they really excel at." (D)

Currently, the procedure of conducting interviews differs from committee to committee:

> "We have some basic questions we ask during the interview. Training sessions have also been organized on the topics of the interview. There could be more training, as the committees now have very different working arrangements. The basic issues of the conversations could also be standardized..." (A2)

The applicants emphasized that, in order to avoid confusion, more detailed information was needed on the content of the expected interview and on what the committee expected of them during the interview. Consequently, harmonization of the content of the interview would be necessary for both applicants and assessors (the common basis for the interview). Thus, all target groups expected that applying for a profession is a process in which the applicants receive constructive and supportive feedback.

4. Discussion

The main value of the work is identifying the assessments, needs and expectations given by the teachers to the qualification application process in order to make it more understandable to them. Previously identified weaknesses prevented teachers from understanding the content and performance of PS. From the results of this study, it is possible to pay attention to the shortcomings that emerged based on the expectations and needs expressed by the participants in the process of applying for the qualification. The practical value is getting an overview of teachers' perceptions of PS as a tool for assessing the professionalism of teachers when applying for a PQ by analysing the results of the research and finding an explanation why the PS of teachers are not being sufficiently used in Estonia. The following are the main shortcomings, needs and expectations relevant to the application process and PS (Table 2):

Table 2. Findings per target groups.

Target Groups	Sub-Category	Shortcomings	Expectations and Needs
All target groups	Interpretation of the PS	A lack of unambiguous interpretation of PS	Improve the wording of the PS performance indicators
All target groups	Self-analysis	Difficulties choosing the level of professional qualification	
Assessors and teachers	Evidence-based	The evidence-based nature of performance indicators is difficult to prove	Add a video lesson to self-analyse (analyse successes and failures)
Assessors and developer of PS	Evidence-based	The e-portfolio does not provide a complete overview of competencies	The interview reveals the applicant's personality traits and provides a more thorough overview of the applicant's professionalism and competency
Assessors and developer of PS	Self-analysis	Weak self-analysis skills	Increase the scope of self-analysis (by supporting the development of various competencies)
Applicants and assessors	Evidence-based	A lack of recognition of the teacher's work by the school management	Put into use a letter of recommendation from the school management
Applicants and assessors	Interview	Applicants do not know what is expected of them in the interview	Standardize the content of the interview
Assessors	Interview	There are no uniform requirements for the structure and conduct of the interview	Need for further training in this area. Harmonize the main points and structure of the interview
Developer of PS	Interview	The quality of the feedback from the professional assessors to the applicant	The quality of the feedback provided, which should support applicants' self-esteem and self-development needs, highlight the best results and motivate applicants to develop further

The shortcomings identified by all three target groups were: difficulties in choosing the level of professional qualification, the lack of unambiguous interpretation of PS, the lack of evidence-based performance indicators (difficult to prove) and weak self-analysis skills. The interviews revealed the reasons for the weak self-analysis, which highlighted the existence of a short time limit by the target groups and the fact that the teaching process is taken for granted and is not documented. It was also stated that teachers cannot write a self-analysis (either they are poor writers or they have weak analysis skills). Teachers tend to perceive self-analysis as self-praise and not as an analysis of one's strengths and weaknesses. It was also revealed that teachers can point out their weaknesses but cannot help them. Assessors stated that the teachers are not very confident, rather they are too self-critical, which in turn indicates a lack of recognition of the teacher's work by the school management (pointed out by applicants). The e-portfolio, however, does not provide a complete overview of the teacher's competencies (PS are compiled based on the ideal teacher), but it is the material on which the applicant is first assessed and additionally questioned. An important part of the research was the expectations and needs of the three target groups for the PS and the application process. The main expectation of all target groups turned out to be the improvement of the wording of the PS performance indicators. It was also considered necessary to introduce the four different levels of PQ; bring in a video lesson as a tool for self-analysis and evidence-based validation, which also supports the development of digital competencies; and put into use a letter of reference from the school management, which means that the management is aware and supports the teacher's willingness to apply for the PQ.

PS are constantly being developed with the aim of making the application process better and more efficient. In the Estonian context, the interest of teachers in applying for a PQ has increased year by year, but PS can only be applied on a voluntary basis at the level of both teachers and schools, which is probably the reason for the lack of interest.

All target groups accepted the current application procedure. On the one hand, their expectations and needs for improvement increase the scope of self-analysis by supporting the development of various competencies. On the other hand, it makes it easier for assessors to assess applicants' capabilities more effectively.

5. Conclusions

This research was based on the opinions and suggestions of three different target groups (teachers who applied for the profession, professional assessors and the developer) regarding the process of applying for a PQ following the requirements of the PS currently valid in Estonia. The main value of the work is identifying the shortcomings, needs and expectations teachers have found while applying for PQ in order to make the process more understandable for them. Previously, the identified weaknesses prevented teachers from understanding the content and performance of PS. Based on the results of this study, it is possible to eliminate the shortcomings based on the expectations and needs expressed by the participants in the process of applying for qualification. Another valuable outcome is getting an overview of teachers' perceptions of PS as a tool for assessing the professionalism of teachers when applying for a PQ. A valuable result is also the consensus on the shortcomings and needs of teachers' professional standards and on proposals to improve the application process for the profession, which, in cooperation with the Teachers' Union, will allow for further improvements.

As the application process for the PQ in Estonia was started as early as in 2014, the re-application process is already underway, which is a new development and requires another review of the PS. Currently, teachers who are seeking higher qualifications are the more skilled active professionals who are open to continuous self-improvement and educational and technological innovation. Their expectations and needs for improvement increase the scope of self-analysis by supporting the development of various competencies. Furthermore, it enables the assessors to assess applicants' capabilities more effectively. In summary, it can be said that raising the qualification of a teacher and thereby defining

one's professional competencies is a free choice for a teacher, which largely depends on the teacher's internal motivation and self-improvement needs. Although this may vary from country to country, improving the qualifications of teachers is a national policy interest everywhere, assuring the professionalism of working teachers and the quality of education in the long term.

This study does not reflect differences in the application process or the objectives of using professional standards across countries. However, it provides insight and benchmarks on the application process and the functioning of professional standards for teachers in different countries. The study showed that teachers need collaborative support, motivation and constructive feedback on their professional development, both from colleagues and at school and national levels. Applying for a PQ is based on an overview, analysis and evidence-based professional activity of a teacher, which forms an important part of a teacher's PD. One of the reasons for the low level of teachers' application for PQ can be the insufficient professional recognition of teachers, which would certainly motivate teachers to apply for a higher PQ level. Currently, teachers who are seeking higher qualifications are the more skilled active professionals who are open to continuous self-improvement and educational and technological innovation. Teachers who have gone through the application process have gained greater self-confidence, value themselves as teachers and continue their in-service training. Therefore, the ones who passed the PS interviews are qualified teachers and are specialists in their field who are able to ensure quality teaching based on their personality.

Further research will focus on creating a learning analytical model which would help teachers in the process of applying for PQ.

Limitations and Implications for Further Research

This work focused on the target group whose members have already applied for the profession and have a higher level of PQ. One of the limitations of the study is that it was based on only working teachers who were upgrading their qualification and excluded those who were still learning to be teachers. The study considered only the views of those who had applied for the PS qualification. We acknowledged that this is a limitation of our sampling strategy. Future research should also include active teachers who have not applied for a qualification as well as those training to be teachers to understand the barriers or hesitancy in applying for the PS qualification. Another limitation is the extremely small sample consisting of only three teachers and three assessors. The findings of the study can therefore only be a snapshot view of these individuals and not necessarily the views of the majority. Nevertheless, their views highlight the challenges involved in the application process. In this case, it would be possible to investigate more precisely, on the one hand, the needs and expectations of teachers regarding the application process and requirements of the profession and, on the other hand, the direction in which the standards of the teaching profession are changing with the development of society. We also highlight the need for involving novice teachers, hoping to expand and repeat our study in the future.

There were cases where teachers had not been awarded the qualification they were applying for (but only professionals applying for the highest qualification were involved).

Author Contributions: Author Contributions: Conceptualization, T.L. and K.S.; methodology, T.L., K.S. and I.-A.C.; valida-tion, T.L., K.S. and I.-A.C.; formal analysis, T.L., K.S. and I.-A.C.; investigation, T.L., K.S. and I.-A.C.; data curation, T.L., K.S. and I.-A.C.; writing—original draft preparation, T.L.; writing—review and editing, T.L., K.S. and I.-A.C.; visualization, T.L.; supervision, K.S. and I.-A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as it involved only adults who all gave informed consent for participation. No sensitive personal data were used in the study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Interview plan.

Research Questions	Interview Questions		
	Target Groups		
	Applicant for the Qualification	Professional Assessor	Developer of Professional Standards
1. How do target groups perceive shortcomings in current PS and professional application procedures?	1. Prior information on applying for a profession? 2. Overview of evaluation criteria (excessive, missing)? 3. Disagreements during the evaluation interview? 4. Process evaluation, suggestions? 5. Satisfaction/non-satisfaction with the invitation application process?	1. Teacher's ability to choose a professional standard? 2. The most difficult criteria for a teacher's self-assessment? Reason? 3. Is an overview of the teacher level sufficient? 4. Shortcomings in analysing e-portfolios? The solution? 5. Shortcomings in the conversation? The solution? Disagreements in the assessment? Solutions?	1. Is a self-assessment overview of the teacher's level sufficient? 2. How does the (initial evaluations) work? 3. Shortcomings in the 's work? 4. Shortcomings in the work of the professional committee? 5. How does the work? Initial assessments of documents? 6. Disagreements in the evaluation (have there been any reasons for reaching solutions?)
2. What are the expectations and needs of the target groups for the current professional standards and the procedure for applying for the qualification	6. Documentation and overlap of conversation goals? 7. Proposals for professional standards and the professional application process 8. The purpose of the interview, was it based on evaluation criteria and documentation? 9. Assessment of the invitation application process	7. Overlap of documentation and conversation goals? 8. The purpose of the conversation/interview? 9. Assessment of professional standards and the application process?	7. Prerequisites for applying as a teacher? 8. How and on what documentation does the Vocational Committee work (final stage)

References

1. Gewirtz, S.; Mahony, P. *Changing Teacher Professionalism: International Trends, Challenges and Ways Forward*, 1st ed.; Hextall, I., Cribb, A., Eds.; Taylor Francis Group: Abingdon, UK; Routledge: London, UK, 2007; pp. 19–30.
2. Hargreaves, A. Four Ages of Professionalism and Professional Learning. *Teach. Teach.* **2000**, *6*, 151–182. [CrossRef]
3. Hargreaves, A.; Goodson, I. Teacher's Professional Lives: Aspirations and Actualities. In *Teacher's Professional Lives*; Goodson, F., Hargreaves, A., Eds.; Falmer Press: London, UK, 1996; pp. 1–27.
4. Whitty, G. Changing modes of teacher professionalism: Traditional, managerial, collaborative and democratic. In *Exploring Professionalism*; Cunningham, B., Ed.; Institute of Education, University of London: London, UK, 2008; pp. 28–49.
5. European Council. Available online: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:302:0006:0009:EN:PDF (accessed on 10 November 2020).
6. Hargreaves, A. Four ages of professionalism and professional learning. In *Education, Globalization, and Social Change*; Lauder, H., Brown, B., Dillabough, J.-A., Halsey, H., Eds.; Oxford University Press: Oxford, UK; New York, NY, USA, 2006; pp. 672–691.
7. Hoyle, E. Professionality, professionalism and control in teaching. *Lond. Educ. Rev.* **1974**, *3*, 13–19.
8. Pratte, R.; Rury, J.L. Teachers, Professionalism, and Craft. *Teach. Coll. Rec.* **1991**, *93*, 59–72.
9. Eraut, M. *Developing Professional Knowledge and Competence*; Falmer Press: London, UK, 1994.
10. Evans, L. Professionalism, professionality and the development of education professionals. *Br. J. Educ. Stud.* **2008**, *56*, 20–38. [CrossRef]
11. Sachs, J. Professional standards: Quality teachers for the future. In Proceedings of the Sharing Experience: Ways forward on Standards Conference, Melbourne, Australia, 21–22 August 2005.

12. Holroyd, C. Are assessors professional? In *Active Learning in Higher Education*, 1st ed.; Open University Press: Buckingham, UK, 2000; Volume 1, pp. 13–19.
13. Theories on and Concepts of Professionalism of Teachers and Their Consequences for the Curriculum in Teacher Education. Available online: http://theories-on-and-concepts-of-professionalism-hugarian-publication.pdf (accessed on 12 November 2020).
14. Snoek, M.; Swennen, A.; Van der Klink, M. The teacher educator: A neglected factor in the contemporary debate on teacher education. In Proceedings of the 3rd Annual Conference on Teacher Education Policy in Europe (TEPE), Umea, Sweden, 18–20 May 2009. Available online: https://www.amsterdamuas.com/binaries/content/assets/subsites/kc-oo/publicaties/the-teacher-educator-a-neglected-factor-in-the-contemporary-debate-on-teacher-education.pdf (accessed on 10 October 2020).
15. Hoyle, E. Changing Conceptions of Teaching as a Profession: Personal Reflections. In *Teaching: Professionalization, Development and Leadership*; Johnson, D., Maclean, R., Eds.; Springer: Dordrecht, The Netherlands, 2008; pp. 285–304.
16. Kleinhenz, E.; Ingvarson, I. Standards for Teaching: Theoretical Underpinnings and Application. 2007. Available online: https://www.researchgate.net/publication/43328197_Standards_for_Teaching_Theoretical_Underpinnings_and_Applications (accessed on 15 September 2020).
17. OECD Library Teachers Matter: Attracting, Developing and Retaining Effective Teachers. Available online: https://www.oecd-ilibrary.org/education/teachers-matter-attracting-developing-and-retaining-effective-teachers_9789264018044-en (accessed on 8 November 2020).
18. Guerriero, S.; Toledo-Figueroa, D.; Révai, D. Teacher professionalism and knowledge in qualifications frameworks and professional standards. In *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*; Guerriero, S., Ed.; OECD Publishing: Paris, France, 2017; pp. 73–91.
19. Darling-Hammond, L. Reshaping Teaching Policy, Preparation, and Practice. In *Influences of the National Board for Professional Teaching Standards*; AACTE Publications: Washington, DC, USA, 1999; pp. 5–39.
20. Hattie, J. Teachers Make a Difference, What Is the Research Evidence? Available online: https://www.researchgate.net/publication/252090858 (accessed on 12 October 2020).
21. Bourke, T.; Rayan, M.; Ould, P. How do teacher educators use professional standards in their practice? *Teach. Teach. Educ.* **2018**, *75*, 83–92. [CrossRef]
22. Standards Frameworks for Teachers and School Leaders. Available online: https://www.academia.edu/8238247/Standards_Framework_for_Teachers_and_School_Leaders (accessed on 10 November 2020).
23. Ingvarson, L.; Kleinhenz, E. A Review of Standards of Practice for Beginning Teaching. 2003. Available online: https://www.researchgate.net/publication/43328189_A_review_of_standards_of_practice_for_beginning_teaching (accessed on 10 November 2020).
24. Hayes, T. Professional teaching associations and professional standards: Embedding standards in the discourse of the profession. *Teach. Aust.* **2006**. Available online: https://apo.org.au/node/1897 (accessed on 10 September 2020).
25. Pedaste, M.; Leijen, Ä.; Poom-Valickis, K.; Eisenschmidt, E. Teacher professional standards to support teacher quality and learning in Estonia. *Eur. J. Educ.* **2019**, *54*, 389–399. [CrossRef]
26. Mayer, D.; Mitchell, J.; Macdonald, D.; Bell, R. Abbreviated Professional standards for teachers: A case study of professional learning. *Asia Pac. J. Teach. Educ.* **2005**, *33*, 159–179. [CrossRef]
27. Mars, E. Professional standards, teacher careers and the conference enactment of professional judgement: An exploration of the discourses. In *AARE 2012 Conference Proceedings, Proceedings of the Joint Australian Association for Research in Education and Asia-Pacific Educational Research Association Conference, Sydney, Australia, 2–6 December 2012*; Australian Association for Research in Education (AARE): Canberra, Australia, 2012.
28. Leibur, T.; Saks, K.; Chounta, A.E. Teachers' and Assessors' Perceptions of the Application Process of Professional Qualification. In Proceedings of the INTED 2020, Valencia, Spain, 2–4 March 2020.
29. Estonian Qualifications Framework (EstQR). Available online: https://www.kutsekoda.ee/en/estonian-qualifications-framework-estqf/ (accessed on 1 October 2020).
30. Ingvarson, L. *Development of a National Standards Framework for the Teaching Profession. ACER Policy Briefs. Issue 1*; ACER Press: Camberwell, Australia, 2002; pp. 3–32.
31. Professional Standards for Teachers and Teacher Education: Avoiding the Pitfalls. Available online: https://img.scoop.co.nz/media/pdfs/0604/avoiding_pitfalls.pdf (accessed on 12 October 2020).
32. Cumming, J. *The Push for Standards in Teaching Profession: Stakeholders, Strategies and Scenarious*; Australian College of Educator: Camberwell, Australia, 2010; pp. 1–14.
33. Bowker, G.; Star, S.L. *Sorting Things out: Classification and Its Consequences*, 2nd ed.; The MIT Press, Massachusetts Institute of Technology: Cambridge, MA, USA, 1999; pp. 1–34.
34. Koster, B.; Dengerink, J.J. Professional standards for teacher educators: How to deal with complexity, ownership and function. Experiences from the Netherlands. *Eur. J. Teach. Educ.* **2008**, *31*, 135–149. [CrossRef]
35. Sykes, G.; Plastrik, P. *Standard Setting as Educational Reform*; American Association of Colleges for Teachers of Education: Washington, DC, USA, 1993; pp. 15–45.
36. Ingvarson, L.C.; Hattie, J. Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards. In *Advances in Program Evaluation*; Ingvarson, L.C., Hattie, J., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2008; Volume 11, pp. 55–93.

37. Kriewaldt, J. Research into relationships between teacher professional learning and teaching standards. In *Proceedings of Conference of the Australian Association for Research in Education, Proceedings of the AARE 2008 International Educational Education Research Conference, Brisbane, Australia, 30 November–4 December 2008*; Jeffrey, P.L., Ed.; Papers collection; Association for Research in Education: Melbourne, Australian, 2008.
38. Ball, S. Performativities and fabrications in the education economy: Towards the performative Society? *Aust. Educ. Res.* **2000**, *27*, 1–23. [CrossRef]
39. Sachs, J. Teacher professional standards: Controlling or developing teaching? *Teach. Teach.* **2003**, *9*, 175–186. [CrossRef]
40. Ingvarson, L.C. Recognising accomplished teachers in Australia: Where have we been? Where are we heading? *Aust. J. Educ.* **2010**, *54*, 46–71. [CrossRef]
41. Sadler, D.R. Interpretations of criteria-based assessment and grading in higher education. *Assess. Eval. High. Educ.* **2005**, *30*, 175–194. [CrossRef]
42. Snoek, M. The teaching profession in the Netherlands: From regulative structures to collaborative cultures. In *The Dutch Way in Education: Teach & Lead the Dutch Way*; Heijmans, J., Christians, J., Eds.; Onderwijs Maak Je Samen: Helmond, The Netherlands, 2017; Volume 3, pp. 91–107.
43. Standards for Graduation and Initial Teacher Certification: The International Experience. Available online: https://works.bepress.com/lawrence_ingvarson1/206/ (accessed on 27 September 2020).
44. Guerriero, S.; Toledo-Figueroa, D.; Révai, D. Teacher' Pedagogical Knowledge: What it is and how it functions. In *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*; Guerriero, S., Ed.; OECD Publishing: Paris, France, 2017; pp. 99–115.
45. Caena, F. Teacher Competence Frameworks in Europe: Policy-as-discourse and policy-as-practice. In *The Changing Role of Teachers*; Wiley Online Library: Hoboken, NJ, USA, 2014; Volume 49, pp. 311–331.
46. Teaching Careers in Europe: Access, Progression and Support. Available online: https://op.europa.eu/en/publication-detail/-/publication/9ef5c37e-1b75-11e8-ac73-01aa75ed71a1/language-en/format-PDF/source-search (accessed on 23 October 2020).
47. Snoek, M.; Dengerink, J.; De Wit, B. Reframing the teacher profession as a dynamic multifaceted profession: A wider perspective on teacher quality and teacher competence frameworks. *Abbreviated Eur. J. Educ.* **2019**, *54*, 413–425. [CrossRef]
48. Gerstner, L.; Ackerman, A.; Barnes, R.; Beattie, R.; Chenault, K.; Condit, D. *Teaching at Risk: Progress and Potholes*; The Teaching Comission: New York, NY, USA, 2006.
49. Estonian Qualifications Authority. Available online: https://www.kutsekoda.ee/en/ (accessed on 1 October 2020).
50. Kutsestandardi Rakendamine Õpetajaks Kujunemisel ja Edasises Professionaalses Arengus. Available online: https://lytorn.files.wordpress.com/2016/10/kutsestandardi-rakendamine-ja-eneseanalc3bcc3bcs.pdf (accessed on 15 March 2020).
51. Robson, C. *Real World Research*, 2rd ed.; Blackwell Publishing: Malden, MA, USA, 2002.
52. Van Manen, M. *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*; State University of New York Press: Albany, NY, USA, 1990.
53. Turner, D.W. Qualitative interview design: A guide for novice investigators. *Qual. Rep.* **2010**, *15*, 754–760.
54. Stationary General Education. Available online: http://andmebaas.stat.ee/Index.aspx?lang=en&SubSessionId=48144bb2-6a50-4d7d-8fec-be8871a3e3ab&themetreeid=6 (accessed on 19 July 2021).

education sciences

MDPI

Review

Remote and Blended Teacher Education: A Rapid Review

Thomas Perry [1,*], Madeleine Findon [1] and Philippa Cordingley [2]

[1] Department of Teacher Education, University of Birmingham, Birmingham B15 2TT, UK; m.a.findon@bham.ac.uk
[2] Centre for the Use of Research & Evidence in Education (CUREE), Penrith CA10 3HN, UK; philippa.cordingley@curee.co.uk
* Correspondence: t.w.perry@bham.ac.uk

Abstract: Initial and continuing teacher education are increasingly making use of remote and blended modes of education. Conducted in the summer of 2020 during the COVID-19 pandemic, this rapid review brings together literature and evidence to inform planning for remote and blended teacher education during restrictions in face-to-face teaching activity. The review consists of three main parts: first, a descriptive framework of modes of remote and blended teacher education; second, an exploratory review of the affordances and limitations of remote and blended approaches connecting the literature on effective teacher education with reviews of remote and blended approaches; third, a rapid review of evidence on the efficacy of remote and blended approaches, including of a small number of studies comparing these to face-to-face equivalents. We conclude that remote and blended teacher education is likely to become an increasingly important part of the teacher education landscape and there are plausible theoretical reasons suggesting that it can be effective with suitable design. However, we find too few studies presenting robust evidence to enable firm conclusions to be drawn on the relative effectiveness of modes and approaches. The review provides a foundation for further research and practice in this area.

Keywords: teacher education; professional development; remote learning; rapid review

Citation: Perry, T.; Findon, M.; Cordingley, P. Remote and Blended Teacher Education: A Rapid Review. *Educ. Sci.* **2021**, *11*, 453. https://doi.org/10.3390/educsci11080453

Academic Editors: Beng Huat See and Rebecca Morris

Received: 30 April 2021
Accepted: 2 August 2021
Published: 23 August 2021

1. Introduction

1.1. Remote and Blended Teacher Education

It is widely acknowledged that teacher quality is the largest single factor affecting student achievement that is amenable to school and policy influence (Slater et al., 2012; Kane et al., 2013) [1,2]. Teacher education, including initial education and continuing professional development (CPD), is therefore arguably the most direct, effective, and cost-effective approach to school improvement (Fletcher-Wood and Zuccollo, 2020) [3]. Here, we define initial teacher education as any training provided to pre-service teachers as part of qualification; we define continuing professional development as any support or training for in-service teachers. In both cases our focus is on efforts to improve teacher effectiveness. We know it is possible to improve teacher effectiveness (Kraft and Papay, 2014; Hill et al., 2020) [4,5], and that there are also wider benefits of doing so including greater collective teacher efficacy (Tschannen-Moran and Hoy, 2001) [6], wellbeing, retention and recruitment, in particular for early-career teachers (Fletcher-Wood and Zuccollo, 2020; Cordingley et al., 2019) [3,7]. The literatures on effective initial teacher education and CPD have not to date been well connected to those on blended and online teacher education; a contribution of this review is to make this connection, bringing together theory, evidence and description of practice for remote and blended teacher education (see Section 1.2, below for further details).

There is ongoing debate about the definitions of 'remote', 'distance' and 'blended' modes of education (Hobbs and Bolan, 2021) [8]. Accordingly, another contribution of this review is the development of a two-part, mode–characteristics framework which

can be used to describe forms of teacher education other than face-to-face provision. Prior to presenting this, we proceed on the basis of simple, broad definitions: we define remote teacher education as any mode other than face-to-face education, typically but not necessarily conducted online (i.e., using computers, tablet devices, smart phones and other web-based technologies). We define blended teacher education as a combination of face-to-face and remote education in any proportion. As we discuss at length in the substantive sections of this review, there is a large amount of variation across and within approaches to teacher education which fall under these broad definitions.

This review is timely. In recent decades, as technology has developed, organizations providing initial and continuing professional development for classroom teachers have increasingly developed blended learning offers, combining distance with face-to-face provision of professional learning content and activities. This slowly increasing trend towards distance and blended forms of teacher education has accelerated rapidly in the context of COVID-19, leading many providers to review their offer in terms of which modes are available for components and content areas in their programs and provision (La Velle, Newman, Montgomery, and Hyatt, 2020) [9]. In our own experience of initial teacher education provision, for instance, teacher educators have had to rapidly prepare and deliver online provision to enable students to continue with their education, and to enable finalists to be recommended for Qualified Teacher Status. Online provision has included video and voice-over to supplement PowerPoint presentations, video-capture lesson recordings, discussion boards, and virtual learning environment chat rooms to ensure that students' knowledge and skills continue to develop while face-to-face contact was not possible.

There is a long history of using blended learning for teacher education, in particular in developing and developed countries (e.g., see Burns, 2011, discussing initial teacher education) [10] and in locations where face-to-face provision presents challenges. In such cases, wireless education for teachers is a venerable forerunner of remote and blended CPD. Australia, for example, has a particularly long history of trying to meet CPD and teacher education needs remotely because of its topography. Dyment and Downing (2020, p.316) [11] report that 'between 2005 and 2016, the number of initial teacher education (ITE) students choosing an online or blended mode of study in Australia nearly doubled from 5412 (22% of total) in 2005 to 12,143 (41% of total) in 2015 (Australian Institute for Teaching and School Leadership (AITSL), 2015) [12]. Of the 47 providers of ITE in 2015 in Australia, 19 (40%) offered some or all of their courses in an online or blended mode of study.' This increase in remote and blended provision is also reflected in the academic literature, with Dyment and Downing being struck by the exponential increase in material when preparing their review in 2019, as compared to a 2013 review. Similarly with the literature on remote and blended continuing professional development: back in 2009, Dede et al. (2009) described much of the literature in the field of online teacher professional development as 'anecdotal', urging researchers to work towards an 'evidence-based conceptual framework that provides robust explanatory power for theory and model building' [13] (p. 8). Five years later, Fishman et al. (2014a, p. 261) [14], were agreeing with Moon et al. that "the field would benefit from scholarship that leads to 'research-based design principles to guide the ongoing development, implementation, and evaluation efforts in online PD (Moon et al., 2014, p. 1) [15]'". Now, another 6 years (and a global pandemic) later, this rapid review takes another look at the approaches, affordances and evidence on remote and blended initial and continuing teacher education. At present there is not an established systematic review focused specifically on effectiveness of or within modes of teacher education. Our initial scoping suggested that the evidence base in this area would be disparate, and there would likely be challenges applying it to a COVID-19 or post-COVID-19 scenario. With few common definitions and considerable variation in practices, we also anticipated challenges delineating between modalities that blend face-to-face and online modes. Nonetheless, we set out to identify relevant studies with the potential to provide practical theory and an evidence base for effective remote and blended teacher education practice. We also

revisited existing teacher education reviews and meta-analyses with a specific focus on identifying evidence and implications relevant to teacher education modality.

1.2. Review Aims, Questions and Contribution

This review was conducted during the summer 2020. It had the specific immediate aim of informing the research and commissioning organizations' teacher education provision for the new academic term beginning in September. The time available of around 4–6 weeks, the commitment of both organizations to evidence-informed practice, and the value of identifying principles for action all necessitated a broad range of literature being brought together in a very limited time frame. As such, we describe the study as a rapid review. As discussed in the methods section, in part this consisted of a simplified, systematic review prioritizing the identification of robust trials of remote and blended teacher education trials (reported in Section 5). Given our broader, practical aims, we also located reviews and empirical studies that describe remote and blended teacher education and/or reported principles for their effective design. While our treatment of these employed elements of systematic review approaches to minimize bias and increase transparency, the time and resource constraints did not allow a wholly systematic treatment. Accordingly, we have described these sections (Sections 3 and 4) as 'exploratory'; our aim with these sections is to bring together, summarize and interpret a large range of literature on CPD, teacher education and remote and blended modes of teacher education into a single account that addressed the immediate aim of informing current provision as well as providing a foundation for further research in this area.

In overview, this rapid review report consists of the following four sections, each with a specific aim and set of questions. We describe the methods employed within these four sections in Section 2 below.

Section 3—Remote and Blended Teacher Education Modes and Their Characteristics

Approaches—What forms of/approaches to remote or blended teacher education have been explored in the research literature? What are their typical characteristics/elements? How can this be organized?

In this section, we summarised reviews and empirical pieces in our data base that described specific modes and approaches to remote and blended teacher education, either individually or with a view to providing an organizing conceptual framework. We synthesize descriptions and frameworks to present our own two-part typology (modes–characteristics) framework, which we proceed to provide illustrative examples for.

Section 4—Affordances and Limitations of Remote and Blended Approaches to Teacher Education

Affordances and Limitations—What are the affordances and limitations of remote and blended teacher education relative to face-to-face variants? What adaptations, restrictions or enhancements are typical relative to face-to-face variants?

a. What, from a theoretical standpoint, can be inferred from the literature on effective continuing professional development and learning (CPDL) and initial teacher education (ITE)?
b. What is advocated as best practice in the remote/blended teacher education literature?

This section examines the theoretical and practical principles behind effective remote and blended teacher education. It does this by organizing, summarizing and analyzing reviews and empirical studies in our database against a set of principles for effective CPD and ITE. We present these principles for effective CPD and ITE in Sections 1.3 and 1.4, respectively, below. By connecting the literature on effective remote and blended provision with a framework for effective CPD and ITE, we are able to present an exploratory but evidence-informed account of the affordances and limitations of conducting teacher education through remote modes. This provides theoretical support for the use of remote modes, identifies adaptations recommended in the literature and in some cases identifies advantages of remote education over face-to-face alternatives.

Section 5—Review of Empirical Evidence

Evidence—What empirical evidence is available that supports, refutes and/or refines our understanding of effective remote or blended teacher education (as per Q2)?

a. Which forms or practices are identified as more/less effective?
b. What are the common features of effective approaches?

In Section 5, we present the evidence from trials of remote and blended teacher education that we have systematically reviewed. This is a subset of the overall review database which consists of studies employing appropriate research designs for causal estimates of the impact of remote or blended teacher education interventions on *pupil* outcomes. We also report a very small subset of these studies that compare the effect of remote or blended teacher education interventions against comparable, face-to-face alternatives. Given the limitation in this evidence base, our findings are restricted to a more general judgements of the effectiveness of remote and blended teacher education. We have presented our original, more granular, questions above for purposes of transparency in relation to our original aims, despite the limitations of the final evidence base.

Section 6—Conclusions

Implications and Unknowns—What, if any, general principles for effective online or blended teacher education are evident?

a. What does the evidence suggest about the relative effectiveness of teacher education modalities (including in comparison with the face-to-face mode)?
b. What are the main areas of uncertainty and gaps in our present understanding?

We have organized the overall paper to provide discussions and summaries of findings in each of the substantive sections above. Section 6, therefore, provides a short, more general conclusion and reflections on the present state-of-the-art in the research and practice of remote and blended teacher education.

1.3. Effective Continuing Professional Development and Learning

Whether teacher education is supported face-to-face, online or through a blend of both modes of activity, it is important for us start with some clarity about the key elements of success likely to applicable to all modes. We therefore consider in this section the principles for effective CPD and then additional considerations and differences for initial teacher education. As we describe above, this provides a conceptual framework for Section 4, where we consider the affordances and limitations of remote and blended teacher education. This connection between the general CPD literature and the remote and blended teacher education literature reflects our assumption that understanding effective remote and online teacher learning requires one to apply what is already known about effective teacher education to a new space, and while utilizing new technologies (Quinn et al., 2019) [16]. In other words, we hold that should not be confusing the medium (and the structure) for the message: Whether online, face-to-face or a blend of these, teachers must go through a familiar set of thinking and practice development processes. Insofar as there are separate principles for effective blended and remote teacher learning, they will relate to how we can employ new approaches, mitigate their limitations, and realize their affordances in recognition of the more general set of teacher learning principles.

In Table 1, below, we provide an overview of principles for effective CPD, continuing professional development learning (CPDL) leadership and then initial teacher education. CPDL is a broader concept than CPD which includes the learning (L) from CPD and ongoing professional learning connected, or in addition, to formal training and development (CPD) activity (Cordingley, 2015) [17]. For leadership of CPDL, we include all staff within educational organizations who have a role or influence on professional development; typically those in senior leadership positions and/or with formal responsibilities relating to professional development. Key reviews sitting behind these principles include those conducted by Cordingley et al. (2015, 2018, and 2020) [18–20], led by the Centre for the Use

of Research and Evidence in Education (CUREE). We have also consulted several other reviews of effective CPD, all with their own foci in areas of teacher education, locations and review methodologies. All reviews of professional development consulted are included in the dedicated section of the review references. Of particular note in terms of both quality and recency are the Darling-Hammond (2017) [21] review, assessing studies employing experimental or quasi-experimental designs for teacher development; the Rogers, Brown and Poblete (2020) [22] systematic review focusing on professional learning in early years of education; and the meta-analysis from Hill et al. (2020) [5] which briefly summarizes effective professional learning in relation to STEM-focused programmes.

Table 1. Overview of Principles for Effective CPD and its Leadership.

Summary of Research on Key Elements of Effective CPD
• *Pupil orientation*—organising CPDL around colleagues' aspirations for pupils including outcomes, • *Diagnostics*—identification and understanding of teachers' starting points (professional identities practices and motivations, beliefs, approaches to learning, existing knowledge and skills and beliefs), • *Differentiation*—tailoring of learning experiences to teachers' starting points, • Theory and practice—emphasizing practical theory alongside content and pedagogy, • *Collaboration/peer support*—ensuring this originates from iterative trialling and testing of new approaches and evidence about how pupils' respond to accelerate trust, deepen ownership, and refine practice, • *Specialist expertise*—mobilising and integrating deep knowledge regarding CPD content, pupils, and community to provide expert support and challenge via: • Challenging of orthodoxies • Illustration of practices from other settings • Securing depth in learning via evaluation and design of CPDL activities • *Contextualization*—ensuring content is contextualised for each subject but also for sub-groups of pupils—generic pedagogic CPD does not work on its own, • *Experimentation*—effective CPDL involves carefully aligned sustained, iterative experimentation, • *Duration*—iterative experimentation needs to be sustained over time to allow for gradual assimilation, testing and refinement. Short term CPDL can work for very tightly specified approaches to particular aspects of teaching for very specific stages in learning (e.g., spelling of a subgroup of words). Time is required to incorporate results into practice.
Selected Summary of Research on Key Elements of Effective CPDL Leadership
• Positioning CPDL as shared responsibility for pupil achievement and wellbeing • Focusing on teachers' professional growth as well as developing knowledge/skills • Modelling openness to professional/leadership learning • Designing structures/systems for managing complexity taking fully into account the cognitive, practical & emotional demands made by CPDL approaches and systems

Adapted from Cordingley et al. (2015, 2018, and 2020) [18–20].

What should be stressed in relation to pupil orientation, in particular, is that both providers and participants take the time and effort to be clear and specific about what the participants' pupils' learning will look like if the CPD is successful and to design and select activities on that basis. Also noteworthy is that one of the biggest practical challenges facing CPD providers is the huge range of starting points in relation to skills, knowledge, expertise and beliefs of the participating teachers, which inevitably makes differentiation demanding—and expensive (see Cordingley et al., 2018) [19].

1.4. Effective Initial Teacher Education

Effective initial teacher education to a large extent aligns with the above. But there are important differences and additional considerations to consider as Hargreaves (1993) [23] points out in his 'Practical Common-sense model of teacher development'. For example, pre-service teachers need rapidly to develop a body of knowledge and set of practical skills

about the business of teaching and learning and to integrate the two within classroom contexts as well as with their specialist knowledge and skills. The literature and evidence base on effective ITE is limited; the following summary therefore is largely based on the experience of the review team and is primarily presented to surface our assumptions prior to the following review section. In overview, and in addition to the principles for effective continuing teacher education, we hold that ITE accentuates the following particular principles set out in Table 2, below:

Table 2. Overview of Additional Principles for Effective ITE.

Summary of Principles for Effective ITE Additional to CPDL
• *Managing Overload*—balance the need for explicit, detailed instruction and the need to meet statutory expectations with avoidance of overload, helping pre-service teachers focus on the fundamentals of practice. • *Foundational Curriculum*—provide an especially coherent and well-sequenced curriculum that covers all foundational knowledge while meeting individual needs of pre-service teachers, allowing pre-service teachers to acquire and refresh a significant body of subject content knowledge (ck), pedagogical knowledge (pk) and pedagogical content knowledge (pck) in tandem, exploring how these come together and their relation to pupil needs and learning. • *Grounded Instruction*—providing well-illustrated, contextualized instruction and the opportunities and skills to closely observe practice and develop as reflective practitioners. • *Manage risk and develop self-efficacy*—helping pre-service teachers manage uncertainty and seek expert (including mentor) support, accessing the practical wisdom and reasoning of experienced teachers, and understanding how this relates to their own developing expertise (Twiselton, 2016), situating themselves in and contributing to supportive professional learning environments. • *Orientation to Professional Identity and Community*—consider the implications of students typically being new to the world of work, professional working and school communities, supporting students to adopt a range of roles within a professional learning environment, begin the development of a holistic professional identity that fosters professional learning (of themselves and others), and advances and critically reflects on educational purposes and values of their communities and more widely. • *Organization of the curriculum to ensure* content and aims are sequenced and prioritized to fit the training timescales, structure, statutory and non-statutory requirements of teacher preparation, and particular learning and practice settings.

There are several distinctions that are likely to have implications for the design of remote and blended approaches to ITE, relative to CPD for in-service teachers. Notably, ITE providers must manage the risk of knowledge overload—but also can build upon widespread recognition that there is a need to explore a large canvas. In contrast, for those supporting experienced teachers there exists an extensive iceberg of knowledge and skills that has can be deployed rapidly in classroom settings. Those supporting CPD need to find ways of helping teachers to surface, review and refine this knowledge base in the light of new approaches and evidence. ITE providers need to ensure they 'carefully craft coherently sequenced curricula that meet the particular needs of their trainee teachers, including the foundational knowledge of what subjects and curricula are', (DfE, 2019, p. 6) [24]. Therefore, education and professional development for pre-service teachers needs to span a very wide 'curriculum' and providers can be relatively confident that pre-service teachers will not already have encountered many of the skills and much of the knowledge they need to acquire.

2. Methods

2.1. Overview of Methods

Scoping work, consisting of initial literature searches using major research databases (as used in the main review, see below), identified relevant reviews and background literature for inclusion in our descriptive and exploratory review sections (Sections 3 and 4,

below). It also informed a review protocol containing a list of search databases, search terms, eligibility criteria and an overview of the evidence review process (Section 5) as described further below. In overview, our scoping work involved, first, initial location of key reviews of CPD and/or Teacher Education, its modality or both—sources were located through database searches and expert advice; second, developing a set of research questions that have been used to structure our main report (as above).

Given the time and budget constraints for this project, two review methods were blended to produce the most useful and highest-quality review:

- A Review of Reviews—to (a) identify findings pertaining to modality in existing reviews and (b) identify implications of CPD and ITE reviews for modality. Reviews were drawn on to develop the review conceptual framework, and to form the descriptive and exploratory discussion of relevant practice and theory underpinning Sections 3 and 4. We also included background pieces identified during scoping, studies from within reviews, and studies within our evidence review database to develop these sections.
- A Rapid Evidence Review—systematically searching reviews and trials of distance and/or blended initial or continuing teacher education interventions. This includes studies of the relative effect of teacher education modality. In line with previous reviews of effective CPD (e.g., Cordingley et al., 2015) [18], we restricted our focus within the systematic element of the review to programmes reporting pupil outcomes (see below for further discussion of this point).

2.2. Review of Review Methods

When planning this review, the review team considered whether a rapid evidence assessment alone would be sufficient for informing practice in this area. We identified two main areas of methodological limitation of solely conducting a review of evidence from trials:

- First, the need to tightly focus search terms to identify studies with a specific modality focus, combined with the likely scenario that the literature is disparate and/or would provide only limited evidence, would result in lack of coverage of the rich combination of practices that comprise CPD; in other words, we were concerned that the review would be limited to discussing the small number of approaches which had been trialled and not be able to provide a broader account of remote and blended teacher education more generally.
- Second, we recognized that there were likely to be difficulties translating and applying the evidence to the specifics of the review and funder organizations' CPD offers and COVID-19 conditions. We felt that a more theoretical analysis would be needed to apply (a) teacher reviews (which are well-based in evidence) and (b) teacher education modality reviews which were unlikely to have a large, robust evidence base to draw on, to address our questions.

In short, we felt that a rapid evidence assessment alone would produce a report which would be too narrowly focused (on the limited causal evidence) and therefore would fail to provide a research-based account to many of the current questions which teacher educators had at the time. This motivated our 'review of reviews', or 'umbrella review' strand. A review of reviews, or 'umbrella review', synthesizes and revisits existing reviews in an area to identify key findings in the literature and give a high-level overview. Drawing on existing review work in an area increases efficiency, but reduces transparency, replicability and ability to define the focus of the review as it must work with the specific review questions used in extant studies. The 'umbrella' review focused on reviews or substantial reports (e.g., government or research organization) in two areas:

1. Reviews of teacher education (ITE or CPD) effectiveness
2. Reviews of blended and/or online teacher education.

This review had three objectives:

- identifying additional remote and blended CPD trials within the studies included in the reviews for inclusion in the evidence assessment,
- identifying specific findings within the reviews pertaining specifically to CPD modality,
- drawing out and discussing implications of general teacher education review findings for CPD modality.

This second strand of our review gives it more breadth and coverage than a rapid review can efficiently do alone and puts it in a position to build on and re-interpret existing work, drawing out implications for CPD modality. For the exploratory and discussion sections of the review (Sections 3 and 4, corresponding to research questions 1 and 2), we focused our attention on reviews in our database. We also drew on information from studies from the main evidence review (e.g., descriptions of remote and blended teacher education programs), pieces identified during scoping work, and pieces identified as potentially valuable to support background section during the searches (but not meeting the eligibility criteria for evidence review set out below). The groups of literature underpinning Sections 3 and 4 are listed below; we provide grouped reference lists for these in a supplementary document associated with this paper.

- Teacher Education, Professional Development and Learning Reviews—These were located during scoping searches and from pieces known to the review team. These are reviews of teacher education without a specific focus on remote or blended approaches.
- Remote or Blended Professional Development and Learning Reviews—These were located during scoping searches and from pieces known to the review team. These are reviews of teacher education with a specific focus on remote or blended approaches.
- Empirical Pieces—These are the main evidence review database of empirical tests of remote and blended approaches. These were located through systematic searches. Descriptions of approaches from within these informed descriptive and exploratory parts of this review.
- Background Pieces—These are neither reviews, nor trails, but retained during the search process as potentially providing descriptive or theoretical information supporting the initial review sections.

The key limitation of the exploratory review sections (Sections 3 and 4) is that, while systematic approaches were adopted, we relied heavily on our judgement and expertise in our selection of relevant literature and findings within it to summarize. This brings in an inevitable degree of subjectivity and reduces transparency. We have clearly cited all sources on which we have based our account and have not knowingly excluded relevant (including contradictory) results. We were guided by common principles of trustworthiness for qualitative analysis including: looking to triangulate empirical evidence and perspectives in the literature to gauge credibility; paying attention to and reporting details of the context of studies to assess transferability; underpinning our review with a systematic searching and screening process (see below, and Appendices A–C) and reporting of all relevant studies and findings to enhance confirmability; and detailed reporting of our methods here and in description of our process of enquiry in the substantive sections to encourage dependability in our results.

2.3. Rapid Evidence Review Methods

A rapid evidence review, or rapid evidence assessment, is a scaled down systematic review, in terms of time, budget and scope. It shares the need to use transparent and reproducible search techniques. It typically seeks to identify, screen and extract data from all relevant studies in a defined area (often focusing on causal evidence of impact from experimental or quasi-experimental work) but reduces the size of that area through narrower inclusion and exclusion criteria in terms of dates, focus and/or quality. The rapid review focused on the most relevant and robust studies that are available and be manageable within the constraints. It was unfeasible to review studies that do not have an explicit focus on CPD modality (i.e., by including all CPD trials and then identifying and assessing modality effects through data extraction and analysis). On the other hand,

we recognized that the evidence base for CPD modality specifically was likely to be quite limited. As a result, we refined the focus and search terms for the rapid review following our scoping work to focus on the areas of the literature most likely to yield experimental or quasi-experimental results to inform practice. In overview, the rapid evidence review (Section 5) involved:

- *Conducting exhaustive, systematic searching, screening and extraction* for areas of the literature identified during scoping as most likely to yield experimental or quasi-experimental results to inform teacher CPD practice in remote and blended modes. Key search terms related to teachers (e.g., teach*), to professional development (e.g., educat*, train*, develop*), to modes (e.g., online, blended, remote, virtual etc.) and to methods (e.g., review, evaluation, trial etc.). Please see Appendix A for full details of the search terms, results and databases.
- *Using research quality and relevance criteria* to identify high quality papers for more detailed data extraction and analysis (see Appendices B and C). We extracted abstracts and key features (e.g., methods, findings) for all studies in the database. Due to the time constraints of the rapid review, we did not calculate inter-rater reliability for our screening process.
- *Producing a narrative summary of the evidence.* In the evidence review section (Section 5), we summarized all papers identified as eligible. For exploratory sections (Sections 3 and 4), we produced a selective narrative summary of papers that we judged to provide relevant findings against our pre-determined research questions and organizational structure.

The review eligibility criteria were as follows (see Appendix B for further details):

1. *Date*—Studies conducted in or after 2005 (searches were conducted in August 2020)
2. *Learners*—All children and young people form age 3–18 (inc. SEN and mainstream),
3. *Teacher learning focus*—Substantial focus on teacher learning,
4. *Non-face-to-face element*—All empirical and theoretical pieces that present findings about a mode of teacher learning other than face-to-face,
5. *Pupil outcomes*—All studies included in the evidence review must present some empirical evidence about the impact of the teacher learning on pupil outcomes.

This review obtained 7354 records from 5 search databases, containing dozens of library collections. We first screened these on titles for relevance and to remove duplicates, leaving 1716 records. We followed this with screening on our eligibility criteria (as above, and see Appendix B for further detail) using titles and abstracts (retaining 989 records) and then applying the same criteria to screening the full texts, retaining 73 documents. We retained a total of 19 background and wider pieces which, while they did not meet all eligibility criteria, were retained to provide background information and illustration of the results (see above). We also included several other pieces that were known to the team (see 'Other References', within the references list) and identified 4 'late entries'—3 reviews and one empirical piece (reporting early results of a program). We provide full search strings and the number of records located through each search in Appendix A. We provide an overview of the search process, with included and excluded records, by criterion in Appendix C. The key limitations of the evidence review were that (1) given the limited time scale, we did not calculate inter-rater reliability, (2) our search terms relied on the studies self-identifying as being blended, remote, online and other synonyms (rather than using general searches for teacher education and assessing mode ourselves), (3) the limited evidence base (n = 24 studies) and limited time led us to adopt a simple narrative review summary approach to synthesis; with more evidence and time it would have been possible to take a more granular, and systematic approach to extracting key details and findings of the studies.

In the Evidence Review section below (Section 5), only the 24 pieces fully meeting eligibility criteria are reported. In our wider discussion of approaches and their affordances (Sections 3–5, Q1, 2 and 4, as above), we draw on the full database, with a focus on the

most up-to-date, relevant, and high-quality reviews. All papers to which we refer are cited throughout the piece. A full list of references by group is provided following the main body of this report.

3. Remote and Blended Teacher Education Modes and Their Characteristics

Question 1 (Q1). *What forms of/approaches to remote or blended teacher education have been explored in the research literature? What are their typical characteristics/elements? How can this be organised?*

Our first question relates to what approaches are evident in the literature and, presumably, also in common practice. Our aim here is to produce a functional framework, accompanied with some brief descriptions and illustration, to identify categories and key concepts that can be used throughout the review to support alignment and discussion of effective teacher learning principles with remote and blended teacher education approaches. We have not had the time nor resources to be systematic or comprehensive in this particular portion of the review, nor to step systematically through the empirical and review papers on which we have drawn. What we offer is a connoisseurial summary of what we found.

3.1. A Working Framework of Remote and Blended Teacher Education Modes

Based on our review database, we have developed the following simple framework of remote and blended teacher education modes. Whether these can accurately be described as modes, or whether we need to distinguish approaches, components, or media (for example) is not a question we pursue here. We simply note that, evident in our review database are numerous different approaches and characteristics of remote and blended teacher education approaches and variants and combinations thereof. Within these modes, there are clear links to be made in terms of overlap and modes that are typically combined (in particular the combination of modes 2 and 3, and of 4 and 5).

We identified six general modes of online or blended teacher education:

1. Lectures, workshops, seminars, discussion groups or conferences, including one-off sessions and series.
2. Coaching and mentoring.
3. Classroom observations with feedback and/or discussion.
4. Resource bases or repositories, with varying degrees of user interaction and content creation.
5. Platforms and self-study programs, ranging from less to more structured programs that give access to curated/designed resources, learning content, assessments and/or directed activities to learners.
6. Virtual reality spaces or simulations.

While these categories capture common types of activity, a large variation exists within and across these broad groups. We have also therefore identified several cross-cutting factors that characterize teacher education in the modes listed above.

Cross-cutting factors we found to characterise these modes are:

1. (A)synchronicity—With most of the above, it is possible to design asynchronous, synchronous and mixed variants. The advantages of each of these is discussed below.
2. Interactivity—Teacher education varies within and across programmes in the opportunities for interaction versus passivity, and the extent to which teachers are placed as consumers or producers.
3. Community—Programmes can be designed to be more or less collaborative, many seeking to form or situate activity within 'Professional Learning Communities' (PLCs). We note that other frameworks (e.g., Little and Housand, 2011, see below) [25] have PLCs as a mode in their own right. In our view, the extent to which teacher education is an individual affair or part of a community is a more cross-cutting question (i.e., PLCs operate across modes rather than being a mode per se).

4. Choice of (Multi-)media—In remote and online teacher education program design, there is often a choice to be made between the use of video, audio and/or text as a medium of expression, for discussion or to convey information. Teacher education designers are now able to choose from several media and are typically opting for multimedia approaches. Moreover, as technology has improved, the inherent benefits of the media can be foregrounded in decisions, rather than the practicalities (e.g., cost and convenience) of the technology.
5. Combining Elements—Many programmes combine the overall modes, e.g., provide an online seminar followed by remote coaching. This is worth considering in relation to affordances below, that programme elements can be combined to ensure 'coverage' of teacher learning aims and principles and allow each medium and element to 'play to its strengths'.
6. Structure, design and facilitation—Programs vary in the extent to which they have been actively designed (as opposed to spontaneous and 'crowd-sourced') and the extent to which the activity is actively kept to this design through facilitation, direction (in person or through technology and activity timings and content). There are also more general organizational issues around the number, length and timing of programme inputs. While these may feel prosaic, as we discuss further below, these can have a significant impact on the impact of the teacher education and affordances around blended modes and activities.
7. Providers/Partnerships—One other consideration is about the provider(s). Many teacher education programs were delivered and/or designed by a combination of one or more teacher groups, schools, universities, and/or local/national authorities working in partnership.
8. Focus and purpose—A final key area of difference apparent in the literature is the extent to which teacher education in different modes is focused on/centred around a practical purpose (e.g., increasing STEM participation), by identity (secondary biology teachers discussion groups) or with a curricular/learning focus (learning and applying the principles of cognitive science).

We have created this two-part typology (modes–characteristics) both from and to structure this rapid review. We do not claim that it will necessarily map on to other and/or more wide-ranging sets of data/practice, nor do we claim that this is necessarily the most instructive way to conceptualize remote and blended teacher education so as to bring key teacher education or practical considerations into focus. One pertinent aspect that might be fruitful to examine in future is the availability of and resources for the technology involved. Resources, expertise and infrastructure for high-, medium- and low-tech approaches are likely to be key factors for applicability of this framework and the findings of this review across contexts. We also note that this typology is at a high level of granularity, with little detail on the underlying techniques and technologies (e.g., Bower, 2016 for a typology of specific Web 2.0 learning technologies) [26]. We compared our framework to other similar frameworks, with which it mostly aligns. Several examples of these are in Table 3, below:

Table 3. Selected Typologies of Remote and blended Teacher Education from the Literature.

Little and Housand (2011) (as Quoted in Elliot, 2017, pp. 120–121) [25,27]	Snell et al. (2019) p. 210 [28]	Horn and Staker, 2011 (as Quoted in Burns, 2011, p. 70) [10,29]
Mode 1—The first mode of online professional development is accessible websites or online resources. Mode 2—The second mode of online professional development is technology for face-to-face interaction with audiences in real time. Mode 3—The third mode of online professional development is professional development supported by asynchronous online discussion. Mode 4—The fourth mode of online professional development delivery is videoconferencing. Mode 5—The fifth mode of professional development is constructing and facilitating an ongoing online community.	(1) Remote, non-live, asynchronous coaching consists of coaches who work with teachers remotely, including via video sent by teachers to coaches, and phone or e-mail exchanges between coach and teacher about the content. (2) Remote live coaching consists of coaches observing and providing feedback to teachers live, using webcams. (3) Online course content allows teachers to access online materials and read about [curriculum content], watch videos, and take quizzes. (4) Online group courses or satellite courses are similar to standard PD, except that teachers attend a viewing of the PD programme offered online or over satellite with other groups of teachers. (5) Online downloadable curriculum or lesson plans consist of resources available online for teachers to use in the classroom.	1. Face-to-Face Driver Model: The face-to-face teacher delivers most of the curriculum and uses online materials to supplement. This model often occurs in a computer lab. 2. Rotation Model: Students rotate equally between face-to-face and online components of the course on a fixed schedule. They have the same teacher for each component. The online component occurs remotely. 3. Flex Model: The online component delivers most of the information, with an in-class teacher present to provide flexible support as needed. This model includes lots of individual and small-group, face-to-face tutoring. 4. Online Lab Model: The online teacher delivers the course in a brick-and-mortar classroom, but with paraprofessional or teacher aides supervising students. 5. Self-blend Model: Individual students take online courses à la carte. Online learning is remote, but traditional instruction is brick-and-mortar. 6. Online Platform Model: Instruction and materials are all online, with students taking the course remotely. Weekly check-ins with a face-to-face supervisor or teacher are required.

3.2. Selected Examples and Illustrations

In this section we discuss and provide selected illustrative examples of remote and blended teacher education approaches, in line with our framework and drawing on the literature we have reviewed. For online initial teacher education, Dyment and Downing (2020) [11] provide an up-to-date and systematic mapping of the literature. This literature is replete with varied examples of practices in online ITE; indeed, as the review authors find in their first theme in the literature, over three quarters (n = 381) of the 492 papers they reviewed profiled an online pedagogical 'innovation'. These included, 'a wide diversity of synchronous and asynchronous innovations . . . web conferences, discussion boards, chat rooms, instant messaging, digital games, social media, virtual laboratories, virtual simulators, discussion boards, portfolios, and Facebook" (Dyment and Downing, 2020, p. 326) [11]. The papers often described the use of these innovations in ITE courses, as well as the affordances they could offer teacher educators and ITE students. As per their third theme, a small number (14%) of articles profiled how the approach had been contextualized in specific subjects or curriculum areas. The focus of most of these papers was more whether it was 'possible' for these areas to be taught online, rather than the technological innovations per se, and identifying the strengths and weaknesses of the online mode as experienced by both ITE students and teacher educators (ibid. p. 327).

Having located this systematic review, our aim is not to unpack these studies to the point of providing further elaboration on the details of innovations and practices. Instead, we refer the reader to the Dyment and Downing (2020) [11] review and the papers therein. In the remainder of this section, we describe each of our modes. We have identified several reviews and case studies that allow us to give an overview and illustrative examples. We

note that we are not advancing any of these as examples of *effective* teacher education, rather they are just examples of programs that fall within our typology and allow us to explore characteristics, variations and issues of the modes before the next section where we consider remote and blended teacher education and the principles for effective teacher education side-by-side.

3.2.1. Mode 1—Lectures, Workshops, Seminars, Discussion Groups or Conferences

Our first mode for blended and remote teacher education consists of approaches that facilitate dialogue between a teacher or pre-service teachers and other teachers, pre-service teachers, instructors, experts, stakeholders, or a combination of these. However, at the boundaries of this are: firstly, lectures, which can often include minimal dialogue and are therefore more akin to video content on platforms and self-study programs (see mode 5); second, conferences, which typically combine numerous elements of which only some resemble a seminar or discussion group. When it comes to conference keynotes and presentations, where the information flow is one directional, these may be thought of as being more similar to static content and our content platform mode (5); when it comes to unstructured discussion between conference delegates, this can have the qualities of more informal professional learning communities (a cross cutting characteristic discussed below); for conference question and answer or discussion-focused sessions/periods, we feel these have a great deal in common with seminars and workshops, sharing the characteristic of facilitated discussion that we have identified for this mode.

Varying interactivity and structure within conferences is one theme discussed in Seddon et al. (2012) [30], who present a model of web conference activity over the course of the conference in terms of their level of social interaction, information provision, internalization of information by participants and co-construction of knowledge. They find, in short, that: 'Social interaction was most noticeable at the beginning of sessions then tailed off as the presentations started, but reappeared at the end, as networking for the future occurred' (p. 445). The more general point, applying across all modes, is that the amount of discussion, and information presentation is variable both across and within modes, but even within individual units or sessions in a specific approach. As we touch on above, as technology has improved, teacher education providers are increasingly able to choose the medium/media of expression according to what is effective and desirable rather than what is practicable. For example, and of particular note, is that in recent years it has become increasingly possible to enhance teacher education through interactive use of video to support observation, discussion, modelling, mentoring and coaching. Recent developments in technology have enabled synchronous, video-based small group activity within larger group workshops. This is an important new affordance that has been widely used during COVID-19 lock-down, for example, to enrich peer support and differentiation but is unlikely, as yet to feature in the research. There are, however, numerous examples in the research of the use of video for remote seminar/workshop or conference discussion.

One extended discussion of this practice in connection with video conferencing is found in Burns (2011) [10]. Burns (2011) [10] particularly noted that the medium gave access to learning experiences that geographically distant teachers might not otherwise be able to participate in, while also reducing misunderstandings caused by the lack of non-verbal cues that may characterize other forms of online learning. However, there remain issues such as audio-visual quality, time lags, drops in service, coordination of activities and suitability of activities for the medium (Burns, 2011) [10]. We note the date (2011, discussing a Canadian network operating in 2001–2002) and the discussion of technological as well as teaching and learning considerations. Finally, Burns (2011) [10] noted that the instructor may not be able to see all the participants and that working with simultaneous face-to-face and remote groups might mean that one group was prioritized over the other.

Another key consideration for lectures, workshops, seminars, discussion groups and conferences is whether these are one-off events, part of a series and/or linked to other forms of activity. The hallmark of activity within this mode is to structure discussion

around focus questions or content. Decisions around the number of sessions, their duration and timings shape how long the 'conversation' can continue, and the depth and quality in which the questions/content can be considered. Choices around frequency and timings also affect who can participate and the extent to which other teacher education activities can take place between the sessions. A seminar series spread across a school term, for example, allows classroom and other teacher education activity to occur alongside the series and links to be made between these. A conference—even if the questions, content, and opportunities for discussion remain the same—is harder to integrate with other activity in this way.

3.2.2. Mode 2—Coaching and Mentoring

Relative to other modes discussed here, coaching and mentoring are relatively well-defined (and have a strong evidence base finding substantial impacts on teacher learning and pupil outcomes (e.g., see Kraft, Blazar and Hogan, 2018) [31], although practice and quality varies considerably (Hill et al., 2020) [5]. The National Framework for Mentoring and Coaching (CUREE, 2005) [32] describes and explains the principles, concepts and skills for mentoring and coaching and compares the overlaps between these (also see Cordingley, and Buckler 2012, and Fletcher and Mullen, 2012) [33,34]. The main difference between remote coaching and mentoring and face-to-face variants is simply that remote versions of coaching and mentoring must use video, audio or textual media to facilitate exchange between the coach/mentor and coachee/mentee. The media ostensibly vary in their suitability and affordances to sustain high quality mentoring and coaching conversations and to the degree they enable synchronous and asynchronous conversations.

3.2.3. Mode 3—Classroom Observations with Feedback and/or Discussion

Closely linked to mentoring and coaching is the observation of classroom teaching. We have included this as a separate mode as video technology allows discussion of classroom teaching in other forms of teacher education, including in discussion groups or on online repositories. So much so that one might argue that rather than being a mode in their own right, classroom observations with discussion and feedback are simply the content for other forms of interaction. Be this as it may, given the centrality of observation of classroom practice to many forms of teacher education and the growing possibilities for (synchronous and asynchronous) observation of classroom teaching afforded by developments in video technology and platforms, we have categorized classroom observation as a mode in its own right.

The literature on the use of classroom video for professional learning is summarized and discussed in Perry et al. (2020) [35]; a short summary of selected points from this is provided below, along with an example. Perry et al. provide an overview of potential affordances of video observations for enabling and enhancing CPD, based on reviews in the literature (e.g., Brouwer, 2011; Gaudin & Chaliès, 2015; Major & Watson, 2018) [36–38]:

- It captures classroom interactions for review outside of the classroom (Brouwer, Besselink, & Oosterheert, 2017; Pehmer, Gröschner, & Seidel, 2015; Tripp & Rich, 2012) [39–41].
- It can foster collaborative approaches to professional development: making it possible for joint learning between colleagues and/or external experts (Brouwer et al., 2017) [39].
- It may assist teachers to recognize previously unnoticed issues, allowing conceptions to change (Gaudin & Chaliès, 2015) [37]; a feature recognised by the 'Lesson Study' (e.g., Lewis, Perry, & Murata, 2006) [42] and 'Learning Study' (Davies & Dunnill, 2008; Holmqvist, 2010) [43,44] literature.
- It can offer convenient opportunities outside the school day for observations and discussions with teachers and their colleagues, reducing supply cover costs (Quinn et al., 2019) [16].

- It is efficient, in that the most instructive videos can be viewed repeatedly by a wider audience.
- The use of video platforms can support processes of teacher reflection, discussion and analysis through tools for editing or annotation (Rich & Hannafin, 2009) [45].

Running through these points one can see the connection with: external expertise and perspectives; discussion groups; enquiry and action research approaches; and platforms—hence why this may be thought of as an activity within more than a mode of teacher education. Questions that arise from this literature include whether it is of greater impact to review one's own teaching, or that of others (including colleagues or experts). When this was systematically studied the answer seemed to be 'both': viewing of the work of others can make teachers more receptive to new ideas and more engaged in the analysis of difficult incidents (e.g., Van Es, 2012; Borko et al. 2008; Kleinknecht and Schneider, 2013) [46–48], but this appears to be especially effective when viewing videos of others is preceded by teachers reviewing their own teaching (e.g., Beisiegel et al., 2017; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011) [49,50]—it seems it is the connection that counts.

One example in this area is Jamil and Hamre (2018) [51] in their study of online professional development on principles of learning from cognitive science. The programme takes a structured approach to teacher reflection asking teachers to observe, analyze, assess and plan, which the authors discuss in relation to cognitive science principles. In relation to observation, they draw attention to the value of videos for identifying specific, situated interactions and opening these up for conscious consideration, examining what are often tacit and unconscious decisions based on automatic schema, making them explicit again (p. 228). They find that the approach can enable teachers to examine objectively the consequences of their actions by distancing them from the immediate, emotional reactions to the situations (Jamil & Hamre, 2018) [51].

In overview, videos of classroom teaching provide teachers with 'objects to think with' (Jamil & Hamre, 2018, p. 74) [51], which connect in a concrete and authentic way to teacher's practice. Use of classroom videos is widely held to hold promise for effective professional learning. The evidence base, however, is not yet sufficiently developed to either support or refute this perception.

3.2.4. Mode 4—Resource Bases or Repositories

Another mode that has not sat easily as a stand-alone item in our typology is that of resource bases or repositories. Should providing access to a resource base even qualify in its own right as a form of teacher education? We are left wondering whether resources bases are better thought of as being content or curricula for teacher education rather than types of it, though we observe that many view curriculum materials as 'an important source of teacher professional learning in and of themselves' (Hill et al., 2020, p. 52) [5]. This question aside, there are numerous examples of resources bases that—to varying degrees—encourage teacher collaboration and professional learning through creation, discussion and use of resources ranging from lesson plans and curriculum schemes of work to worksheets, ideas, stimulus material and assessment tools for use in the classroom.

The Times Educational Supplement (TES) teaching resources page provides over 900,000 resources made by teachers, sorted into curriculum areas and topics, with the option to review resources and join in with over 250,000 discussion threads in the community section. More recent examples include Teacherly, which emphasizes to a greater extent the collaborative and professional learning aspects of resource creation and sharing, as well as providing a platform for remote teaching. Similarly, linking with the previous section, is that there are examples of resource bases that include classroom videos, such as those investigated in Bates, Phalen and Moran (2016) [52], which at the time they were writing housed over 350 video clips used by over 37,000 members. As noted by Dede et al. (2016) [53], resource collections can also include tools and apps; also, as groups get large, 'crowd sourcing' even specialist, topic or subject-specific resources becomes more

viable. Members of the review team are increasingly seeing schools and groups of school developing their own approaches to collaboration and sharing of resources (such as using software such as SharePoint, Google Drive and other similar platforms). Before moving on, we note the parallels here with the questions that arise in relation to PLCs (below) relating to quality control, depth of engagement and participation.

3.2.5. Mode 5—Platforms and Self-Study Programmes

Many programs we have reviewed are based around an online platform, or virtual learning environment. This is the area that is most varied in terms of content, activities (and apparently quality) and there are many points of contact with other modes. Indeed, it is possible to describe most online platforms for teacher education as being a curated, structured resource base (often with added instructional material) with wrap-around collaboration and expert input in the form of seminars, mentoring, discussion groups and or (less commonly) lesson observations. Moreover, where a platform and its associated activities bring together teachers and pre-service teachers for discussion, sharing and support, it can be said to be an organizational focal point for a PLC. In short, teacher education online platforms and programs that use them can and often do incorporate one or more (other) modes of teacher education. Typical examples include Dana et al. (2017) [54], who provide a detailed description of the Prime Online program, a year-long program designed to develop subject (pedagogical) content knowledge and pedagogical content knowledge in mathematics elementary (primary) classrooms through practitioner inquiry. The Prime Online experience consisted of "three distinct, yet integrated, segments of weekly modules consisting of content and experiences designed to build on one another" (Dana et al., 2017, p. 215) [54]. It was designed and delivered via Moodle™ (an online learning platform), with a consistent, four-component format for each module.

The learners begin each week by logging in to find an introduction and an overview of activities for that week, including aims and resources (Dana et al., 2017) [54]. They have an Anticipatory Activity that consists of some sort of provocation to enable them to reflect on prior experiences and connect them to the new material plus readings, videos and 'web quests' (ibid p. 215). The learning is assessed through assignments where participants have to apply what has been learned to their classroom practice and whole group learning and discussion is summarized and analyzed by the facilitator (ibid). Dana et al. (2017) [54] go on to discuss the program, its content and how participants engaged with it in detail. While their focus is more on practitioner inquiry than the affordances of online teacher learning approaches, it nonetheless provides a rich illustration of how online approaches to teacher learning can be designed and implemented. Although space does not permit us to pursue them further here, other rich examples of programs in this area are provided in Owston et al. (2008) [55], who review three program evaluations of blended teacher professional development and exploring, for example, issues around program structure and its links to expectations and relevance. As Owston et al. (2008, p. 208) [55] describe, "the more structure that a program imposed, the less flexibility it provided teachers to experiment with activities in the classroom at same time they were planning on teaching them".

In sum, learning platforms are clearly performing multiple functions, bringing together and organizing teacher education modes, content and activities. 71 out of 89 studies of professional development programs for STEM teachers reviewed by Hill et al. (2020) [5] for example combined new curriculum materials for teachers to use in classrooms with professional development. Elliot (2017, p. 121) [27], in their review, discuss how 'course management systems' or 'learning management systems', which they define as a 'collection of online learning tools contained in one system' can be used to achieve multiple aims, also cautioning against the assumption that conventional face-to-face programs can be straight-forwardly transferred into the differing dynamics of online environments.

3.2.6. Mode 6—Virtual Reality Spaces or Simulations

The final mode that we have identified but will not be pursing further past this brief section is the use of virtual reality (VR) or simulation spaces for teacher education. This includes the use of virtual environments such as the popular 'Second Life' virtual world as a space for developing teaching skills or its components and the use of VR to immerse teachers in observation of virtual reality classrooms (i.e., 360-degree videos of real classroom activity). Burns (2011) [10] describes the "immersive, highly synchronous attributes" (p. 100) of one popular virtual reality space, Second Life, as having huge potential as a distance learning tool. We refer readers interested in finding out more about virtual worlds to systematic reviews by Theelen, Van den Beemt and den Brok (2019) [56]—focused on interpersonal competence for pre-service teachers—and Billingsley et al. (2019) [57], who systematically examine VR in pre-service and in-service programs.

When it comes to VR classroom experiences, we feel that many of the affordances and questions linked to the use of classroom video, as discussed above, apply. Like classroom video, VR technology promises the potential for immersive and concrete classroom experience, which when organized around principles for effective teacher education apparently have great potential to support teacher learning.

3.3. Professional Learning Communities

Before moving on to discuss the affordances and limitations of remote and blended modes, we consider one of the cross-cutting factors in our framework: the extent to which the teacher education program is delivered within and/or forms a professional learning community (as opposed to individual-level, or small group teacher education). While other cross-cutting factors are discussed in specific sections throughout the report, as PLCs are in other reviews considered a mode and, moreover, raise many common issues around the nature of collaboration, expertise and focus, this area warranted a dedicated section.

Remote and blended professional learning communities have attracted considerable attention from both practitioners and researchers. It is an area of this literature that comprises not only numerous research pieces, including many case studies, but a growing body of meta-analyses, reviews and reports. This is a sign of the growing maturity of both research and practice in this area. This said, what is not yet evident in the literature (or at least from our searches) are many examples of attempts to evaluate the impact of online (or indeed any) professional learning communities in terms of their impact on pupils, as opposed to changes in teacher practices or teacher satisfaction. PLCs have evolved for numerous purposes and have not always been well aligned to principles for effective teacher education. While we can provide—via reviews identified in the study—a descriptive and theoretical account, some caution is needed in drawing any conclusions about the design principles for effective teacher education within PLCs.

We identified numerous case studies and reviews focused on online or blended professional learning communities. There was huge variety evident, something that the reviews in the area have also grappled with (e.g., Macià and García, 2016; Lantz-Andersson, Lundin and Selwyn, 2018; Khalid and Strange, 2016) [58–60]. Macià & García (2016) [58] found that the range of online communities and networks for teacher PD is wide and still in fairly early development, the corresponding theoretical frameworks are quite varied and hard to compare, and that the practical impact on teacher PD is unproven and the mechanics of such impact unclear. Remote and blended professional learning communities vary in terms of (for example):

- Formal or informal organization (Lantz-Andersson et al., 2018) [59].
- Extent of time commitment desired and achieved by individuals and the combined community.
- The media/technologies employed for PLC activities and discussion. However, much of the research in this area fails to mention or describe the technological basis for the PLC in any detail (Lantz-Andersson et al., 2018, p. 305) [59]. Technology, particularly

in relation to issues of access and technological understanding of potential PLC members, is perceived as a key barrier to participation (Khalid and Strange, 2016) [60].
- Their focus and activities (see below).
- Their size and geographical reach.
- The heterogeneity of the community in terms of their phase, subject expertise and so on.
- The balance of online to face-to-face contact for blended communities.

This list, while far from exhaustive, already casts PLCs as a fairly nebulous concept. What seems to be common to PLCs is that they bring groups of educators together around a purpose (e.g., a school improvement initiative) and/or identity (e.g., secondary biology teachers interested in research-informed practice). It is this that motivates our decision to classify PLCs as a factor rather than a mode. PLCs could therefore be a group that meets to observe and discuss teaching (mode 1) or one based around a teacher education programme using a virtual learning platform (mode 5), and so on. PLCs frequently extend other teacher activity, creating additional opportunity for collaboration, discussion and sharing. They also exist in their own right, again varied, and characterized in part by the above factors. PLCs existing in their own right often centre on sharing and peer support. In their systematic review of formal and informal PLCs, Lantz-Andersson et al. (2018) [59] describe the sharing function of PLCs as distinct from the construct of professional learning as focusing on a narrow syllabus because it can distil information from the knowledge and experiences of the group. Nevertheless, they do caution that PLCs may not lead to deeper level learning as many often became " . . . sites for 'superficial' sharing of information, quick exchanges and a 'smash-and-grab' approach to becoming informed" (Lantz-Andersson et al., 2018, p. 311) [59].

Capturing some of the possible benefits of PLCs, as discussed above, Frumin et al. describe the nature, organization and timings of activity in an Advanced Placement Teacher Community (APTC). The online community is generally sustained by the activity of the teachers themselves, though each APTC does have a moderator. In contrast to other PD offers, the APTC can be accessed flexibly throughout the year and around the clock. They add: "In addition to being 'bottom-up' and ongoing, the APTC has the following two primary attributes, based on survey and case study data, that complement and extend top-down forms of professional development: (1) personalisation of content and (2) a shared, affective community" (Frumin et al., 2018, p. 413) [61].

Despite (or perhaps because) of their bottom-up and ongoing nature, PLCs can have issues with engagement levels from all members of the community. Frumin et al. (2018) [61] found that only about half of their survey respondents used the APTC and many of the participants in the online community were what is known as 'lurkers' (i.e., individuals who observe but do not actively participate—something we discuss further below). They also find (and this is echoed in Khalid and Strange, 2016) [60] that PLC members can struggle to feel safe (e.g., from criticism or judgement) or able (e.g., technological, cultural or expert knowledge) to fully participate in the community. Some survey respondents in Frumin et al. (2018, p. 415) [61] 'note that the online community does not always feel safe given the employed moderation techniques (or lack thereof) and/or the domination (or bullying) by a few strong voices.' Other rich case studies we have examined in our review include Holmes (2013) [62], who presents a case study of eTwinning professional learning community, and Dede et al. (2016, Chapter 6) [53], who describe a 'Just-in-time' professional development community called the 'Active Physics Teacher Community' that "has the specific intention of helping teachers plan their daily lessons and 'providing them with formal instruction that is directly related to the lessons they are teaching; share their knowledge, experiences, successes, and challenges with other teachers who are using the same lesson plans and curriculum; and compare the effectiveness of their teaching . . . " (Dede et al., 2016, p. 162, Chapter 6) [53].

4. Affordances and Limitations of Remote and Blended Approaches to Teacher Education

Question 2 (Q2). *What are the affordances and limitations of remote and blended teacher education relative to face-to-face variants? What adaptations, restrictions or enhancements are typical relative to face-to-face variants?*

We tackle this question from two perspectives:

a. What, from a theoretical standpoint, can be inferred from the literature on effective continuing professional development and learning (CPDL) and initial teacher education (ITE)?
b. What is advocated as best practice in the remote/blended teacher education literature?

It should be stressed that this is an exploratory treatment of the question. While reviews such as Surrette and Johnson (2015) [63] and several other pieces we discuss below provided an excellent starting point, the literature (even when taken as a whole) did not yield an extended or comprehensive account. So, in what follows we draw on our own expertise as well as the studies and reviews in the review database, where these discuss or present relevant principles and issues. This is an exercise in synthesis and interpretation that is not (unlike our evidence review) based on exhaustive and systematic treatment of relevant studies. Our aim is to entertain and explore hypotheses and plausible connections between the effective teacher education principles and the affordances, limitations and practice of remote and blended teacher education.

Our opening observation is that remote and blended teacher education is a broad church. The accounts of many of our review studies tell us something about a program, but not about the modality in general. As discussed above, there is huge variation in practice and the technologies employed with each mode, and the apparent quality in their implementation. Reflecting this, and also our response to it, is a comment from Dede et al. (2016) [53] in relation to technology:

> "When compared to context, goals, or pedagogical design, technology is arguably the least important component of a learning environment. That being said, the affordances of different technologies enable new variations and forms of organizing activity. There are activities which are much easier to do with particular technologies when they are used well."
>
> (Dede et al., 2016, p. 36) [53]

Therefore, we think there is currently great value in reviewing modes and technologies of teacher education in relation to and while not losing sight of teacher education principles, not least because of the possibilities and potential for missteps of the rapidly developing potential for remote and blended teacher education being opened up by technological development and circumstance (including the present COVID-19 pandemic). The reviews we have examined reveal a landscape which is 'a divided, unsettled, and challenging space with pockets of acceptance, but characterized by epistemological and pedagogical questions, doubts, and uneasiness' (Dyment and Downing, 2020, p. 327, commenting on online/blended ITE but in our view more widely applicable) [11].

While one modality may not be necessarily superior to others and the implementation is as important as the design, the affordances and limitations of the modes do seem to vary. Indeed, the fact that many programs are combining modes and approaches is often in deliberate recognition of the strengths and weaknesses of each (Elliot, 2017) [27]. There is, therefore, value in attempting to grasp these. We continue this discussion in connection with specific teacher education principles and perspectives presented in the review introduction (Sections 1.2 and 1.3). To ease this discussion, the principles have been grouped with each other and related CPDL, leadership of CPDL and ITE principles connected where possible. While there is obviously no perfect grouping and delineation possible, these do bring together some connected considerations pertaining to teacher

education modality for pre-service teachers, teachers and leaders for a more focused discussion.

4.1. Pupil Orientation

Perhaps surprisingly, there is very little focus on pupil outcomes in the teacher education online/blended literature, much of which is focused on changes in teacher practices or thinking and/or teacher perceptions on and experiences of new ways of learning; this is especially true for the initial teacher education literature (Dyment and Downing, 2020) [11]. This can also be seen in our searches: when we screened 989 papers using full-texts 749 (over 75%) were excluded from the review as they reported no empirical data on pupil outcomes arising from the teacher education of any kind. This runs counter to the finding in the effective teacher education literature (e.g., see Cordingley et al., 2015) [18] that the organization of CPDL around teachers' aspirations for pupils and how teaching practices affect learners and learning was the only common feature of all effective CPDL. It is therefore a concern that such a large proportion of pieces describing and evaluating teacher education programs paid such little attention to what, if any, benefits and impacts there were for pupils.

For an example of a programme where a pupil orientation was evident, we can return to the previously discussed 'Just-in-time' Advanced Placement Teacher Community (APTC) reported in Dede et al. (2016, Chapter 6) [53]. The programme platform was structured around a 'Prepare, Share, Compare' pattern, linked to the 'Active Physics' curriculum and its associated textbook and typical teaching structure. On the prepare portion of the platform (see p. 162 for further details) teachers could access videos, lesson plans and were provided with information on the science content and common misconceptions etc.; Share was designed to allow teachers to share experiences, ask questions and receive responses from other members of the community in relation to content, pedagogy (and pedagogical content knowledge); on the Compare section, and pertinent here, is the space and expectation for teachers to use pupil data from quizzes to discuss their impact on pupils and inform their future instruction. The detailed discussion of the Compare feature is particularly relevant to this section. This feature allows the participants to examine the progress and scores of their own students with those of other teachers, as well as explore why the differences may have arisen, why students may have been drawn to incorrect answers, and use the data to improve their practice. These characteristics are unique to his forum and are not replicable in classic forms of PD as there is usually a delay between the classroom activities and the analysis (Levy, Eisenkraft and Fields, in Dede et al., 2016) [53]. It is noted that this collaborative approach was adapted from the Japanese practice of Lesson Study, but the online format enables wider engagement in a "real-time, just-in-time context" (Levy, Eisenkraft and Fields, in Dede et al., 2016, p. 174–175) [53]. We are also reminded of the previously mentioned discussion of video observation in relation to cognitive science principles in (Jamil and Hamre, 2018) [51], who describe the analysis of classroom observations to 'make an honest appraisal of the consequences of these interactions for [the] students' (Jamil and Hamre, 2018, p. 229) [51].

A pupil orientation goes beyond including some consideration of pupil outcomes (e.g., via for example analysis of pupil work or assessments during initial planning or following a lesson observation). Having a pupil orientation is to use aspirations for pupil experience and outcomes as a way of focusing and making coherent, all aspects of the teacher education activities, whether one is designing a curriculum, developing questioning technique or learning about effective assessment. 'How *will* children/young people benefit from, react to, and/or experience this?' and, 'How *have* children/young people benefited from, reacted to, and/or experienced this?' are the central, yet often neglected, questions at the heart of all teacher education activity.

We now turn to the general question of which teacher education modes are a good 'fit' for achieving this feature of teacher education. The general principle of 'distance to practice' highlights that technology (and especially video) has the potential to bring classroom

interactions into a teacher education space for examination, discussion and reflection, without the expense of face-to-face observations (i.e., around release time) and the reliance on memory—which will be most fresh immediately after the teaching and increasingly distant as time passes. We note that in Levy et al. [53], 'student *results* and *memory* of what transpired in classes' is the *basis* for discussion and practice development, whereas Jamil and Hamre discuss 'precise' and 'specific' interactions and their consequences. Remote and blended modes provide—to a greater or lesser extent—a multimedia space to which it is possible to bring rich practice and pupil outcome stimuli for detailed discussion; some of these (e.g., planned seminars and platforms) tend to allow more structure (which allows pupil orientation to be 'built-in'); whereas for others (and especially informal online professional learning communities), it may be difficult to ensure a pupil orientation.

Overall, we judge that there are no modes within our framework for which it is *not* possible to include substantial consideration of pupil outcomes, and many which have advantages (such as cost and ability to create a larger community) when compared to face-to-face methods. Discussion of pupil outcomes is made easier when student assessments or artefacts (e.g., pupil work) as stimulus material and classroom video are built into the teacher education to make links between teaching and learning.

4.2. Collaboration and Support

A common characteristic of teacher education is collaboration. Being supportive of collaboration is one the elements of effective professional development identified in Darling-Hammond et al. (2017) [21], who explain that:

> "High-quality PD creates space for teachers to share ideas and collaborate in their learning, often in job-embedded contexts. By working collaboratively, teachers can create communities that positively change the culture and instruction of their entire grade level, department, school and/or district."
>
> (Darling-Hammond et al., 2017, p.v.) [21]

With this principle and others, it is important not to take the simplistic interpretation that its mere presence will make for effective CPDL, or indeed that it is either a necessary or sufficient condition for effective CPDL. A pupil orientation (see previous) provides the most concrete example of another factor that may or may not be evident within collaborative activity. That is, collaboration can be a superficial 'sharing of best practice' or a rich exchange orientated towards how pupils are responding to teaching practices and curriculum, examining theory and practice side-by-side through iterative classroom 'experiments', and that is sensitive to participants' level of experience (cf. the other teacher education principles). It is not the presence of collaboration *per se* that matters; it is the who, what, where, why and how of it, and how it aligns to the other effective teacher education principles (and especially pupil orientation) which requires consideration.

Putting the detail of what constitutes effective teacher education aside, the focus here is on the 'fit' of particular principles with remote and blended modes of teacher education, and what new affordances, limitations and considerations come to light when putting teacher education principles and what we know about remote and blended teacher education modes side by side.

4.2.1. Presence, Participation and Facilitation

A concept that appears often in the literature we have reviewed is that of 'presence', its importance and the potential difficulties of achieving it remotely. Elliott (2017, p. 118) [27] argues that PLCs ('Communities of Inquiry' in their terms) provide a model for effective practice across any online learning experience which includes 'social presence'. Eliot also discusses 'cognitive presence', and 'teaching presence' (p. 118), which refer respectively to how learners interact with the content in the environment, and the instructional design and delivery within the course. Surrette and Johnson's 2015 [63] meta-analysis suggests that it is *possible* to achieve all of these within an online space. Their analysis of 20 empirical studies (ranging from single-case qualitative studies to randomised controlled trials)

indicated that 'online PD that facilitated opportunities for collective participation enabled the teachers to (a) experience a high level of social connection with their peers, (b) engage in multiple opportunities to reflect on their teacher practice and the practice of their peers, (c) share knowledge and experiences with peers, and (d) increase their knowledge and understanding of unique instructional strategies' (Surrette and Johnson, 2015, p. 266) [63]. Similarly, there are examples (e.g., Lin, 2008, in Keengwe and Kang, 2013, p. 487) [64,65] of students feeling a sense of 'connection and active participation through online discussion' even within a text-only format of an online discussion board. In some cases, interactions being online and/or asynchronous even has advantages for participation, including less confident students being 'able to take the time to compose a response' and participation being less tied to geography and timings of face-to-face interactions.

While high 'presence' appears to be possible, it also seems to need careful consideration when designing remote and blended teacher education. High participant attrition seems to be a general issue for online modes of delivery (e.g., see Goldenberg et al., 2014) [66] as well as the issue of non- or highly-passive participation (Keengwe and Kang, 2013) [64], with interactions being overly tutor-initiated and student responses limited in number or depth (Lantz-Andersson et al., 2018, p. 309) [59]. The inherent limitation of all non-face-to-face interaction, as Quinn et al. (2019, p. 410) [16] explain, is that 'in online settings, the capacity to read body language and facial expression is limited, especially if communication is primarily through text, which challenges the intimacy and immediacy required for social presence (Tu and McIsaac 2002)' [67]. This sense of connection with present, responsive others is not always achieved, as the teacher quotation in Owston et al. (2008) [55] vividly captures:

> "The one thing I found about [the online discussions] was that I get all these ideas and I do some writing and stuff and then press the button and it goes. Mentally it goes out there somewhere. I don't know does anyone see it? No response: does anyone care?"
>
> (Owston et al., 2008, p. 207) [55]

Also, often linked to discussions of presence, participation and responsiveness is the question of instructor and facilitation roles for online courses. Many pieces discuss the roles and values of facilitators to support discussion and collaboration. Dede et al. (2016) [53] devote a chapter (10) to the role and 'importance of moderators' within online STEM CPD, they describe their roles in online communities as being to: 'facilitate organization, welcome newcomers, encourage contributions, support top contributors, foster commitment, address problems and model good behavior' (p. 274). Expectations around participation, and preconceptions about these, seem to vary with teacher education modes as well as potentially being influenced by facilitators and formal or informal rules and etiquettes. Common rules relate to the expected frequency of contributing to fora and the need to respect the opinions and ideas of others (Keengwe and Kang, 2013) [64]. As with presence above, active participation is something often discussed and sometimes, but not always achieved in practice (Keengwe and Kang, 2013, p. 488) [64].

Getting the balance right between coordinated and/or mandated activity and spontaneous participation seems to be a common consideration in the literature. Lantz-Andersson et al. (2018, p. 310) [59] discuss how informally-developed online communities include a 'majority of participants that observed, without making any visible contribution (so-called lurkers).' And, at the other extreme, that the most active communities can often encroach on teachers' free time and leave many 'feeling 'overwhelmed' by the regular flow of information'. Whether or not this balance is well struck and/or whether communities are more formal or informal, *as per* Dede et al. (2016) [53], Lantz-Andersson et al. found moderators had a key role:

> "Clearly no online community is ever spontaneous and an entirely un-organized 'bottom up' affair. In the studies on formally-organized online teacher communities, clear forms of leadership are seen as an implicit aspect of the institutional

> setup. However, even the most informally developed communities being studied were also found to depend on the sustained efforts of moderators . . . In this sense, the apparent difficulty of controlling and planning the activities of all members in large online-groups challenges the extent to which opportunities exist for fostering active participation (cf., Macia & Garcia, 2016) [58]."
>
> (Lantz-Andersson et al., 2018, pp. 311–312) [59]

4.2.2. Community Formation

A theme within the teacher education principles relating to collaboration is the value of forming a professional learning community and working to orient teachers, and especially pre-service teachers, to participate within it. As touched on above, it is possible to wrap community elements around other (including more individualistic) teacher education approaches. Much attention is paid to culture and community development in the literature and reviews generally paint a positive picture of the potential to achieve this in online spaces. This can be seen in relation to online professional learning communities in Lantz-Andersson et al. (2018, p. 310) [59] where communities with a 'friendly, participatory culture' are supportive of experimentation, risk taking and searching discussion and feedback. This culture can take time to develop (as discussed in relation to video sharing and discussion in Perry et al., 2020) [35]. It is a relatively widespread view that community building and functioning benefits from (a) there being an 'established history of sharing and working together' outside of the program (Seddon et al., 2012, p. 436) [30] and (b) there being face-to-face sessions which take place prior to or in between online activity (i.e., within a blended format) (Seddon et al., 2012; Owston et al., 2008; Matzat, 2013; Keengwe and Kang, 2013) [30,55,64,68]. Face-to-face experience was considered by many as the 'glue' that creates a community, and increases rates of participation in subsequent online interaction (e.g., see Owston et al., 2008, p. 207) [55]. It is also worth noting that online spaces are discussed both as a hindrance for community development (as above) and as a benefit: Seddon et al. (2012, p. 435) [30] relate Beatty and Allix's (2005) [68] description of the potential advantages WebEx video conferencing technology in terms of its ability to 'transcend space and time', sustaining a larger group, and providing the ability to catch up on earlier content.

Overall, our reviews suggest that collaboration, including (emotional and pedagogical) support and belonging within professional learning communities is both possible and evident in remote and blended modes of teacher education, both formal or informal (but see Lantz-Andersson et al., 2018) [59], and for both novice and established teachers (Lantz-Andersson et al., 2018; Surrette and Johnson, 2015) [59,63]. The question is whether it is designed effectively to compensate for the lack of face-to-face interaction and as an intrinsic and important, structured element of the CPD that aligns with and realizes other principles of effective CPD.

4.3. Diagnostics, Differentiation and Teacher Starting Points

How can we 'differentiate' teacher education for teachers and their starting points? In relation to CPD more generally, we know for example that mentoring programs are particularly effective for providing personalized, targeted support and instruction and this can improve retention rates for early-career teachers (Fletcher-Wood and Zuccollo, 2020) [3]. To what extent are teacher diagnostics and differentiation possible within remote teacher education?

Modes that are built around the teacher education participants, such as mentoring or teacher-led discussion groups, lend themselves to diagnostics and differentiation in relation to teacher staring points. Similarly, where leaders and initial teacher educators can meaningfully guide and contribute to the teacher education, this allows the wider picture of a teacher's professional growth to be brought into focus and can manage overload, enabling teacher educators and leaders to focus in on fundamental points. Here we return to issues from previous sub-sections around (a) programme structure and its links to ownership

and participation and (b) the ability to use video and/or multimedia to bring classroom practice and pupil outcomes to light within the teacher education space. Trade-offs are apparent in relation to content, structure and assessment:

Content—the greater range of content available, the more tailored it can be to support individual teacher needs; but with more content comes a greater risk of overload and difficulties supporting participants to focus in on and discuss fundamentals. Perhaps the most extreme examples here are the large resource bases/repositories and informal, open networks such as twitter communities. While we may find that there is 'something for everyone', including particularly specialist and specific content to be found—many (and particularly pre-service teachers) may feel overloaded, and content may not be accompanied with instruction (to help the teacher navigate and identify the key pedagogical features of the resource) nor opportunity to discuss these with others. Moreover, and again returning to the example of the APTC, the Share and Compare sections of the online programme were deliberately not pre-populated by instructors. There are benefits in terms of both ownership and depth of thinking of creating and developing ideas, resources and communities—or contextualising more generic content for one's own subject (Cordingley et al., 2018) [19] as compared to being met with 'off-the-shelf' content for all needs and pre-existing communities into which teachers must orientate and establish themselves.

Structure—The other point touched on in previous sections (e.g., Jamil and Hamre, 2018; Owston et al., 2008; Dede et al., 2016, Chapter 6) [51,53,55] is the use of structure to 'build-in' effective teacher education principles, including in this case opportunities for diagnostics (such as initial needs assessments, or implementation planning processes). Like with content, too much structure—particularly when used to maintain focus and mandate content (as opposed to design principles)—can reduce opportunities to tailor teacher education according to teachers' aspirations and starting points. Examples include self-study courses with set content. Similarly, our experience of MOOCs is that many have set content structured into weeks or modules, and provided by instructors, which are then 'softened' by the presence of community building features such as discussion boards. Equally, there are advantages to greater levels of structure, especially for initial teacher education, where it is possible to curate, sequence and ensure coverage of a core curriculum.

Assessment—Looking at the literature, great claims are made about the potential for personalized learning via technology in teacher education and education more generally. A widely advocated variant of this idea is that a combination of learning analytics (from simple scores to analysis through artificial intelligence and machine learning) and 'micro-credentials' or units can make even the most structured and content-heavy programs highly accessible and differentiated for individual needs. Dede et al. (2016) [53], for example discuss the use of micro-credentials in the form of 'digital badges':

> "A unique feature of digital badges is that they can easily be shared or aggregated, much like a professional portfolio, and the evidence behind the badge, since it is often also digital, can be examined to verify the value of the badge ... Technologies such as badges and analytics can help teachers map their personalized learning pathways, identify their progress along them and point to appropriate resources to support needed next steps."
>
> (Dede et al. 2016) [53] (pp. 44–45)

There are several points within this general description: first, that micro-structures and credentials may allow more bespoke teacher education activity, perhaps at a cost of collaboration and guided support. Second, that multimedia technology can be used to connect and make accessible evidence relating to the learning to foster discussion and/or as a summative assessment. Third, micro-structures may indulge the concept of gamification. This is not something we discuss at any length here, but note that points, badges, levels and leader-boards are becoming ubiquitous features of many online spaces and are widely

held to be beneficial features of learning design to encourage engagement and performance (Subhash and Cudney, 2018) [69].

A point raised in Qian et al. (2018) [70] worth noting here—one to which we return in relation to our discussion of curriculum and expertise below—is that the starting points of teachers in relation to subject knowledge and experiences within a topic area can influence their engagement with PD. Qian et al. (2018, p. 164) [70] found that novice, but subject-specialist teachers focused on material to develop their pedagogical content knowledge (PCK), non-subject-specialists focused on the content knowledge, and experience teachers were less likely to engage, believing that they had little need for the PD. This is an interesting point of contact between this section on differentiation for teachers, earlier discussions about the need for more experienced teachers to dial into and re-examine their extensive (and often tacit) expertise and experience, and the need for careful consideration of curriculum and employment of expertise.

One final point to note regarding teacher starting points not encountered within face-to-face teacher education is that teachers may be at different starting points with regards to their attitudes towards, experience of and skill with using technology and/or conducting teacher education remotely. This is born out in the available evidence around teacher satisfaction with online teacher education:

> "Research indicates that course management systems are effective for those familiar with technology, but that user satisfaction diminishes as user familiarity with the technology declines."
>
> (Elliot 2017, p. 121) [27] (Taylor, 2011; Tella, 2011; Thomas, 2010) [71–73]

4.4. Theory and Practice, Experimentation and Contextualisation

The next principle of effective teacher education relates to connecting theory with practice, and the reflection, and practical 'experimentation' that is needed to achieve this. This is often described in terms of active and contextualized learning (e.g., Darling-Hammond, 2017, p. v–vi) [21]. Again, rather than the details of this, our concern is the extent to which this can take place through remote or blended teacher education approaches. While the focus has been more on reflections than iterative development of practice as part of or alongside teacher education, the literature was broadly positive (Surrette and Johnson, 2015, p. 264; Lantz-Andersson's et al., 2018) [59,63]. Lantz-Andersson's et al. (2018) [59] review of formal and informal professional learning groups for example identifies numerous examples of groups that fostered reflection on practice. In some cases, it was felt that writing (rather than saying) responses 'was also seen to stimulate forms of reflection and self-analysis not evident in face-to-face meetings (Unwin, 2015) [74]' (p. 308). Moreover, links were found between online and face-to-face interaction for communities that had both an online and physical presence:

> "For teachers working together in the same schools, increased visibility of colleagues' beliefs, thoughts, problems and evolving practices online created entry points for subsequent face-to-face dialogue."
>
> (Lantz-Andersson et al., 2018, p. 308) [59]

Picking up themes noted in our earlier sections, Surrette and Johnson (2015) [63] found studies noting the benefits of being able to 'rewind and review the video and record more complete and accurate comments than is possible during real-time observations' (West et al., 2009, p. 384, quoted in Surrette and Johnson, 2015, p. 263) [63,75]. Using video, there is also a greater opportunity to conduct the reflections and discussions collaboratively, drawing on a greater number of examples (Perry et al., 2020) [35]. Other authors such as Keengwe and Kang (2013) [64] conclude that technological tools are important for integrating online and face-to-face learning and supporting learning to be put into practice, and also—as above—discuss the value of using technology to allow teachers to 'actively participate, communicate, and create their own materials' (p. 488). Tools discussed in the studies used by Keengwe and Kang (2013) [64] include multimedia CD-ROM, 'blackboard', web-based

models and sites, wikis, online discussion and lectures, interactive white boards and Ning, a community building platform. We are also reminded of discussions of tools for teachers in Timperley et al. (2007) [76] and of 'smart tools' in Robinson et al. (2007) [77].

A key idea here is that (a) to combine theory and discussion, expert models of practice, and one's own practice requires focus and experimentation over time (see points around duration, below) and (b) sustaining this within teacher education benefits from artefacts, tools and technologies that bring practice and pupil learning (see pupil orientation section) together and across classroom and professional education spaces. This is as true for face-to-face as online/remote teacher education and, arguably, the latter has the affordance of a greater range of technology at one's fingertips to be able to achieve this. Related, and discussed further in the section below about sustained activity, is that the asynchronous teacher education approaches and content allows professional learning to run alongside and shape to classroom activity over time. Returning to the point made above, there is no guarantee that these exchanges lead to deep and critical reflection and discussion and—alongside the positive examples noted above—reviews provide numerous counter-examples:

> "While these benefits were mostly intrapersonal in nature, studies did note a lack of critical discussion in the exchanges between teachers (e.g., Chen et al., 2009; Donnelly, 2010; Jarosewich et al., 2010; Prestridge, 2009) [78–81]. For example, participants were found to seldom challenge peers or engage in higher levels of analysis or reflection. When reviewing teachers' interactions in discussion forums, for example, the vast majority of responses to others' comments were found to be supportive without addressing the content of the course or furthering the discussion (Jarosewich et al., 2010) [80]"
>
> (Lantz-Andersson et al., 2018, p. 308) [59]

4.5. Curriculum and Specialist Expertise

Here we consider a set of issues around how easy it is to incorporate expertise and create and sequence curriculum content within remote and blended teacher learning. When considering initial teacher education specifically, greater attention is required to the foundational curriculum and whether this 'covers' the fundamental skills and knowledge for teaching within the chosen subject area and educational context, as well as meets statutory duties, thereby providing a strong basis for ongoing early-career development and beyond. Moreover, there is a need to develop subject knowledge, which varies considerably by subject area and by phase (e.g., primary and secondary phases) (Cordingley et al., 2018, p. 19) [19]. Bringing this to our central question: to what extent is a subject-specific curriculum incorporating expert content and support possible in remote and blended modes of teacher education? Several responses to this question are apparent in the literature: first, authors note that online modes and technology can be used to bring together larger groups, which are more likely to incorporate and/or find it more economical to draw on specialist expertise (including top scientists, public figures and outstanding educational leaders and practitioners etc.), as well as develop subject-specific content. With larger-group modes such as MOOCs this is further accentuated, although Dede et al. (2016) [53] note that for MOOCs, splitting up by subject specialism sometimes reduces opportunities for cross-pollination of ideas and connection, which can come by splitting large groups by roles, interests and geography (e.g., when trying to establish a blended community in a local area). At the extreme, as they suggest, it is possible and, in many ways, beneficial to make group splits even more arbitrary, splitting by the first letter of the participant's city for example.

Overall, there seems to be a strong case that online teacher education, particularly when built around platforms and/or resource bases, and a strong curriculum content offer with specialist input is eminently possible, and indeed has many advantages over exclusively face-to-face approaches (including the ability of teachers to locate specific (e.g., 'just-in-time') content, fitted to their needs, and work through and engage with it

individually or collectively in their own time. This connects to the above discussion around content and structure for differentiation, also to connecting theory and practice in their classroom. This is a familiar theme, which is echoed, as elsewhere, when our reviews come to consider curriculum content in online or blended teacher education, with both positive and negative examples discussed, such as points from Surrette and Johnson (2015, pp. 262, 264, 266) [63] about content that is 'disconnected with their classroom practices' or 'are not aligned with district and state standards and assessments'.

4.6. Duration and Organization

The final teacher education principle addressed in this section is the value of approaches that are sustained over time and thereby enable fulfilment of earlier principles around meaningful connections with classroom practice and the need to gradually incorporate new thinking and practices into teaching repertoires. We hasten to add, however, that there are examples of more short-term teacher education programs, which when well-targeted and tightly specified can have significant and lasting impact (see Lauer et al., 2014, for a review focused on this specific question) [82]. Be that as it may, effective teacher education typically is sustained over time and establishes a rhythm of activities that give teachers time to practice, reflect on and embed techniques in their classrooms (Cordingley et al., 2015; Darling-Hammond, 2017) [18,21].

We note that the additional flexibility around timings (including remote and asynchronous participation) afforded in remote and blended learning approaches, combined with the use of technology and pedagogical/reflection 'smart' tools (see earlier) is a supportive factor for practicably sustaining teacher education over time. In this section, now this has been noted, we move on to discuss the question of flexibility versus structure for remote and blended teacher education design. Often the structure for professional learning is provided by a school curriculum, such as the AP Physics programme discussed above, or (and especially in the context of initial teacher education) a set of teacher standards (Elliott, 2017) [27]. There is an extended discussion of 'flexible design' for remote initial teacher education in Burns (2011, p. 138–139) [10], and how this can be accompanied with 'flexible delivery', where instructors look to tailor the programme to learner needs (see discussion of differentiation, above). Flexibility is frequently linked with convenience and the ability to tailor learning to meet individual teacher needs. For programs tightly structured around particular curricula, along with the benefits discussed above in terms of 'just-in-time' learning came practical problems of timing. As discussed in Owston et al. (2008) [55], teachers were often not teaching the same topics at the same time and a degree of alteration to teachers' plans was needed to fit in and align activity. As a general rule, flexibility around timings and online approaches helps teachers fit learning around their professional commitments and school timetables, avoid the need to use weekend days for group activity and better fit with personal circumstances. The only negative point in this area that we found in reviews was the concern in Lantz-Andersson et al. (2018, p. 312) [59] that 'flexibility also introduced a risk of 'never-ending' engagement, and online professional learning encroaching into teachers' non-work time and becoming an additional source of over-work'.

5. Review of Empirical Evidence

Question 3 (Q3). *What empirical evidence is available that supports, refutes and/or refines our understanding of effective remote or blended teacher education (as per Question 2)?*

a. Which forms or practices are identified as more/less effective?
b. What are the common features of effective approaches?

This section reports results from 24 empirical studies that met our inclusion criteria. The first 19 present empirical results (including evidence of impact or otherwise on pupils) of remote and/or blended teacher education programs. The final 5 include consideration of more than one mode (remote, blended and/or face to face) and thereby enable a

form of comparison between modes. We summarize these 24 pieces within three broad groups. We also summarize results in relevant sections from one recent systematic review (Snell et al., 2019) [28], which presents a systematic review of 'technology-mediated language and literacy PD models' for teachers of early years (reception/kindergarten and pre-school) children. This review focused on random or quasi-randomized experimental studies, finding 11 in total.

5.1. The Efficacy of Remote and Blended Teacher Education

There are a group of studies that have evaluated remote or blended teacher education programs, measuring impacts on teachers and students. They are reviewed in this section. These do not seek to compare modality; indeed, for some, the fact that a remote or blended mode is used is incidental and raised only in specific results or discussion points. The studies reviewed below, provide us with some of examples of remote or blended programs that have been evaluated, and evidence of varying quality about whether they work or not.

5.1.1. Coaching and Mentoring Interventions

The evidence base for blended or remote coaching/mentoring of teachers was limited; however, there were a few high-quality, highly relevant studies. Allen et al. (2015) [83] present a study based on a stratified randomized controlled trial of the *My Teaching Partner–Secondary (MTP-S)* program, a web-mediated coaching program involving 86 secondary school teachers and 1194 secondary students. The program was conducted over 2 years involving 12 coaching cycles, where coaches selected extracts from the teachers' own lessons that they had videoed for the teachers to reflect on their pedagogical practices before discussing by telephone (Allen et al. 2015) [83]. Students' scores on the Virginia state standardized tests rose significantly (Effect Size = 0.48) compared to controls.

In contrast, Kraft and Hill (2020) [84] reported on a web-based coaching programme designed to improve teaching of mathematics in relation to a *Mathematic Quality of Instruction (MQI)* framework using a 'MQI Coaching' model involving 'Cycles' of coaching using the MQI framework. The MQI is a well-validated instrument which provides a theory of action for the programme and a measure of specific aspects of teaching to target:

> "A central element of our theory of action was that if teachers could learn to analyze instruction using the MQI's items and practice descriptors, they would (a) use more of the practices contained in those items and (b) strengthen the quality of these practices as implemented in their classrooms."
>
> (Kraft and Hill, 2020. p. 5) [84]

The program involved 142 elementary and middle school teachers from 51 schools, supported by 24 expert MQI coaches (all with experience with the MQI, as classroom teachers and/or instructional coaches). Their randomized field trail indeed showed improved teaching using the MQI instrument and a higher likelihood of teacher retention for teaching maths, and in particular teaching maths in a grade (year group) with a high-stakes maths examination. Student gains (as effect sizes) however were estimated to be minimal, with estimated effects of 0.00 (2 dp) on state tests, and 0.07 on a Measures of Academic Progress (MAP) test, neither of which were statistically significant. There were not clear reasons why student gains were not achieved/evidenced: the authors suggest issues of low statistical power, issues with standardized tests capturing all student learning and/or lack of program efficacy.

Snell et al. (2019) [28], the systematic review included in this section, reviewed evidence for various professional development programmes in language and literacy for teachers in early years settings. There were several areas of reporting:

- 'Remote, non-live, asynchronous coaching'—this included further studies of the My Teaching Partner (MTP) programme (as per Allen et al., 2015, above) [83]. Results were mixed, with Pianta et al. (2017) [85] and Mashburn et al. (2010) [86] finding

mixed results on student learning outcomes, and Early et al. (2017) [87] not finding an impact on language or literacy outcomes.

- 'Remote live coaching' programmes, finding only one effective trial: Vernon-Feagans et al. (2015, 2013) [88,89] used remote live coaching with webcams and found effect sizes around 0.5 for four measures of reading ability for early years. Note that these studies also included other interventions, such as online materials and even a 3-day workshop for this last study.
- 'Enhancing' Tech PD with in-person coaching', where Landry et al. (2009) [90] found that in-person coaching i.e., a blended approach improved pre-school children's reading skills with effect sizes around 0.2 (CIRCLE study). But Jackson et al. (2006) [91] found mixed effects.

Our final study in this section is Rosenberg et al. (2020) [92], who reported on individual coaching for paraprofessionals (teaching assistants) about self-advocacy for students with disabilities. An initial 45-min individual teaching session was followed by distance bug-in-ear (BIE) coaching via the Zoom platform. They report improved student behavior but there were just four adult-student pairs and no controls. Nonetheless, it is interesting to note the study's use of advancing technology in the form of video conferencing and 'bug-in-ear' coaching tools.

5.1.2. Mixed Component Interventions

Having looked at mentoring and coaching interventions separately, we now turn our attention to other programs, noting the variety of approaches in use. The following studies are mostly reporting programs that combine instructional content and sessions, which deliver this with wrap-around teacher discussion and/or implementation tasks to promote application to classrooms. Given the relative small number of studies, we provide a short narrative summary of each study, providing an italicized summary of the main finding for each.

Ascetta, Harn & Durán (2019) [93] reported on a brief (4–6 weeks, $\leq$1 h per week) online program that focused on: (a) teachers' self-monitoring and goal setting; and (b) language enhancement strategies to use in practice. They were interested in the impact of the type of feedback strategies used with the teachers, who were then randomly assigned to the type of feedback they received across the intervention: performance or self-report.

Finding: Regardless of feedback condition, all participating teachers increased their frequency of language enhancement strategy use, and all children had enhanced vocabulary skills performance.

Dash et al. (2012) [94] examined the impact of participation in online PD on 5th grade mathematics teaching. The PD consisted of 1 course per semester over three semesters, equating to approximately 70 h work per teacher. The delivery was asynchronous, with trained facilitators to monitor the completion of activities. It utilized a learning community model including readings, web-based resources, interactive on and offline activities, video, and peer-to-peer discussions. Teachers were required to implement a final project in their classrooms.

Finding: Participants in the PD scored significantly higher overall for pedagogical content knowledge than teachers in the control group and had greater gains in scores for overall pedagogical practices. However, no significant differences in changes in students' maths achievement were identified.

Our next study is Fishman et al., (2014b) [95] who report results from a survey of 6796 (Advanced Placement Curriculum) Biology teachers, which are matched to subsequent student scores. Teachers were provided with 'a range of professional development (PD) options, from week-long summer workshops to short face-to-face (FtF) courses, online self-paced courses, downloadable resources, and online peer-learning communities' and 'two high-quality PD offerings from outside providers included in this study.' (p. 3) There

are limitations in methodology, including the inevitable self-selection in the extent to which teachers engage with the professional development and also complete the survey.

Finding: Overall, they find (p. 15) that 'some of the patterns of PD selected by teachers were predictive of student outcomes on the AP Biology exam, although effect sizes were small'.

Frumin et al. (2018) [61] undertook a 5-year longitudinal study of the participation of US high school science teachers in an online teacher community supporting curriculum change. The study focused on teachers' patterns of usage, the impact of usage on the performance of their students, and how the PLC might harmonize with established programmes of PD.

Finding: a clear association was found between teachers' participation and higher student examination scores, though causality could not be assumed. Interestingly, the frequency and duration of participation did not have a clear association with test scores. This project built on the earlier study by Fishman et al. (2014b) [95], as immediately above.

Goldenberg et al. (2014) [66] note the challenge of convening and supporting role-alike groups of teachers to provide specific training. They implemented a randomized control trial that focused on PD for Biology teachers to improve content knowledge, pedagogical knowledge and use of digital media in teaching. *Teaching High School Biology* was a fully online programme that included 4 sessions on teaching genetics and evolution using digital resources and inquiry-based approaches, utilizing readings, video clips, notebook work, discussion boards, and assignments.

Finding: The study found significant teacher knowledge increases despite high (50%) teacher attrition rates, albeit no effect on student outcomes—perhaps because of difficulties measuring pupil outcomes from inquiry-based learning.

Griffin et al. (2018) [96] studied 32 elementary maths teachers for children with disabilities in the US participating in a PD programme called Prime Online. This consisted of 35 Moodle modules over 12 months with clear taught content, plus online discussions etc. with a focus on teacher beliefs (about maths teaching, students with disabilities and teacher inquiry) and teachers' mathematical knowledge.

Finding: Positive teacher impacts were found, but there was no impact on the students. There were no control conditions, and it is not clear which aspects of the programme were beneficial.

Steve Shaha has been involved in several related studies in this section. Shaha and Ellsworth (2013) [97] compared comparative gains (from one year to the next) in mathematics and reading standardized assessments for school with teachers with high levels of engagement with *PD 360*, which 'offers a large library of expert-produced training videos, powerful support tools and resources, and an online professional learning community of nearly one million educators' (p. 176). 169 schools were selected and compared with average gains for their Districts.

Finding: Students in these schools experienced 18.9% increases in mathematics scores versus 4.2% for their districts, and 15.3% increases in reading versus 2.5% for their districts. Shaha, Glassett and Copas (2015b) [98] showed that these gains were further extended with an additional follow-up year. Similarly, Shaha, Glassett, Ellsworth (2015d) [99] suggest the greater the number of years of high participation in the programme, the bigger the effect (also see Shaha et al., 2015a; 2015c; 2016) [100–102]. With all of these studies, without randomization to receive the program, these gains may reflect teacher self-selection into their degree of engagement with the programs and/or general school characteristics rather than PD 360 per se.

Walker et al. (2012) [103] looked at 2 designs for online PD to help teachers learn to create online activities for students. One focused solely on technology skills to design student activities using online resources, while the other coupled technology skills with learning to design PBL activities for their students using online resources. The quasi-experimental study involved 36 mathematics and science junior high school teachers (51 initially signed up) and 1247 students over a sustained period of 3 months.

Finding: Both designs showed significant gains in 'self-reported knowledge, skills, and technology integration' for teachers. There were also 'self-reported gains in behaviour, knowledge, and attitudes' for all students, though those taught by PBL teachers scored greater gains on behavior and knowledge. PBL teachers had larger gains in PBL knowledge and use than the tech-only teachers. Limitations included the fact that teachers selected the class to administer the student survey; students may have provided socially desirable responses and there was potential for pre-existing differences between treatment groups, such as prior exposure to PBL.

The final study reported in this sub-section is Weschke, Barclay & Vandersall (2011) [104], who undertook a small-scale study of 31 elementary teachers who completed a fully online master's programme focused on pedagogy and literacy matched with a comparative sample of 39 teachers with an existing masters level qualification who did not complete the online programme. The study looked at the impacts on oral reading fluency of students as a measure of efficacy by undertaking 3828 student observations. It was an attempt to establish the impact on the education community of a high demand and fully online programme.

Finding: Hierarchical linear modelling was used to understand the longitudinal effects of teachers on student learning. The findings indicated a significant positive effect of the online program on student achievement. However, the results were not constant across different elementary grades. The study suggests more research needs to be conducted on why student outcomes resulting from online PD vary according to age and stage.

5.1.3. Interventions for Student Behaviour

There were two studies that focused on student behavior that met our general eligibility criteria, but these are different in kind to other studies, so are not reported here (Marquez et al., (2016); Hoffman et al., 2020).

5.2. Comparing the Effectiveness of Teacher Education Modes

In this section we review five studies, a meta-analysis (Hill et al., 2020) [5] and a systematic review (Snell et al., 2019), all of which present results that allow comparison of the effectiveness of teacher education modes or combinations of them.

An early study in this area is Fisher et al. (2010) [105] who compare a virtual workshop with an actual workshop, both focused on developing a Concept Mastery Routine (CMR). The study recruited 59 certified teachers and randomly allocated them to one of the two groups. The virtual workshop (VW) made use of a school's computer 'lab' and a multimedia software programme. The face-to-face session (the 'actual workshop'—AW) in the classroom was set up with 25 to 30 seats in rows. To ensure comparability in other respects, both VW and AW had an instructor's computer, data projector, a screen and whiteboards; the AW translated this material into PowerPoints and other documents, and made use of the video clips form the multi-media CD in the session.

Finding: Comparisons of pre and post test data show that there is no difference in student tests of concept knowledge between the students of teachers who participated in the face to face or virtual PD, although teachers expressed slightly more satisfaction from face-to-face PD.

Another early study, one that received considerable attention in the literature (see discussion section) is Fishman et al. (2013) [106]. This was a randomized experiment involving a total of 49 teachers that explicitly set out to compare online and face to face professional development. The PD aims to prepare high school teachers to implement a year-long environmental science curriculum, increasing the probability that curriculum enactment is consistent with original design and outcomes. Teachers either received a week-long (48 h) course spread over 6 days, consisting of face-to-face workshop sessions or an online alternative 'to be completed by teachers asynchronously at their own pace', consisting of a series of 'short courses with a facilitator guiding teachers and answering questions as they worked through the materials' (pp. 429–430).

Finding: Fishman et al. (2013) [106] find increases in subject knowledge for both online and face-to-face groups, with little difference in the gains between the two (controlling for pre-test scores) and no differences between the conditions in relation to a range of teacher attitudinal measures. In relation to student outcomes on an environmental science scale, the students of teachers in the online condition improved by just over half a standard deviation, the students in the face-to-face condition, one smaller than half a standard deviation, but the difference between these was not statistically significant. One especially notably finding within the study was the lack of a relationship between 'contact hours' and student performance in the online condition.

> "Because we found no relationship between amount of time teachers spent on online PD and our outcome measure, we conclude that teachers who completed online PD more rapidly required less time to benefit from materials. Online PD does not have to be 'one size fits all' in terms of participation."
>
> (Fishman et al., 2013, p. 435) [106]

The next study we examine is Rock (2017) [107]. This doctoral thesis presents results retrospectively comparing student assessment scores for students of teachers who participated in online or face to face state-grant-funded professional development. The study looks at whether the mode of PD delivery was associated with student outcomes of pupils taught by the sample of 216 elementary teachers and 216 high school teachers. In overview, 105 of these teachers engaged in F2F PD and 327 took part in online sessions.

Finding: The results found no statistically significant difference between the two modes of PD in terms of student impact, concluding that, 'when the content of professional development is comparable, the form of professional development and the level of instruction have minimal effect on student learning outcomes.' (p. 186)

Webb, Nickerson and Bush (2017) [108] present comparative analysis of student survey responses (n = 4832) between teachers (n = 75, K-12) who undertook different modes of professional development in the context of a computer science (game design) curriculum.

Finding: There were very little difference across the survey items when comparing students of teachers who received professional development in different modes. The authors do make a helpful distinction between Massive, Open, Online Courses (MOOCs) and managed online PD with smaller groups. They state that their completion rate (60%) is higher than that found in typical MOOCs and accredit this to the ability to develop a sense of community and provide facilitator support in smaller groups.

The final individual study reported in this section is Russell et al. (2009) [109]. The study compared the effects of a professional development course for teachers of maths in grades one to five in an online and a face-to-face format. Participants were randomly assigned to one of the conditions, both delivered over the same eight-week period (see Russell et al., 2009, pp. 74–75, for further details of the intervention) [109].

Finding: Using a range of data sources, the authors find that, 'both formats of the course showed significant impacts on teachers' mathematical understanding, pedagogical beliefs, and instructional practices' (p. 71). Student outcome data were minimal, although enough to meet our eligibility criteria, and were only collected through a student survey. This student survey asked about instructional practices of the teachers and their quality; no statistically significant differences were found between online and face-to-face groups.

Finally, we review results from Hill et al.'s (2020) [5] meta-analysis of STEM PD and Snell et al. (2019) [28]. We only provide a single result from Hill et al. (2020) [5] as the report is focused on CPD in general—but this result is a significant one. They present meta-analysis of the effect of STEM professional learning programs according to various PD characteristics relating to its development, type, focus and—pertinent for our review—format. Regarding the latter, standardized effects from a total of 95 studies of professional development (including six which provided curriculum materials) were summarised in terms of the average percentile rank for intervention students compared to control/comparison students in the 50th percentile.

Finding: Their results on program format, looking across the 95 studies within the meta-analysis suggest that 'Any Online PD' (i.e., not entirely face-to-face) tends to have positive, but lower effects than other formats, with about half the overall increase (see Hill et al., 2020, Figure 4, p. 54) [5]. Whether this is reflective of the quality of the online programs or is something related to the mode per se is difficult to determine from this brief and 'high-level' view meta-analysis. A fair comparison between modes would deliver the same content through more than one mode, rather than programs that 'naturally' do or do not include an online element (which may be of a different nature or quality for reasons other than the mode).

Finally, we examine the systematic review reported in Snell et al. (2019) [28]. This reviewed evidence for various professional development programs in language and literacy for teachers in early years settings. They report studies from across our remote and blended approaches (Q1, above). Here we focus only on results that compare CPD modes or difference remote and blended approaches.

Finding: In overview, Snell et al. (2019) [28] conclude their review as follows:

> "These results also showed that Tech PD programs can be as effective as in-person PD programs, with effect sizes in the moderate to large range for impacts on teaching practices (i.e., Cohen's d > 0.5) and in the small to moderate range for impacts on student learning (i.e., Cohen's d = 0.2–0.5). These studies may also suggest that the PD content was well developed and tested and that the technology was successfully used as method of dissemination . . . In most of the Tech PD studies, not all measured outcomes showed positive results. Sometimes the effects were similar between in-person PD and Tech PD, and sometimes this similarity was reflected in null effects for either condition."
>
> (Snell et al., 2019, p. 216) [28]

6. Conclusions

6.1. Main Findings and Contributions of This Review

In each of the sections above, we have summarized and drawn conclusions about the main findings arising from our analysis. Here we consider the more general question of whether remote and online teacher education can be considered to be effective, based on theory and evidence, and—if so—in what respects it is likely to be superior or inferior to face-to-face teacher education. Given the limited robust evidence to base this answer on (see Seciton 5), our conclusion is inevitably only tentative and general. We also, therefore, briefly comment on our present understanding, in terms of theory and evidence, of effective remote and blended teacher education.

Question 4 (Q4). *Implications and Unknowns—What, if any, general principles for effective online or blended teacher education are evident?*

a. What does the evidence suggest about the relative effectiveness of teacher education modalities (including in comparison with the face-to-face mode)?
b. What are the main areas of uncertainty and gaps in our present understanding?

6.2. Relative Effectiveness of Teacher Education Modality

Remote and blended teacher education is a relatively new field of practice and study. There are few studies that enable firm conclusions to be drawn on the relative effectiveness of modes and approaches. We therefore return to our discussion of the question itself that we began near the opening of the Approaches and Limitations (Q2) section, where we asked whether the focus should be on the 'medium' (i.e., the mode and approaches to remote and blended teacher education) or the 'message' (i.e., the quality of the teacher education itself with respect to research-informed principles). We concluded that—while there is value in examining the former (especially given recent technological advances and

the present circumstances surrounding the COVID-19 pandemic), we must not lose sight of the latter.

Having now reviewed the literature and evidence, we believe that this position still stands. We also note a similar reticence to draw conclusions about modes *per se*—given too little systematic testing and the large variation in the nature and quality of remote and blended programs—is also evident in the literature:

> "We do not believe that the PD described in this study, whether online or face-to-face, is representative of all PD. Yet there is a tendency, especially among educational decision makers, to treat 'online' PD as if it represents a particular approach, rather than a delivery vehicle. There is a need for studies that examine many different designs for PD: 'online' and 'face-to-face' are no more descriptive in the end than 'workshop' in terms of understanding the nature of the teaching and learning activities contained within the PD."
>
> (Fishman et al., 2013, p. 436) [106]

We agree. While we hold that there is value in examining 'delivery vehicles' and their respective affordances and limitations, it is likely—especially when considering blended or multi-approach teacher education and the growing ubiquity of technology (especially video for conferencing and lesson observation)—that differences in medium will not be the main consideration for teacher education quality: programs will tend towards a blend of media and approaches and their distinguishing quality factors will relate to teacher education principles and whether these have been achieved rather than the mode.

That said, let us address the question nonetheless: what *does* the evidence suggest about the relative effectiveness of teacher education modes? The few studies we have to go on (i.e., which allow fair comparison between similar content in different modes or combination of approaches) suggest that there is little difference in effectiveness (Fisher et al., 2010; Fishman et al., 2013; Rock, 2017; Webb et al., 2017; Russell et al., 2009) [105–109]. This includes studies presenting results from randomized experimental designs assessing impact on teachers' student's attainment (Fisher et al., 2010; Fishman et al., 2013) [105,106]. In the closest systematic review to our own, focused on language and literacy PD in the early years, Snell et al. (2019, p. 216) [28] conclude 'that Tech PD programs can be as effective as in-person PD programs'. There are tantalizing findings about combining components such as coaching and mentoring with video lesson observations, curriculum materials and/or CPD—but with such a limited evidence base, drawing conclusions would be over-reaching.

6.3. Design Principles for Teacher Education Modality

Anticipating limitations in the evidence, our approach has included more theory-rich and exploratory sections where we have drawn on reviews and selected empirical pieces to explore what effective remote and blended teacher education might look like in terms of its use of technology and learning design. As we first noted in the introduction, back in 2014, researchers (e.g., see Fishman et al., 2014a; Moon et al., 2014) [15,95] were in agreement that the field would benefit from scholarship that leads to 'research-based design principles to guide the ongoing development, implementation, and evaluation efforts in online PD'. Six years on, we have been able to report on a larger, but still limited literature. Even in the most highly researched areas, such as around professional learning communities (see Lantz-Andersson et al., 2018) [59], more research is needed. A similar appraisal of the literature on online ITE is provided by Dyment and Downing (2020) [11], as follows:

> " ... the present research base is fragmented and consists of disparate and unrelated studies that rarely cross reference each other. Given that no particular author is dominating the publishing landscape for ITE, we suspect that, for most authors, these articles represent scholarship of their own teaching and learning, as they report through scholarly publications on a particular innovation or trend in their practice. We do not wish to be overly critical of this fragmented

> approach to research; we feel confident that these studies would nevertheless have had a small sphere of influence on the practices of the teacher educators who performed the research, also on their local colleagues and, hopefully, other readers of the journal . . . We believe that there is an urgent need for a more strategic, cohesive, and collaborative approach to researching online ITE. Researchers must be encouraged to work together to establish a shared research agenda that tackles the 'harder-to-reach' variety of research questions. This will require leadership, coordination, and collaboration. If successful, this would allow interdisciplinary teams, with diverse research experience and cross-cultural expertise, to develop broader and more relevant insights than is possible when individual researchers simply profile their own 'innovation'—again and again . . . "
>
> (Dyment and Downing, 2020, p. 329–330) [11]

There were, however, notable exceptions: we did find strong examples which provide the rigorous evidence and theoretical depth needed for progress, including Surrette and Johnson's (2015) [63] meta-analysis which concludes favorably about the question of whether online environments have the ability to facilitate key professional development principles; Dyment and Downing's (2020) [11] systematic review of the literature on online ITE; the systematic review from Snell et al. (2019) [28] identifying 11 experimental or quasi-experimental studies; and well-designed evaluation studies including Kraft and Hill (2020) [84] and Allen et al. (2015) [83]; and several other pieces which we discussed in previous sections. It is on such reviews and studies that conclusions about effectiveness can start to be founded. We hope we have added to this through our blend of theory application, review synthesis and systematic evidence assessment.

The affordances and limitations section is replete with principles and considerations for the design of remote and blended teacher education. This, however, is almost exclusively based on perspectives from across the literature coupled with interpretations of our own, which enjoy varying levels of evidential support—support that becomes more tenuous as one takes the principle further from the study context and focus from which it originated. Of particular concern when it comes to evidential support was the lack of studies that reported student outcomes (in any form whatsoever). Given that a pupil orientation is a feature of effective professional learning, this lack of attention to how teacher education affects pupils impairs our ability to assess both process and impact of teacher education programmes. Evaluating the impact of teacher education on pupils is central to establishing evidence-based principles for effective policy, programmes and practice.

The set of principles, affordances and limitations described in this report, therefore, can be seen as both a statement from the literature of *working principles for online or blended teacher education design* at the same time as *a set of hypotheses and research questions* that are yet to be established and need to be tested and developed through teacher education research and practice. So, while many plausible principles are apparent, we simply do not have firm evidence back these up at present. However, we do think that the literature suggests that remote and blended teacher education show considerable promise; that they have distinct advantages and disadvantages relative to solely face-to-face approaches; and that they already are and are likely to increasingly become important parts of the teacher education landscape.

Supplementary Materials: The following are available online at https://www.mdpi.com/2227-7102/11/8/453/s1, References organized by review database group.

Author Contributions: Conceptualization, T.P., M.F. and P.C.; methodology, T.P. and P.C.; formal analysis, T.P., M.F. and P.C.; data curation, T.P.; writing—original draft preparation, T.P., M.F. and P.C.; writing—review and editing, T.P., M.F. and P.C.; project administration and funding acquisition, T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by STEM Learning, commissioned through University of Birmingham Enterprise Limited.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank STEM Learning for commissioning this review, and for their support with review scoping, screening and write up, colleagues in the Department of Teacher Education, University of Birmingham, and Centre for the Use of Research and Evidence in Education (CUREE), UK.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Search Term Record

Database	Search String/Details	Results
Web of Science (10 collections)	(TI = (teach* OR educat* OR profession*) AND AB = (educat*OR learn* OR CPD OR train* OR develop*) AND TI = (review OR "meta-analysis" OR synthesis OR Evaluation OR Trial OR Impact OR effect) AND AB = (modality OR mode OR online OR "face-to-face" OR blended OR remote OR virtual OR distance)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan = 2005–2020	3314 initial 1400 after exclusion by WoS categories
Proquest (9 databases see record)	TI: (teach* OR educat* OR profession*) AND TI: (educat* OR learn* OR CPD OR train* OR develop*) AND TI: (modality OR mode OR online OR "face-to-face" OR blended OR remote OR virtual OR distance) AND AB: (review OR "meta-analysis" OR synthesis OR Evaluation OR Trial OR Impact OR effect) Language: English Date: 2010 onwards	3638 results
Supplementary Searches		
Google Scholar	allintitle: (teacher OR educator OR practitioner) (modality OR mode OR online OR "face-to-face" OR blended OR remote OR virtual OR distance) Date: 2005+	6390 results (Limited to first 1000 results)
Google Scholar	allintitle: (teacher OR educator OR practitioner) (modality OR mode OR online OR "face-to-face" OR blended OR remote OR virtual OR distance) (trial OR RCT OR random OR experiment OR randomised OR quasi-experiment)	24 results (19 retrieved)
Google Scholar	'Pearl growing' using citation data for 15 relevant papers	839 results (64 retrieved)
Ingenta Connect	Title, Keywords or Abstract contains: teacher AND (development OR learning OR training) AND (modality OR mode OR online OR "face to face" OR blended OR remote OR virtual OR distance) AND (review OR "meta-analysis" OR synthesis OR Evaluation OR Trial OR Impact OR effect)	193 results (Initial screening on title—26 met criteria, 24 items retrieved)
JSTOR	(((ti:((teach* OR educat* OR profession*)) AND ti:((educat* OR learn* OR CPD OR train* OR develop*))) AND ti:((modality OR mode OR online OR "face-to-face" OR blended OR remote OR virtual OR distance))) Subject: Education Date: 2010 onwards	214 results (Initial screening on title—17 results selected; 16 results retrieved)

Appendix B. Inclusion and Exclusion Criteria

Area	Include/Exclude	When Applied?
Date	Include—Studies conducted in or after 2005	Searching
Learners	Include—All age 3–18 (Inc. SEN and mainstream)	Screening title and abstract
Teacher learning focus	Include—Substantial focus on teacher learning and/or improving pupil outcomes via teacher learning. There must be a teacher learning element. Include—Teachers defined as all subjects, all career states (qualified, unqualified), all statuses (including teaching assistants).	Screening title and abstract
Non-face-to-face element	Include—All empirical and theoretical pieces which present findings about a mode of teacher learning other than face-to-face. This can be a trial of an online or blended CPD programme. It would ideally include a comparison of face-to-face and another mode of teacher learning (but an evaluation of a single remote or blended mode is acceptable). Also Include—Reviews, meta-analyses or reports—if they are focused on effective CPDL or ITE and have potential implications (either due to sub-results or by inference from principles) for remote or blended teacher education.	Screening full text
Pupil outcomes	Include—All studies included in the main review must present some empirical evidence about the impact of the teacher learning on pupil outcomes. This is to be interpreted broadly and can include, e.g., classroom observations or other data collections reporting pupil behaviours, perspectives and/or learning outcomes.	Screening full text

Appendix C. Record of Search Results and Screening

Search Databases		
Web of Science	1400	7354 records, 5 databases, dozens of collections
Proquest	3638	
Google Scholar	1000	
Google Scholar	24	
Goole Scholar (pearl)	839	
Ingenta Connect	193	
JSTOR	214	
Other Additions	46	
Total Records	7354	

	Before	Exclusion										Retained
		Not 3–18	Not TD	No Remote PD Element	Duplicate	No Pupil Outcome	Pre-2005	Inaccessible	Not English	Total Exc		
After round 1—title screening	7354			5638						5638		1716
After round 2—title and abstract	1716	160	515	52						727		989
After round 3—full text	989	28	50	32	45	749	5	4	3	916		73
											Late Entries	
Reviews and Reports	25				6					6	3	22
Empirical Studies	25				2					2	1	24
Background and Wider	23				4					4	0	19
										Total		65

References

1. Slater, H.; Davies, N.M.; Burgess, S. Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxf. Bull. Econ. Stat.* **2012**, *74*, 629–645. [CrossRef]
2. Kane, T.J.; McCaffrey, D.F.; Miller, T.; Staiger, D.O. *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*; Bill & Melinda Gates Foundation: Seattle, WA, USA, 2013.
3. Fletcher-Wood, H.; Zuccollo, J. *The Effects of High-Quality Professional Development on Teachers and Students: A Rapid Review and Meta-Analysis*; Education Policy Institute: London, UK, 2020.
4. Kraft, M.A.; Papay, J.P. Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. *Educ. Eval. Policy Anal.* **2014**, *36*, 476–500. [CrossRef] [PubMed]
5. Hill, H.C.; Lynch, K.; Gonzalez, K.E.; Pollard, C. Professional development that improves STEM outcomes. *Phi Delta Kappan* **2020**, *101*, 50–56. [CrossRef]
6. Tschannen-Moran, M.; Hoy, A.W. Teacher efficacy: Capturing an elusive construct. *Teach. Teach. Educ.* **2001**, *17*, 783–805. [CrossRef]
7. Cordingley, P.; Crisp, B.; Johns, P.; Perry, T.; Campbell, C.; Bell, M.; Bradbury, M. Constructing Teachers' Professional Identities. Available online: https://issuu.com/educationinternational/docs/2019_ei_research_constructing_teach (accessed on 18 August 2021).
8. Hobbs, A.; Bolan, F. Distance Learning, POSTNote Parliamentary Research Briefing. Available online: https://post.parliament.uk/research-briefings/post-pn-0639/ (accessed on 18 August 2021).
9. La Velle, L.; Newman, S.; Montgomery, C.; Hyatt, D. Initial teacher education in England and the COVID-19 pandemic: Challenges and opportunities. *J. Educ. Teach.* **2020**, *46*, 596–608. [CrossRef]
10. Burns, M. *Distance Education for Teacher Training: Modes, Models, and Methods*; Education Development Center Inc.: Washington, DC, USA, 2011.
11. Dyment, J.E.; Downing, J.J. Online initial teacher education: A systematic review of the literature. *Asia-Pac. J. Teach. Educ.* **2019**, *48*, 316–333. [CrossRef]
12. Le Cornu, R. *Key Components of Effective Professional Experience in Initial Teacher Education in Australia*; Australian Institute for Teaching and School Leadership: Melbourne, Austrilia, 2015.
13. Dede, C.; Jass Ketelhut, D.; Whitehouse, P.; Breit, L.; McCloskey, E.M. A Research Agenda for Online Teacher Professional Development. *J. Teach. Educ.* **2008**, *60*, 8–19. [CrossRef]
14. Fishman, B.; Konstantopoulos, S.; Kubitskey, B.W.; Vath, R.; Park, G.; Johnson, H.; Edelson, D. The Future of Professional Development Will Be Designed, Not Discovered. *J. Teach. Educ.* **2014**, *65*, 261–264. [CrossRef]
15. Moon, J.; Passmore, C.; Reiser, B.J.; Michaels, S. Beyond comparisons of online versus face-to-face PD: Commentary in response to Fishman et al.,"Comparing the impact of online and face-to-face professional development in the context of curriculum implementation". *J. Teach. Educ.* **2014**, *65*, 172–176. [CrossRef]

16. Quinn, F.; Charteris, J.; Adlington, R.; Rizk, N.; Fletcher, P.; Reyes, V.; Parkes, M. Developing, situating and evaluating effective online professional learning and development: A review of some theoretical and policy frameworks. *Aust. Educ. Res.* **2019**, *46*, 405–424. [CrossRef]
17. Cordingley, P. The contribution of research to teachers' professional learning and development. *Oxf. Rev. Educ.* **2015**, *41*, 234–252. [CrossRef]
18. Cordingley, P.; Higgins, S.; Greany, T.; Buckler, N.; Coles-Jordan, D.; Crisp, B.; Saunders, L.; Coe, R. *Developing Great Teaching: Lessons from the International Reviews into Effective Professional Development*; Durham University: Durham, UK, 2015.
19. Cordingley, P.; Greany, T.; Crisp, B.; Seleznyov, S.; Bradbury, M.; Perry, T. *Developing Great Subject Teaching: Rapid Evidence Review of Subject-Specific Continuing Professional Development in the UK*; CUREE Limited: London, UK, 2018.
20. Cordingley, P.; Higgins, S.; Greany, T.; Crisp, B.; Araviaki, E.; Coe, R.; Johns, P. Developing Great Leadership of CPDL. Available online: http://www.curee.co.uk/files/publication/%5Bsite-timestamp%5D/Developing%20Great%20Leadership%20CPDL%20-%20final%20summary%20report.pdf. (accessed on 18 August 2021).
21. Darling-Hammond, L.; Hyler, M.E.; Gardner, M. *Effective Teacher Professional Development*; Learning Policy Institute: Palo Alto, CA, USA, 2017.
22. Rogers, S.; Brown, C.; Poblete, X. A systematic review of the evidence base for professional learning in early years education (The PLEYE Review). *Rev. Educ.* **2019**, *8*, 156–188. [CrossRef]
23. Hargreaves, D.H. A common-sense model of the professional development of teachers. In *Reconstructing Teacher Education: Teacher Development*; Elliot, J., Ed.; Falmer: London, UK, 1993.
24. DfE. ITT Core Content Framework. Available online: https://www.gov.uk/government/publications/initial-teacher-training-itt-core-content-framework (accessed on 18 August 2021).
25. Little, C.A.; Housand, B.C. Avenues to professional learning online: Technology tips and tools for professional development in gifted education. *Gift. Child Today* **2011**, *34*, 18–27. [CrossRef]
26. Bower, M. Deriving a typology of W eb 2.0 learning technologies. *Br. J. Educ. Technol.* **2016**, *47*, 763–777. [CrossRef]
27. Elliott, J.C. The Evolution From Traditional to Online Professional Development: A Review. *J. Digit. Learn. Teach. Educ.* **2017**, *33*, 114–125. [CrossRef]
28. Snell, E.K.; Hindman, A.H.; Wasik, B.A. A review of research on technology-mediated language and literacy professional development models. *J. Early Child. Teach. Educ.* **2019**, *40*, 205–220. [CrossRef]
29. Horn, M.B.; Staker, H. The Rise of K-12 Blended Learning. Available online: http://www.innosightinstitute.org/innosight/wp-content/uploads/2011/01/The-Rise-of-K-12-Blended-Learning.pdf (accessed on 18 August 2021).
30. Seddon, K.; Postlethwaite, K.; James, M.; Mulryne, K. Towards an understanding of the learning processes that occur in synchronous online seminars for the professional development of experienced educators. *Educ. Inf. Technol.* **2011**, *17*, 431–449. [CrossRef]
31. Kraft, M.A.; Blazar, D.; Hogan, D. The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Rev. Educ. Res.* **2018**, *88*, 547–588. [CrossRef]
32. National Framework for Mentoring and Coaching. Available online: http://www.curee.co.uk/resources/publications/national-framework-mentoring-and-coaching. (accessed on 18 August 2021).
33. Cordingley, P.; Buckler, N. *Professional Learning Environments in Primary and Secondary Contexts*; University of Manchester: Manchester, UK, 2012; pp. 11.00–12.30.
34. Fletcher, S.; Mullen, C.A. *Sage Handbook of Mentoring and Coaching in Education*; Sage: Thousand Oaks, CA, USA, 2012.
35. Perry, T.; Davies, P.; Brady, J. Using video clubs to develop teachers' thinking and practice in oral feedback and dialogic teaching. *Camb. J. Educ.* **2020**, *50*, 615–637. [CrossRef]
36. Brouwer, C. *Equipping Teachers Visually*; Stichting Kennisnet: Zoetermeer, The Netherlands, 2011.
37. Gaudin, C.; Chaliès, S. Video viewing in teacher education and professional development: A literature review. *Educ. Res. Rev.* **2015**, *16*, 41–67. [CrossRef]
38. Major, L.; Watson, S. Using video to support in-service teacher professional development: The state of the field, limitations and possibilities. *Technol. Pedagog. Educ.* **2018**, *27*, 49–68. [CrossRef]
39. Brouwer, N.; Besselink, E.; Oosterheert, I. The power of video feedback with structured viewing guides. *Teach. Teach. Educ.* **2017**, *66*, 60–73. [CrossRef]
40. Pehmer, A.-K.; Gröschner, A.; Seidel, T. How teacher professional development regarding classroom dialogue affects students' higher-order learning. *Teach. Teach. Educ.* **2015**, *47*, 108–119. [CrossRef]
41. Tripp, T.; Rich, P. Using video to analyze one's own teaching. *Br. J. Educ. Technol.* **2012**, *43*, 678–704. [CrossRef]
42. Lewis, C.; Perry, R.; Murata, A. How should research contribute to instructional improvement? The case of lesson study. *Educ. Res.* **2006**, *35*, 3–14. [CrossRef]
43. Davies, P.; Dunnill, R. 'Learning Study'as a model of collaborative practice in initial teacher education. *J. Educ. Teach.* **2008**, *34*, 3–16. [CrossRef]
44. Holmqvist, M. Teachers' learning in a learning study. *Instr. Sci.* **2010**, *39*, 497–511. [CrossRef]
45. Rich, P.J.; Hannafin, M. Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *J. Teach. Educ.* **2009**, *60*, 52–67. [CrossRef]

46. Van Es, E.A. Examining the development of a teacher learning community: The case of a video club. *Teach. Teach. Educ.* **2012**, *28*, 182–192. [CrossRef]
47. Borko, H.; Jacobs, J.; Eiteljorg, E.; Pittman, M.E. Video as a tool for fostering productive discussions in mathematics professional development. *Teach. Teach. Educ.* **2008**, *24*, 417–436. [CrossRef]
48. Kleinknecht, M.; Schneider, J. What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teach. Teach. Educ.* **2013**, *33*, 13–23. [CrossRef]
49. Beisiegel, M.; Mitchell, R.; Hill, H.C. The Design of Video-Based Professional Development: An Exploratory Experiment Intended to Identify Effective Features. *J. Teach. Educ.* **2017**, *69*, 69–89. [CrossRef]
50. Seidel, T.; Stürmer, K.; Blomberg, G.; Kobarg, M.; Schwindt, K. Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teach. Teach. Educ.* **2011**, *27*, 259–267. [CrossRef]
51. Jamil, F.M.; Hamre, B.K. Teacher Reflection in the Context of an Online Professional Development Course: Applying Principles of Cognitive Science to Promote Teacher Learning. *Action Teach. Educ.* **2018**, *40*, 220–236. [CrossRef]
52. Bates, M.S.; Phalen, L.; Moran, C.G. If you build it, will they reflect? Examining teachers' use of an online video-based learning website. *Teach. Teach. Educ.* **2016**, *58*, 17–27. [CrossRef]
53. Dede, C.; Eisenkraft, A.; Frumin, K.; Hartley, A. *Teacher Learning in the Digital Age: Online Professional Development in STEM Education*; Harvard Education Press: Cambridge, MA, USA, 2016.
54. Dana, N.F.; Pape, S.J.; Griffin, C.C.; Prosser, S.K. Incorporating practitioner inquiry into an online professional development program: The Prime Online experience. *Prof. Dev. Educ.* **2017**, *43*, 212–231. [CrossRef]
55. Owston, R.; Wideman, H.; Murphy, J.; Lupshenyuk, D. Blended teacher professional development: A synthesis of three program evaluations. *Internet High. Educ.* **2008**, *11*, 201–210. [CrossRef]
56. Theelen, H.; Van den Beemt, A.; den Brok, P. Classroom simulations in teacher education to support preservice teachers' interpersonal competence: A systematic literature review. *Comput. Educ.* **2019**, *129*, 14–26. [CrossRef]
57. Billingsley, G.; Smith, S.; Smith, S.; Meritt, J. A systematic literature review of using immersive virtual reality technology in teacher education. *J. Interact. Learn. Res.* **2019**, *30*, 65–90.
58. Macià, M.; García, I. Informal online communities and networks as a source of teacher professional development: A review. *Teach. Teach. Educ.* **2016**, *55*, 291–307. [CrossRef]
59. Lantz-Andersson, A.; Lundin, M.; Selwyn, N. Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teach. Teach. Educ.* **2018**, *75*, 302–315. [CrossRef]
60. Saifuddin, K.M.; Strange, M.H. School teacher professional development in online communities of practice: A systematic literature review. In Proceedings of the 15th European Conference on e-Learning, Prague, Czech Republic, 27–28 October 2016; Academic Conferences International Limited: Reading, UK, 2016; pp. 605–614.
61. Frumin, K.; Dede, C.; Fischer, C.; Foster, B.; Lawrenz, F.; Eisenkraft, A.; Fishman, B.; Jurist Levy, A.; McCoy, A. Adapting to large-scale changes in Advanced Placement Biology, Chemistry, and Physics: The impact of online teacher communities. *Int. J. Sci. Educ.* **2018**, *40*, 397–420. [CrossRef]
62. Holmes, B. School Teachers' Continuous Professional Development in an Online Learning Community: Lessons from a case study of an e T winning Learning Event. *Eur. J. Educ.* **2013**, *48*, 97–112. [CrossRef]
63. Surrette, T.N.; Johnson, C.C. Assessing the ability of an online environment to facilitate the critical features of teacher professional development. *Sch. Sci. Math.* **2015**, *115*, 260–270. [CrossRef]
64. Keengwe, J.; Kang, J.-J. A review of empirical research on blended learning in teacher education programs. *Educ. Inf. Technol.* **2012**, *18*, 479–493. [CrossRef]
65. Lin, Q. Student satisfactions in four mixed courses in elementary teacher education program. *Internet High. Educ.* **2008**, *11*, 53–59. [CrossRef]
66. Goldenberg, L.; Culp, K.; Clements, M.; Anderson, A.; Pasquale, M. Online professional development for high-school biology teachers: Effects on teachers' and students' knowledge. *J. Technol. Teach. Educ.* **2014**, *22*, 287–309.
67. Tu, C.-H.; McIsaac, M. The relationship of social presence and interaction in online classes. *Am. J. Distance Educ.* **2002**, *16*, 131–150. [CrossRef]
68. Beatty, B.R.; Allix, N. Being there: Exploring extending and enriching distance education in real time with WebEx. In Proceedings of the Open and Distance Learning Association of Australia (ODLAA) Conference 2005, Adelaide, SA, Australia, 5–9 November 2005.
69. Subhash, S.; Cudney, E.A. Gamified learning in higher education: A systematic review of the literature. *Comput. Hum. Behav.* **2018**, *87*, 192–206. [CrossRef]
70. Qian, Y.; Hambrusch, S.; Yadav, A.; Gretter, S. Who Needs What: Recommendations for Designing Effective Online Professional Development for Computer Science Teachers. *J. Res. Technol. Educ.* **2018**, *50*, 164–181. [CrossRef]
71. Taylor, J.M. Characteristics of Effective Online Professional Development: A Case Study Analysis of an Online Professional Development Course Offered via Blackboard. Ph.D. Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 2011.
72. Tella, A. Reliability and factor analysis of a blackboard course management system success: A scale development and validation in an educational context. *J. Inf. Technol. Educ. Res.* **2011**, *10*, 55–80. [CrossRef]
73. Thomas, T.S. *Online vs. Face-to-Face: Educator Opinions on Professional Development Delivery Methods*; University of Alabama Libraries: Tuscaloosa, AL, USA, 2010.

74. Unwin, A. Developing new teacher inquiry and criticality: The role of online discussions. *Br. J. Educ. Technol.* **2015**, *46*, 1214–1222. [CrossRef]
75. West, R.E.; Rich, P.J.; Shepherd, C.E.; Recesso, A.; Hannafin, M.J. Supporting induction teachers' development using performance-based video evidence. *J. Technol. Teach. Educ.* **2009**, *17*, 369–391.
76. Timperley, H.; Wilson, A.; Barrar, H.; Fung, I. *Teacher Professional Learning and Development. Best Evidence Synthesis iteration (BES)*; Ministry of Education: Wellington, New Zealand, 2007.
77. Robinson, V.M.; Hohepa, M.; Lloyd, C. *School Leadership and Student Outcomes: Identifying What Works and Why*; Australian Council for Educational Leaders: Winmalee, NSW, Australia, 2007; Volume 41, pp. 1–27.
78. Chen, N.-S.; Wei, C.-W.; Wu, K.-T.; Uden, L. Effects of high level prompts and peer assessment on online learners' reflection levels. *Comput. Educ.* **2009**, *52*, 283–291. [CrossRef]
79. Donnelly, R. Interaction analysis in a 'Learning by Doing'problem-based professional development context. *Comput. Educ.* **2010**, *55*, 1357–1366. [CrossRef]
80. Jarosewich, T.; Vargo, L.; Salzman, J.; Lenhart, L.; Krosnick, L.; Vance, K.; Roskos, K. Say what? The quality of discussion board postings in online professional development. *New Horiz. Educ.* **2010**, *58*, 118–132.
81. Prestridge, S. Teachers' talk in professional development activity that supports change in their ICT pedagogical beliefs and practices. *Teach. Dev.* **2009**, *13*, 43–55. [CrossRef]
82. Lauer, P.A.; Christopher, D.E.; Firpo-Triplett, R.; Buchting, F. The impact of short-term professional development on participant outcomes: A review of the literature. *Prof. Dev. Educ.* **2013**, *40*, 207–227. [CrossRef]
83. Allen, J.P.; Hafen, C.A.; Gregory, A.C.; Mikami, A.Y.; Pianta, R. Enhancing Secondary School Instruction and Student Achievement: Replication and Extension of the My Teaching Partner-Secondary Intervention. *J Res. Educ. Eff.* **2015**, *8*, 475–489. [CrossRef]
84. Kraft, M.A.; Hill, H.C. Developing Ambitious Mathematics Instruction Through Web-Based Coaching: A Randomized Field Trial. *Am. Educ. Res. J.* **2020**, *57*, 2378–2414. [CrossRef]
85. Pianta, R.; Hamre, B.; Downer, J.; Burchinal, M.; Williford, A.; Locasale-Crouch, J.; Howes, C.; La Paro, K.; Scott-Little, C. Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Educ. Dev.* **2017**, *28*, 956–975. [CrossRef]
86. Mashburn, A.J.; Downer, J.T.; Hamre, B.K.; Justice, L.M.; Pianta, R.C. Consultation for teachers and children's language and literacy development during pre-kindergarten. *Appl. Dev. Sci.* **2010**, *14*, 179–196. [CrossRef]
87. Early, D.M.; Maxwell, K.L.; Ponder, B.D.; Pan, Y. Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Child. Res. Q.* **2017**, *38*, 57–70. [CrossRef]
88. Vernon-Feagans, L.; Bratsch-Hines, M.; Varghese, C.; Bean, A.; Hedrick, A. The Targeted Reading Intervention: Face-to-face vs. webcam literacy coaching of classroom teachers. *Learn. Disabil. Res. Pract.* **2015**, *30*, 135–147. [CrossRef]
89. Vernon-Feagans, L.; Kainz, K.; Hedrick, A.; Ginsberg, M.; Amendum, S. Live webcam coaching to help early elementary classroom teachers provide effective literacy instruction for struggling readers: The Targeted Reading Intervention. *J. Educ. Psychol.* **2013**, *105*, 1175. [CrossRef]
90. Landry, S.H.; Anthony, J.L.; Swank, P.R.; Monseque-Bailey, P. Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *J. Educ. Psychol.* **2009**, *101*, 448. [CrossRef]
91. Jackson, B.; Larzelere, R.; Clair, L.S.; Corr, M.; Fichter, C.; Egertson, H. The impact of HeadsUp! Reading on early childhood educators' literacy practices and preschool children's literacy skills. *Early Child. Res. Q.* **2006**, *21*, 213–226. [CrossRef]
92. Rosenberg, N.E.; Artman-Meeker, K.; Kelly, E.; Yang, X. The Effects of a Bug-in-Ear Coaching Package on Implementation of Incidental Teaching by Paraprofessionals in a K-12 School. *J. Behav. Educ.* **2020**, *29*, 409–432. [CrossRef]
93. Ascetta, K.; Harn, B.; Durán, L. Comparing Self-Reported and Performance-Based Online Feedback on Early Childhood Teachers' Implementation of Language Strategies. *Early Child. Educ. J.* **2019**, *47*, 353–365. [CrossRef]
94. Dash, S.; Magidin de Kramer, R.; O'Dwyer, L.M.; Masters, J.; Russell, M. Impact of online professional development or teacher quality and student achievement in fifth grade mathematics. *J. Res. Technol. Educ.* **2012**, *45*, 1–26. [CrossRef]
95. Fishman, B. *Professional Development for the Redesigned AP Biology Exam: Teacher Participation Patterns and Student Outcomes*; American Educational Research Association: Washington, DC, USA, 2014.
96. Griffin, C.C.; Dana, N.F.; Pape, S.J.; Algina, J.; Bae, J.; Prosser, S.K.; League, M.B. Prime online: Exploring teacher professional development for creating inclusive elementary mathematics classrooms. *Teach. Educ. Spec. Educ.* **2018**, *41*, 121–139. [CrossRef]
97. Shaha, S.H.; Ellsworth, H. Predictors of success for professional development: Linking student achievement to school and educator successes through on-demand, online professional learning. *J. Instr. Psychol.* **2013**, *40*, 19.
98. Shaha, S.H.; Glassett, K.F.; Copas, A. The impact of teacher observations with coordinated professional development on student performance: A 27-state program evaluation. *J. Coll. Teach. Learn. (TLC)* **2015**, *12*, 55–64. [CrossRef]
99. Shaha, S.; Glassett, K.; Copas, A.; Ellsworth, H. I Schools: The Student-Based Impact of Online, On-Demand Professional Development on Educators. *Contemp. Issues Educ. Res.* **2015**, *8*, 227–234. [CrossRef]
100. Shaha, S.; Glassett, K.; Copas, A.; Huddleston, T.L. Maximizing educator enhancement: Aligned seminar and online professional development. *Contemp. Issues Educ. Res. (CIER)* **2016**, *9*, 15–22. [CrossRef]
101. Shaha, S.H.; Glassett, K.; Copas, A. Sustaining Student Gains from Online On-Demand Professional Development. *J. Int. Educ. Res.* **2015**, *11*, 163–172. [CrossRef]

102. Shaha, S.H.; Glassett, K.F.; Ellsworth, H. Long-term impact of on-demand professional development on student performance: A longitudinal multi-state study. *J. Int. Educ. Res. (JIER)* **2015**, *11*, 29–34. [CrossRef]
103. Walker, A.; Recker, M.; Ye, L.; Robertshaw, M.B.; Sellers, L.; Leary, H. Comparing technology-related teacher professional development designs: A multilevel study of teacher and student impacts. *Educ. Technol. Res. Dev.* **2012**, *60*, 421–444. [CrossRef]
104. Weschke, B.; Barclay, R.D.; Vandersall, K. Online Teacher Education: Exploring the Impact of a Reading and Literacy Program on Student Learning. *J. Asynchronous Learn. Netw.* **2011**, *15*, 22–43. [CrossRef]
105. Fisher, J.B.; Schumaker, J.B.; Culbertson, J.; Deshler, D.D. Effects of a Computerized Professional Development Program on Teacher and Student Outcomes. *J. Teach. Educ.* **2010**, *61*, 302–312. [CrossRef]
106. Fishman, B.; Konstantopoulos, S.; Kubitskey, B.W.; Vath, R.; Park, G.; Johnson, H.; Edelson, D.C. Comparing the Impact of Online and Face-to-Face Professional Development in the Context of Curriculum Implementation. *J. Teach. Educ.* **2013**, *64*, 426–438. [CrossRef]
107. Rock, H.M. *The Effect of Face-to-Face versus Online Pedagogy-Based Professional Development on Student Learning Outcomes*; Grand Canyon University: Phoenix, AZ, USA, 2017.
108. Webb, D.C.; Nickerson, H.; Bush, J.B. *A Comparative Analysis of Online and Face-to-Face Professional Development Models for CS Education*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 621–626.
109. Russell, M.; Carey, R.; Kleiman, G.; Venable, J.D. Face-to-face and online professional development for mathematics teachers: A comparative study. *J. Asynchronous Learn. Netw.* **2009**, *13*, 71–87. [CrossRef]

education sciences

MDPI

Article

Impact of Contextual Predictors on Value-Added Teacher Effectiveness Estimates

Ismail Aslantas

School of Education, Durham University, Durham DH1 1TA, UK; ismail.aslantas@durham.ac.uk

Received: 4 November 2020; Accepted: 16 December 2020; Published: 18 December 2020

Abstract: It is widely believed that the teacher is one of the most important factors influencing a student's success at school. In many countries, teachers' salaries and promotion prospects are determined by their students' performance. Value-added models (VAMs) are increasingly used to measure teacher effectiveness to reward or penalize teachers. The aim of this paper is to examine the relationship between teacher effectiveness and student academic performance, controlling for other contextual factors, such as student and school characteristics. The data are based on 7543 Grade 8 students matched with 230 teachers from one province in Turkey. To test how much progress in student academic achievement can be attributed to a teacher, a series of regression analyses were run including contextual predictors at the student, school and teacher/classroom level. The results show that approximately half of the differences in students' math test scores can be explained by their prior attainment alone (47%). Other factors, such as teacher and school characteristics explain very little the variance in students' test scores once the prior attainment is taken into account. This suggests that teachers add little to students' later performance. The implication, therefore, is that any intervention to improve students' achievement should be introduced much earlier in their school life. However, this does not mean that teachers are not important. Teachers are key to schools and student learning, even if they are not differentially effective from each other in the local (or any) school system. Therefore, systems that attempt to differentiate "effective" from "ineffective" teachers may not be fair to some teachers.

Keywords: value-added model; teacher performance evaluation; teacher accountability; effectiveness; contextual predictors

1. Introduction

With growing interest in identifying and improving student achievement through large-scale assessments, several researchers have conducted a considerable number of studies to uncover the factors that influence student academic performance. The common consensus is that of school-related factors, it is teachers that are the most significant [1–3]. Accordingly, politicians have tended to develop educational policies that consider schools and teachers responsible for the performance of pupils. One of the key concerns of decisionmakers is to ensure that effective teachers are hired in classrooms and, similarly, parents also desire their children to be enrolled in the best schools and be taught by well qualified teachers. There is, therefore, a need for a performance appraisal framework that can evaluate the role of teachers in achieving student targeted success. While it is accepted that evaluating teacher performance is beneficial in enhancing teacher development and student outcomes, it is a complex process and there is no perfect measure [4]. Since longitudinal student achievement results have become readily accessible, academics and decisionmakers have considered alternative measures that focus upon the students' achievement growth rather than the percentage of students in the class or the school reaching a threshold value determined by teacher and school effectiveness indicators.

Since teachers' performance results are heavily influenced by the context of students and other factors that are beyond the control of teachers [5], value-added models that take into account the

cumulative nature of students' learning and that allow students' progress over time to be monitored, have been recently favored. Value-added models (VAMs) are statistical approaches used to estimate the effectiveness of individual teachers and schools by employing students' longitudinal test scores (and some covariates). While a variety of value-added models, with their own specific advantages and drawbacks, have been implemented to estimate an individual teacher's contribution to student attainment, the fundamental concept behind all VAMs is to determine the changes in the students' school achievement over the years. In all student standardized assessment-based VAMs, a particular teacher's performance is statistically estimated using the test scores of their students by subject and grade. Changes in student achievement in tests performed for at least two consecutive years are then attributed to teacher effects. Since students can be taught by different teachers, it is difficult to say if any improvement in student test scores is the result of any particular teacher's effort. However, by using student test scores for the same teacher for multiple years, as in this current study, it can be ensured that the effects (if any) can be attributed to the teacher. In the VAM concept, teacher effect has a special meaning that relates to evaluating the discrepancies between expected and observed student test scores [3,6].

VAMs are one of the most contentious and critical matters for education policy in the assessment of teacher performance. Theoretically, VAMs separate a particular teacher's contribution to the achievement of students from other factors that are beyond the control of the teacher. These include student characteristics, school and classroom characteristics and students' prior year test scores. On the basis of such performance evaluations, high-stakes decisions regarding teachers' professional careers are made, such as pay rises, promotions, or even the loss of a job. The VAM score estimates how much a teacher has contributed to enhancing their students' performance. Based on this score, the teacher is ranked by how much difference he/she makes in their students' achieved test scores compared to their predicted scores. Such a ranking is often used to reward or penalize teachers. However, the premise of this effectiveness evaluation has been found to be flawed by many educators [7–9]. For example, VAMs assume that students are randomly assigned to teachers, which is rarely the case in reality. Nonrandom assignment of teachers to students can bias the estimate of teacher effects [10]. Other studies argue that standardized student achievements alone are not a sufficiently reliable metric to establish the link between student learning and what teachers do. The Economic Policy Institute (EPI) recommend that standardized academic performance be considered as just one factor among many factors, for a more accurate understanding of the actual teaching performance of teachers in the classroom [11]. Despite these concerns, several scholars have come up with different models, each trying to overcome the limitations of the others.

Nevertheless, the literature agrees that lagged test scores may provide a better measure of VAM estimates. Hu [12] recorded that on average, 57% and 59% of the changes in student math and reading test scores can be explained by the closest lagged score in the related subjects, respectively. Similarly, Kersting et al. [6] found that just one test score from a previous year explained 68% of the variation in the students' actual scores. In accordance with the scope of this study, Rothstein [13] analyzed the impact of variables on the value-added estimates through modifying the models' R-squared by employing 28 contextual variables, such as ethnicity, sex, free/reduced lunch status, parental education, etc. The integration of the 28 variables into the equations contributed to changes of 0.05 and 0.01 in R^2 depending on the model to which they were applied. While existing literature suggests that lagged scores may be a better measure in explaining the variation in a student's actual test score [1,14–18], there is no consensus regarding the association between contextual predictors, such as observable student, teacher, classroom and school characteristics and student outcomes.

Although Alban [14] and Gagnon [15] report that adding student-level variables in the estimates improves predictability, most studies suggest that the influence of variables at the student level employed in the equations is negligible [6,12,19–23]. Gagnon [15] analyzed the use of different predictors at the student level in the value-added teacher effectiveness estimates, including lagged scores, poverty, ethnicity, sex, English as a second language, disability status, attendance and suspension.

The results show that the previous test scores in the third through to the seventh grade explain roughly 76% of the variance in test scores in the eighth grade. The researcher concluded that the lagged test scores are better predictors of potential achievement. Other student factors employed in the equations, except ethnicity, are shown to be significant predictors. Hu [12], on the other hand, used up to three years of pupil test scores in math and reading, as well as student demographic characteristics, such as sex, race, language learner status, gifted and disability status, and class size and found that the inclusion of all variables explained only 2% of the variance in the students' actual test scores.

Other studies looked at the use of teacher-level predictors in value-added estimates. These studies show a positive relationship between the teacher variables, such as Grand Point Average (GPA) score [24], permanent contract [25], being certified [26] and their prior year performance [27]. Conversely, some studies have suggested little or no advantage in using teacher-/classroom-level variables [18,28–30]. For instance, Leigh [31] examined the impact of teacher characteristics, such as gender, age, experience and DETA rating score (The Queensland Department of Education, Training and the Arts), on the teacher effectiveness estimates in terms of changes in the R-squared values of the models. Student achievement gains were regressed on the teacher characteristics, and it was found that the explained variance in the achievement gain, employing all teacher characteristics, never exceeded 1%.

In line with the scope of the study, the literature reviewed also focused on school-related variables. This revealed that adding school-level predictors makes little difference to the teacher value-added effectiveness estimates [18,24,32,33], except for one study suggesting that including the percentage of students receiving special education services at the school level in the estimates has a benefit [14]. To determine the important variables in the teacher value-added effectiveness estimates, Alban [14] employed a range of student-, teacher- and school-related variables, including sex, ethnicity, prior attainment, language learner status, teaching experience, highest qualification held, the socio-economic level of the school and the proportion of students receiving special education services. The study found that prior success is the only significant variable for each estimate, followed by gender and the percentage of students receiving special services at school.

In light of the findings of the existing literature, the aim of this study is to examine the contribution of contextual predictors and observable student, teacher/classroom and school characteristics, on teacher effectiveness estimates calculated by a value-added model.

The following research questions were addressed:

1. To what extent can the value-added estimates for mathematics teachers be explained by student characteristics, other than prior attainment?
2. To what extent can the value-added estimates for mathematics teachers be explained by school characteristics, over and above student characteristics?
3. To what extent can the value-added estimates for mathematics teachers be explained by teacher/classroom characteristics, over and above student and school characteristics?

2. Materials and Methods

2.1. Participants

The research involved 8th grade mathematics teachers. The participants of this study were 230 mathematics teachers from 145 secondary schools and their 7543 students in 8th grade during the 2016–2017 school year in the Samsun Provincial Directory of National Education, Turkey. Students that could be academically tracked from Grades 6 to 8 (Key stage 3–Years 7 to 9) were the target population in this study. However, as with most longitudinal studies, the inclusion of each lagged test score in the study results in a loss in the study population. The main reasons for these losses may be that, during the testing time, students did not take the test, or they moved away from the province. In a target population for which 8th grade test scores are available, the average loss of employing one previous year's test score was approximately five percent.

2.2. Data

To conduct value-added teacher performance estimates, it is essential that student data can be linked to teacher data longitudinally [34,35]. The research employed the longitudinal data from schools in Samsun province, Turkey, to examine the contribution of a range of contextual variables at student, teacher/classroom and school levels on maths teacher effectiveness estimates. The Samsun Provincial Directory of National Education has been running the "Step by Step Achievement" project (SBSA) since 2015 [36] and the exam scores from 2016–2017 in maths were used as outcome variables. All the SBSA exam scores over the years, students' background information, their school and class information (i.e., teacher's name) are stored electronically in the Samsun Provincial Directorate of National Education's private electronic systems. The longitudinal test scores of all 8th grade students registered on the system were downloaded and these data were then merged with other student-level data, including their names, unique school number, gender and language learner status.

Since the SBSA did not contain teacher-/classroom-level data, these were obtained from each of the schools via the Provincial Directorate. The teacher/classroom information included gender, classroom size, number of years of teaching experience, number of years teaching in the current school, teaching appointment field, teachers' major degree subject and their highest level of qualification and its field.

School-level data were obtained from the Ministry of National Education's official website [37]. A list of all secondary schools was first downloaded from the website of Samsun Provincial Directory of National Education [38]. Only schools that were included in the project were retained. School-level data included school type (private or state-funded), school category (general, regional boarding or vocational secondary school), location (urban, suburban or rural) and school service scores. After all three data files were merged to maintain confidentiality, participants' identities were removed from the data set and identification numbers were assigned to each student, teacher, school and school location.

Data used in the study included student longitudinal achievement data, spanning up to three consecutive years, students' characteristics, teacher/classroom background information and school information. Table 1 summarizes the outcome variables and independent student, teacher/classroom and school-level variables included in this study.

Table 1. Summary statistics for data used in value-added estimates.

	Mean	**SD**
Outcome variable		
Math test score in Grade 8	9.18	4.348
Independent Variables		
Panel A: Student characteristics		
Female	48.3%	
Language Learner	0.2%	
Math test score in Grade 7 (t-1)	9.52	4.888
N (student)	7543	
Panel B: Teacher/Classroom characteristics		
Female	53.5%	
Class size	22.93	5.702
Percentage of female students in class	0.48	0.177
Total teaching experience	10.13	5.363
Experience in the current school	3.330	2.232
Graduation field (whether related to math?)	98.8%	
Appointment field (whether related to math?)	99.8%	
Having master or higher degree	2.9%	

Table 1. *Cont.*

		Mean	SD
Field of master's (or higher) degree	Related to math	57.9%	
	Not related to math	17.6%	
	Unknown	24.45	
Class average test score in Grade 7		9.48	2.715
Class average test score in Grade 6		8.66	2.290
N (teacher)		230	
Panel C: School characteristics			
Categories	General	84.3%	
	Regional Boarding	3.5%	
	Vocational	12.2%	
Service score (1-highest, 6-lowest score)		2.01	1.433
Location	Rural	17.9%	
	Suburban	19.8%	
	Urban	62.3%	
School average test score in Grade 7		9.48	2.193
School average test score in Grade 6		8.67	1.823
N (school)		145	

2.3. Data Analysis

The study analyzed the contribution of various contextual variables at student, teacher/classroom and school levels on teacher effectiveness estimates obtained by VAM. The multiple linear regression analysis was performed using the forward-selection approach to assess the contribution of each of the contextual variables to the value-added estimates of mathematics teachers. By using the forward-selection approach, significant changes to the model-fit can be checked by including new predictor(s) in contrast with the model employed in the previous step.

In RQ1, the basic model was generated using just one prior achievement score (t-1) and then student characteristics (gender and language learner status) were added to the next stage in order to figure out how much change exists in the model-fit when utilizing the student characteristics. The highest R-squared value can be generated if all of the predictors are used in the model, so both gender and language learner identity variables were added to the basic model with the enter method, and the highest R-squared value that can be obtained in RQ1 was found. Finally, the same student characteristics were again added to the basic model using the forward-selection method. Once a model with the largest R-squared value (determined by enter method) was proposed, the model(s) to be suggested in the next step(s) (if any) was ignored in order to include only variables with a predictive impact on the estimates. The variables excluded from predictions, because they have no predictive ability, or are too small to be considered in this research question, will also not be included in the analysis of next two research questions.

Similarly, in RQ2, to determine which school characteristics make a worthwhile contribution to mathematics teachers' value-added estimates, the model obtained in research question 1 was used as the baseline model in this research question. The same steps were followed to create the final model in research question 2. For the last stage of this study, RQ3 was built on the model derived from RQ2, to which student and teacher-/classroom-level predictors were added. Unlike with the other previous two research questions, before conducting the regression analyses, the presence of any relationship between the teacher/classroom characteristics and the mathematics teachers' value-added effectiveness scores was checked. For this, individual student residual scores (the difference between predicted and actual attainment level), obtained through the final model proposed in RQ2, were aggregated at the teacher level. The mean of the residuals at the teacher level was attributed to a teacher's individual value-added effectiveness score. Then, the effectiveness scores were correlated with the teacher/classroom characteristics. After the correlation statistics, the model generated in RQ2 was used as the baseline model for this last research question to decide whether the characteristics of teacher

and classroom contribute greatly to the value-added estimates of mathematics teachers. Then, the ultimate regression model was created by pursuing the same steps in the previous research questions.

The *p*-value and confidence interval of statistical significance tests, such as the *t*-test or chi-square, etc., are still widely reported in the social sciences, regardless of whether the data ensures the assumptions required to use these statistics. As this current study involves a nonrandom sample from the study population that also has missing data, the main assumption of reporting the *p*-value is not met [39,40]. Therefore, as the key decision on analyzing the data, *p*-values of the significance tests used are not reported in this study [41–45].

3. Results

The analyses were mainly centered on the consistency of teacher value-added effectiveness estimates regarding model specifications. A series of analyses were conducted to answer each research question. Models were created by excluding and/or including contextual variables at student, school and teacher/classroom levels and changes in model fit were evaluated by checking changes in R^2 values for each research question by conducting the forward-selection method.

3.1. Research Question 1: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by Student Characteristics, Other Than Prior Attainment?

The students' math test scores in Grade 8 were used as the outcome variable, while their gender, language learner identity and seventh grade math test scores (prior attainment at Time 1) were employed as predictors.

The baseline model indicates that the previous achievement score of students alone can explain 47% (R^2 = 0.470) of the variability in Grade 8 math test scores (Table 2). Adding other student characteristics (gender and language status) to the baseline model using the enter method increased the R^2 value of the new model to 0.471 (see in Appendix A). This increase infers that students' gender and their language status contribute 0.1% to explaining the variance in their current test scores. To find out whether this minor change in the explanation of the variation in the result variable is due to both of the two variables, or just one of them, another regression analysis was performed by employing the forward method. The forward method suggested a model using the prior attainment and gender variables with the largest R^2 values, which was the same as the highest R^2 value achievable with the full model (0.471); therefore, the language learner identity variable was excluded from the final model created using the student-level variables.

Table 2. Value-added estimates for mathematics teachers using student characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment	0.470	
Forward method			
Model 1	Prior attainment	0.470	-
Model 2	Prior attainment Gender	0.471	0.001

Dependent Variable: 8th grade math test score.

3.2. Research Question 2: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by School Characteristics, over and above Student Characteristics?

In this research question, students' eighth grade maths test scores were employed as the outcome variable. Their seventh grade maths test scores, gender and five school-level variables: school categories (general, regional boarding, vocational), service scores (1—highest, 6—lowest score), locations (rural, suburban, urban), the school-level average maths test scores for seventh grade and school-level average maths test scores for sixth grade were used as predictors in this research question.

The final model created in the previous research question was used as a baseline model for this research question, where the eighth grade students' math test scores were regressed on students' prior attainment scores (t-1) and gender. All school-related variables were added to the baseline model of RQ2 at the same time using the enter method to discover the highest R-squared value achievable at the school level. Adding all school-level variables to the baseline model increased R-squared by 19 points (R^2 = 0.490) (see in Appendix B). To minimize uncertainty and consider the variables that have a predictive impact on estimates, it is important to find out whether the 19-point increase in R-squared of the full model is due to the inclusion of all, or only some of the characteristics. Therefore, a regression analysis using the forward-selection method was conducted using the same variables employed in the full model. The forward method suggested a model with the largest R^2 value, which was the same as the highest R^2 value achievable with the full model (R^2 = 0.490), where prior attainment, school-level average test scores in Grades 6 and 7 and student gender variables were employed. School categories, service scores and school location variables were excluded from the final model created for this research question about math teachers' value-added effectiveness estimates. These exclusions can also be interpreted as giving no indication that the students' current attainment in maths is linked to the school service score, the school location and the type of school attended, once the prior attainment, school-level average test scores in Grades 6 and 7 and student gender have been taken into account (Table 3).

Table 3. Value-added estimates for mathematics teachers adding school characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment Student gender	0.471	
Forward method			
Model 1	Prior attainment	0.470	−0.001
Model 2	Prior attainment 6th grade average school test score	0.487	0.016
Model 3	Prior attainment 6th grade average school test score 7th grade average school test score	0.489	0.018
Model 4	Prior attainment 6th grade average school test score 7th grade average school test score Student gender	0.490	0.019

Dependent Variable: 8th grade maths test score.

3.3. Research Question 3: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by Teacher/Classroom Characteristics, over and above Student and School Characteristics?

The research question focused on mathematics teachers' value-added estimates regarding using teacher/classroom characteristics, in addition to student- and school-level variables identified in the previous research questions. Seven observable teacher characteristics were employed as the teacher-/classroom-level predictors in this research question: gender, number of years of teaching experience, number of years teaching in the current school, teachers' major degree subject, teaching assignment field and the highest level of qualification and field, in addition to four classroom-level variables: class size, percentage of female students, and sixth and seventh grade classroom-level average maths test scores.

To determine whether there is a relationship between the teacher/classroom characteristics and teacher effectiveness scores, individual math teachers' value-added effectiveness scores were estimated by aggregating each student's residual scores (the difference between predicted and actual attainment level) obtained through the final model proposed in RQ2 at the teacher level. Pearson's r coefficients show that there is no substantial association between the teacher effectiveness score

and the teacher-/classroom-level continuous variables (see Table 4). There was barely an association between class size and teacher VAM score and, surprisingly, bigger classes had slightly more "effective" teachers (r = 0.079). Another interesting finding is that classes with a higher female student ratio are taught by less effective mathematics teachers (r = −0.101). Experience, regardless of whether in total or just in their current schools, has a negative link with their effectiveness scores, so more experienced mathematics teachers tend to be assigned a lesser effectiveness score from value-added estimates. Lastly, with very small correlation coefficients, unsurprisingly, a positive relationship appeared for classroom-level prior attainment. However, the classroom average at a two lagged year (t-2, Grade 6) was a better predictor for the effectiveness of maths teachers than the average score at the prior year (t-1, Grade 7).

Table 4. Correlation between teacher/classroom characteristics and teacher value-added effectiveness.

Teacher/Classroom Characteristics	Correlation Coefficient (r)
Class size	0.079
Percentage of female students	−0.101
Total teaching experience	−0.029
Experience in the current school	−0.017
Classroom average test score in Grade 7	0.064
Classroom average test score in Grade 6	0.106

Similarly, Cohen's effect sizes, shown in Table 5, were also calculated for each subcategory of teacher characteristics variables. On average, the value-added scores of female maths teachers were marginally worse than those of male teachers (d = −0.10). Moreover, teachers with a bachelor's degree in a field related to mathematics tend to have lower effectiveness scores, even if the effect size is very fractional (d = −0.08). More interestingly, teachers who were first appointed as primary school teachers, but later became maths teachers, have remarkably higher value-added effectiveness scores than those appointed originally as mathematics teachers (d = −0.92). This result may be attributed to the unbalanced subcategories in this variable, so this result needs to be confirmed with a variable containing a balanced subcategorical distribution. Another interesting finding is that contrary to what is believed, having a master's degree does not contribute to the effectiveness estimates for mathematics teachers, and math teachers with master's degrees even had, on average, lower effectiveness scores than those teachers with just a bachelor's degree (d = −0.25). Finally, almost no link was found between having a master's degree in a math-related field and the VAM scores of mathematics teachers (d = 0.01).

Table 5. Comparison of means for teacher/classroom characteristics.

Category	Yes	No	Cohen's d
Gender (Female)	−0.015	0.017	−0.10
Graduation field (related to math)	−0.000	0.027	−0.08
Appointment field (related to math)	−0.001	0.298	−0.92
Terminal degree (master or higher degree)	−0.080	0.002	−0.25
Field of master's degree			
Related	0.003	−0.000	0.01
Not Related	−0.312	0.002	−0.96
Unknown	−0.109	0.001	−0.34

Overall standard deviation = 0.325.

After disclosing the association between teacher-/classroom-level characteristics and teacher VAM scores, as for previous research questions, a best-fit regression model was created and had the highest R-squared that can be obtained by using as few variables as possible.

The final model created in RQ2 was again used as a baseline model in this research question (see Table 6). Using the enter method, all teacher-/classroom-related variables were added to this baseline

model at the same time to discover the highest R-squared value achievable at teacher/classroom levels. Adding all teacher-/classroom-level characteristics to the baseline model increased the variance explained by 12 points ($R^2 = 0.502$). To include only variables that have predictive impacts on estimates, another regression analysis was carried out with the forward method, employing the same predictors used in the full model.

Table 6. Value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment 6th grade average school test score 7th grade average school test score Student gender	0.490	
Forward method			
Model 1	Prior attainment	0.470	−0.020
Model 2	Prior attainment 6th grade average classroom test score	0.497	0.007
Model 3	Prior attainment 6th grade average classroom test score 7th grade average classroom test score	0.500	0.010
Model 4	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender	0.501	0.011
Model 5	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender Class size	0.501	0.011
Model 6	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender Class size Percentage of female students	0.502	0.012

Dependent Variable: 8th grade maths test score.

The forward method proposed a model with the largest R^2 value achievable at teacher/classroom levels (revealed by the full model), which includes students' prior attainment scores (t-1), classroom-level average test scores in Grades 6 and 7, student gender, class size and percentage of female students' variables ($R^2 = 0.502$) (see Appendix C). School-level variables and sixth and seventh grade average school test scores were included in the final model in RQ2. However, once the teacher/class characteristics had been taken into account, these variables' predictive power on value-added estimates disappeared. Therefore, these variables were removed from the final model of this research question. In addition, teacher gender, major degree subject, teaching field, number of years of teaching experience, number of years teaching in the current school and highest level of qualification and field variables were also excluded from the final model, as they did not contribute to the variance explained in the students' current test scores. Interestingly, these exclusions show that none of the teacher characteristics are considerably related to the achievements of students, but all classroom characteristics are included in the final model. Therefore, it can be interpreted that student achievement is more related to class characteristics than the observable characteristics of the teacher.

The full list of standardized coefficients for the predictors employed in the final model were also investigated (see Table 7). The overall conclusion is that when the hold of the other factors is constant, there is the largest positive relationship between students' prior mathematics attainment and their recent mathematics outcomes; for each correct answer increase in the prior mathematics test score, the number of the correct answers in the Grade 8 test increases by 0.615. The second largest

relationship was found between the eighth grade maths score and the class average maths score at two lagged years (t-2). For each correct answer increase in the sixth grade maths average classroom test score, the recent maths attainment increases on average by 0.269. Interestingly, a negative relationship was revealed between the nearest prior year's classroom average (t-1) and the eighth grade math scores. When the negative standardized coefficient accounts for each one point increase in the seventh grade maths average classroom test score, the recent maths attainment decreases by 0.118, on average. In the current maths test, it is expected that females would be 0.039 points higher than male students. Another unexpected conclusion related to class size was that the standardized coeffect indicated that, on average, larger classes were more successful in the eighth grade maths test. Finally, although female students were, on average, more successful in the mathematics test in Grade 8, the classes with more female students were less successful in the same test. For each increase in the percentage of female students in the classroom, the number of correct answers in the eighth grade maths test decreases by 0.024.

Table 7. Standardized regression coefficients for the variables used in the final model.

Variables	Standardized Coefficient β
Prior attainment (t-1)	0.615
6th grade average classroom test score	0.269
7th grade average classroom test score	−0.118
Student gender (female)	0.039
Class size	0.026
Percentage of female students	−0.024

$R^2 = 0.502$.

4. Discussion

A student's academic achievement is linked with various student-related, teacher-/classroom-related and school-related factors and in a growing number of studies there is a consensus that, among school-related resources, the most crucial factor is teachers [1–3,46]. Consequently, the accountability of teachers for students' school achievement is among the educational issues on which policymakers and researchers have recently focused.

However, teacher effectiveness is not an easy attribute to measure. Recently, measures of teacher effectiveness have relied on more objective measures, such as student performance in high-stakes tests. The use of student data in measuring teacher performance is one of the most controversial and important issues of educational policy. Significant decisions affecting teachers' professional careers, such as pay raises, promotions, or redundancies, are made based on such performance evaluations. Schools and teachers are penalized and even shamed based on such measures. A well known example of a teacher performance evaluation measure based on student achievement is that of value-added models. VAMs, theoretically, isolate a particular teacher's relative effectiveness on his/her students' achievement from other factors outside the teacher's control. These include student, school, teacher, classroom characteristics and students' prior year test scores. The main purpose of this study is to examine the contribution of contextual predictors, such as observable student, teacher/classroom and school characteristics to teacher value-added effectiveness estimates.

Consistent with the findings of the literature review, this study suggests that the strongest student-related factor in explaining the variation in a student's current test score is their nearest prior attainment (math score at Grade 7). The results show that approximately half of the variance in the students' Grade 8 math test scores (47%) can be explained by their Grade 7 math results alone. The literature in the review also agrees that the strongest contribution to value-added estimates is from previous years' test scores. For example, Hu [12] wrote that the nearest prior year's achievement score alone accounted for an average of 57% and 59% of the variance in students' current achievements in

math and reading, respectively, while Kersting et al. [6] found that 68% of the variance in the students' current scores was explained by controlling only test scores from one previous year.

Another conclusion reached in this study is that, in agreement with the literature, student gender is not an important factor in explaining differences in the teacher effectiveness estimates. Similar to this current study, Heistad [22] found that adding gender to a model that already controlled for students' prior attainment and race increased the explanatory power by 0.1% to 0.4% depending on the testing year. Tobe [26] excluded gender in the analyses as it did not make a statistically significant contribution to explaining the variance in their model. This current study of teacher effectiveness estimates, which employed data from Turkey, came to the same conclusion as the studies in the review—that is, that the strongest predictor in explaining students' test scores in the current year is their prior test scores, and a student's gender is not an important factor in the value-added teacher effectiveness estimates when the prior attainment is controlled [12,13,21].

Adding the language learner identity variable (i.e., whether Turkish is a second language or not) in value-added estimates also showed no impact on students' test scores. This variable was therefore removed from further analysis. This is likely because only a very small proportion (0.2%) of the students in the study population identified Turkish as their second language. However, the literature in the review also revealed that students' language identity did not considerably contribute to the teacher value-added effectiveness estimates [12,20,21,23].

Regarding changes in the R^2 value, the findings suggest that the most important school-related variables, in terms of explaining the variation in students' current test scores, are the average math test scores in Grades 6 and 7. Adding sixth and seventh grade average math test scores to the model obtained in RQ1 increased the R^2 of the new model to the highest R^2 value that could be achieved with the school characteristics. Therefore, it was determined that the other three characteristics, category, service scores and location, do not have a considerable predictive impact on value-added estimates. This result can also be interpreted to mean that there is no indication that students' current attainment in math is linked to school service score, school location and the type of school attended, once their prior attainment, sixth and seventh grade average test scores and gender have been taken into account. Contrary to this, Sander and Horn [47] summarized previous research findings and reported that cumulative academic gains for schools in the entire state are unrelated to schools' average attainment. Supporting the finding about the ineffectiveness of school-level variables, the retrieved literature on the effectiveness of teachers also indicates that school-level measures, the percentage of students receiving free/reduced-price lunches, class size, racial/ethnic composition, students with special educational needs and those with English as a second language (ESL) accounted for very little of the variance in student attainment [33]. Ballou et al. [32] suggest that controlling the percentage of students eligible for free/reduced-price lunches in schools has a substantial impact on TVAAS (the Tennessee Value-Added Assessment System) estimates in some grades and subjects; however, the authors also emphasize the precision of the models used and advise caution in terms of this finding. Alban [14], using the socio-economic level of the school, the percentage of students receiving special services, enrolment, and mobility and ethnicity as school-level predictors, found that prior achievement is the only significant variable in each estimation, followed by the percentage of students receiving special education.

Many recent studies report that the value-added estimates of teachers are loosely tied to the observable characteristics of teachers. To determine whether a relationship exists between teacher/classroom characteristics and the effectiveness of mathematics teachers, the residual scores of individual students were aggregated at the teacher level, and the aggregated residuals at the teacher level (teacher effectiveness score) were then correlated with teacher/classroom characteristics. The results suggest that no considerable relationship exists between teacher effectiveness and the teacher-/classroom-level variables. Interestingly, the experience of teacher, regardless of whether over their total career or only in their current schools, has a negative relationship with the effectiveness scores of teachers; thus, more experienced mathematics teachers tend to acquire lower effectiveness scores from value-added estimates. The findings related to teacher/classroom characteristics are consistent

with previous research in that no clear relationship can be found between teacher/class characteristics and teacher effectiveness [24,29,48].

The existing literature presents mixed findings on experience; therefore, the association between experience and effectiveness remains unanswered. In the current research, both total and current teacher experience have a slightly negative relationship with teachers' effectiveness scores. Conversely, Wayne and Youngs [49] found that most studies claim a positive relationship in this respect [34,50]. More robust evidence is therefore needed to ascertain whether a close relationship exists between teacher effectiveness and experience.

Aside from correlation analyses, the changes in R^2 indicate that the most important teacher-/classroom-level variables in explaining variations in students' current test scores are classroom-level average test scores in Grades 6 and 7, class size and the percentage of variables associated with female students. The exclusion of the other teacher-/classroom-level variables, teacher gender, graduation field, teaching assignment field, total teaching experience, experience in the current school, terminal degree and field of terminal degree indicate the lack of obvious links with students' current math attainment when the proposed predictors in the final model (model 6) were used. Interestingly, although the final model contains all classroom-level characteristics, consistent with a large body of literature, the findings also reveal that none of the teacher characteristics are directly linked to students' attainment in mathematics [1,13,18,27,29,31,33].

Nye et al. [30] examined the impact of teacher characteristics, experience and education on teacher effectiveness estimates and found that none of them have a considerable impact. Their contribution to the variance explained in each estimate never exceeded 5%. Conversely, few studies have revealed the impacts of certain teacher characteristics on the effectiveness estimates. For instance, Kukla-Acevedo [24] found that among teacher characteristics employed, only the overall undergraduate performance (GPA) of teachers in math has a consistent and positive impact on students' mathematics achievement. Goel and Barooah [25] suggested that only permanent employment status (tenured) positively affects teacher effectiveness estimates. Moreover, Tobe [26] reported that apart from certification by the state, none of the other teacher characteristics have an impact on students' attainment. Munoz et al. [51] also discovered that teacher experience is the only teacher characteristic that has predictive power on teacher effectiveness estimates.

The finding related to class size based on the standardized coefficient is another surprising conclusion drawn from this current study. Only a modest correlation was found between class size and teacher effectiveness, and, interestingly, larger classes performed better in eighth grade math tests on average. Taking into account debates on class size, with support from more precise evidence, this finding may contribute to efforts towards a reconsideration of the policy of class size reduction, which involves considerable costs [2,52]. On the other hand, completely ignoring this policy is not advocated, considering the existing evidence from many studies suggesting that reducing class size can increase the academic success of students [53,54].

To sum up, the study reveals that, although approximately half of the variance in student current math test scores can be explained by their prior attainment alone (47%), including other contextual predictors, such as teacher or school characteristics, makes a very limited contribution to teachers' value-added estimates. Classroom-level average test scores in one and two lagged years, student gender, class size and percentage of female students variables are the other contextual predictors that contribute to the value-added effectiveness estimates, but R^2 change analyses suggests that the contribution of the contextual variables to the variance explained in math has never exceeded 4%. This result shows that student achievement mostly depends on the performance of the student in early education. The focus of education and investment, therefore, should clearly be in the early years.

Funding: This research was funded by the Ministry of National Education, Turkey.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The author declares no conflict of interest.

Appendix A

Full model of value-added estimates for mathematics teachers using student characteristics.

Table A1. Value-added estimates for mathematics teachers using student characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment Gender Language Learner ID	0.471

Appendix B

Full model of value-added estimates for mathematics teachers adding school characteristics.

Table A2. Value-added estimates for mathematics teachers adding school characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment Student gender School categories Service score School locations 7th grade average school test score 6th grade average school test score	0.490

Appendix C

Full model of value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Table A3. Value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment 6th grade average school test score 7th grade average school test score Student gender Teacher gender Class size Percentage of female students Graduation field Appointment field Total teaching experience Experience in the current school Terminal degree Field of terminal degree 7th grade average classroom test score 6th grade average classroom test score	0.502

References

1. Aaronson, D.; Barrow, L.; Sander, W. Teachers and student achievement in the Chicago public high schools. *J. Labor Econ.* **2007**, *25*, 95–135. [CrossRef]
2. Rivkin, S.G.; Hanushek, E.A.; Kain, J.F. Teachers, schools, and academic achievement. *Econometrica* **2005**, *73*, 417–458. [CrossRef]
3. Sanders, W.L.; Wright, S.P.; Horn, S.P. Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *J. Pers. Eval. Educ.* **1997**, *11*, 57–67. [CrossRef]

4. See, B.H. Challenges in Using research evidence in improving teaching quality. *Bera Res. Intell.* **2020**, *144*, 20–21.
5. Wei, H.; Hembry, T.; Murphy, D.L.; McBride, Y. *Value-Added Models in the Evaluation of Teacher Effectiveness: A Comparison of Models and Outcomes*; Pearson: New York, NY, USA, 2012.
6. Kersting, N.B.; Chen, M.; Stigler, J.W. Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Educ. Policy Anal. Arch.* **2013**, *21*, 1–39.
7. Rothstein, J. *Can Value-Added Models Identify Teachers' Impacts*; (Policy Brief); Institute for Research on Labor and Employment: Berkeley, CA, USA, 2016.
8. McCaffrey, D.F.; Koretz, D.; Lockwood, J.R.; Hamilton, L.S. *The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness*; RAND Corporation: Boston, MA, USA, 2004.
9. Pivovarova, M.; Amrein-Beardsley, A.; Broatch, J. Value-added models (VAMs): Caveat emptor. *Stat. Public Policy* **2016**, *3*, 1–9. [CrossRef]
10. Rothstein, J. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Q. J. Econ.* **2010**, *125*, 175–214.
11. Baker, E.L.; Barton, P.E.; Darling-Hammond, L.; Haertel, E.; Ladd, H.F.; Linn, R.L.; Ravitch, D.; Rothstein, R.; Shavelson, R.J.; Shepard, L.A. *Problems with the Use of Student Test Scores to Evaluate Teachers*; Economic Policy Institute: Washington, DC, USA, 2010.
12. Hu, J. Teacher Evaluation Based on an Aspect of Classroom Practice and on Student Achievement: A Relational Analysis between Student Learning Objectives and Value-Added Modeling. Ph.D. Thesis, Boston College, Boston, MA, USA, 2015.
13. Rothstein, J. Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Educ. Financ. Policy* **2009**, *4*, 537–571. [CrossRef]
14. Alban, T.R. Evaluating School and Teacher Effectiveness: A Comparison of Analytic Models. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2002.
15. Gagnon, D.J. Understanding the Distribution of Teacher Effectiveness. Ph.D. Thesis, University of New Hampshire, Durham, NH, USA, 2014.
16. Gallagher, H.A. The Relationship between Measures of Teacher Quality and Student Achievement: The Case of Vaughn Elementary. Ph.D. Thesis, The University of Wisconsin–Madison, Madison, WI, USA, 2002.
17. Kane, T.J.; Staiger, D.O. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*; The National Bureau of Economic Research: Cambridge, MA, USA, 2008.
18. Slater, H.; Davies, N.M.; Burgess, S. Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxf. Bull. Econ. Stat.* **2012**, *74*, 629–645. [CrossRef]
19. Chetty, R.; Friedman, J.N.; Rockoff, J.E. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *Am. Econ. Rev.* **2014**, *104*, 2593–2632. [CrossRef]
20. Cunningham, P.L. The Effects of Value-Added Modeling Decisions on Estimates of Teacher Effectiveness. Ph.D. Thesis, The University of Iowa, Iowa City, IA, USA, 2014.
21. Ehlert, M.; Koedel, C.; Parsons, E.; Podgursky, M. The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Stat. Public Policy* **2014**, *1*, 19–27. [CrossRef]
22. Heistad, D.J. Stability and Correlates of Teacher Effects in Grade Two Reading Achievement. Ph.D. Thesis, University of Minnesota, Minneapolis, MN, USA, 1999.
23. Johnson, M.T.; Lipscomb, S.; Gill, B. Sensitivity of teacher value-added estimates to student and peer control variables. *J. Res. Educ. Eff.* **2015**, *8*, 60–83. [CrossRef]
24. Kukla-Acevedo, S. Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Econ. Educ. Rev.* **2009**, *28*, 49–57. [CrossRef]
25. Goel, D.; Barooah, B. *Drivers of Student Performance: Evidence from Higher Secondary Public Schools in Delhi*; Global Labor Organization (GLO): Geneva, Switzerland, 2018.
26. Tobe, P.F. An Investigation of the Differential Impact of Teacher Characteristics and Attitudes on Student Mathematics Achievement Using a Value-Added Approach. Ph.D. Thesis, University of Houston, Houston, TX, USA, 2008.

27. Goldhaber, D.; Hansen, M. *Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions*; National Center for Analysis of Longitudinal Data in Education Research (CALDER): Washington, DC, USA, 2010.
28. Buddin, R. *Measuring Teacher and School Effectiveness at Improving Student Achievement in Los Angeles Elementary Schools*; University Library of Munich: München, Germany, 2011.
29. Muñoz, M.A.; Chang, F.C. The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *J. Pers. Eval. Educ.* **2007**, *20*, 147–164. [CrossRef]
30. Nye, B.; Konstantopoulos, S.; Hedges, L.V. How large are teacher effects? *Educ. Eval. Policy Anal.* **2004**, *26*, 237. [CrossRef]
31. Leigh, A. Estimating teacher effectiveness from two-year changes in students' test scores. *Econ. Educ. Rev.* **2010**, *29*, 480–488. [CrossRef]
32. Ballou, D.; Sander, W.; Wright, P. Controlling for student background in value-added assessment of teachers. *J. Educ. Behav. Stat.* **2004**, *29*, 37–65. [CrossRef]
33. Germuth, A.A. Comparing Results from Value-Added HLM and OLS Models to Assess Teacher Effectiveness. Ph.D. Thesis, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 2003.
34. Clotfelter, C.T.; Ladd, H.F.; Vigdor, J.L. Teacher-student matching and the assessment of teacher effectiveness. *J. Hum. Resour.* **2006**, *XLI*, 778–820. [CrossRef]
35. Goldhaber, D. Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *J. Hum. Resour.* **2007**, *42*, 765–794. [CrossRef]
36. Step by Step Achievement Project. Available online: https://samsun.meb.gov.tr/www/adim-adim-basari-projesi/icerik/570 (accessed on 10 November 2017).
37. Schools and Other Institutions. Available online: https://www.meb.gov.tr/baglantilar/okullar/index.php (accessed on 11 November 2017).
38. The List of Institutions. Available online: https://mebbis.meb.gov.tr/KurumListesi.aspx (accessed on 11 November 2017).
39. Figueiredo Filho, D.B.; Paranhos, R.; Rocha, E.C.; Batista, M.; Silva, J.A.; Santos, M.L.W.D.; Marino, J.G. When is statistical significance not significant? *Braz. Political Sci. Rev.* **2013**, *7*, 31–55. [CrossRef]
40. Gorard, S. Significance testing with incompletely randomised cases cannot possibly work. *Int. J. Sci. Res. Methodol.* **2018**, *11*, 42–51.
41. Cohen, J. The earth is round ($p < 0.05$). *Am. Psychol.* **1994**, *49*, 997–1003.
42. Gorard, S. Damaging real lives through obstinacy: Re-emphasising why significance testing is wrong. *Sociol. Res. Online* **2016**, *21*, 102–115. [CrossRef]
43. Gorard, S. Do we really need confidence intervals in the new statistics? *Int. J. Soc. Res. Methodol.* **2019**, *22*, 281–291. [CrossRef]
44. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350. [CrossRef]
45. White, P.; Gorard, S. Against inferential statistics: How and why current statistics teaching gets it wrong. *Stat. Educ. Res. J.* **2017**, *16*, 5565.
46. Opper, I.M. Does helping John help Sue? Evidence of spillovers in education. *Am. Econ. Rev.* **2019**, *109*, 1080–1115. [CrossRef]
47. Sanders, W.L.; Horn, S.P. Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *J. Pers. Eval. Educ.* **1998**, *12*, 247–256. [CrossRef]
48. Buddin, R.; Zamarro, G. Teacher qualifications and student achievement in urban elementary schools. *J. Urban. Econ.* **2009**, *66*, 103–115. [CrossRef]
49. Wayne, A.J.; Youngs, P. Teacher characteristics and student achievement gains: A review. *Rev. Educ. Res.* **2003**, *73*, 89–122. [CrossRef]
50. Rice, J.K. *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*; Economic Policy Institute: Washington, DC, USA, 2003.
51. Munoz, M.A.; Prather, J.R.; Stronge, J.H. Exploring teacher effectiveness using hierarchical linear models: Student- and classroom-level predictors and cross-year stability in elementary school reading. *Plan. Chang.* **2011**, *42*, 241–273.

52. Hanushek, E.A. The failure of input-based schooling policies. *Econ. J. Lond.* **2003**, *113*, F64–F98. [CrossRef]
53. Krueger, A.B. Economic Considerations and Class Size. *Econ. J. Lond.* **2003**, *113*, F34–F63. [CrossRef]
54. Nye, B.A.; Archilles, C.M.; Boyd-Zaharias, J.; Fulton, B.; Wallenhorst, M. Small is far better. *Res. Sch.* **1994**, *1*, 9–20.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Education Sciences Editorial Office
E-mail: education@mdpi.com
www.mdpi.com/journal/education

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com

ISBN 978-3-0365-5446-4