education sciences

Special Issue Reprint

Innovations in Mathematics Education

Evaluation, Research and Practice

Edited by
Christopher R. Rakes, Robert N. Ronau and Jon Saderholm

mdpi.com/journal/education

MDPI

Innovations in Mathematics Education: Evaluation, Research and Practice

Innovations in Mathematics Education: Evaluation, Research and Practice

Guest Editors

Christopher R. Rakes
Robert N. Ronau
Jon Saderholm

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors
Christopher R. Rakes
Department of Education
University of Maryland
Baltimore County
Baltimore, MD
USA

Robert N. Ronau
College of Education and
Human Development
University of Louisville
Louisville, KY
USA

Jon Saderholm
Education Studies
Department
Berea College
Berea, KY
USA

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Education Sciences* (ISSN 2227-7102), freely accessible at: https://www.mdpi.com/journal/education/special_issues/4UOF4L44UR.

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

Lastname, A. A., & Lastname, B. B. Article Title. *Journal Name*, **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-4929-1 (Hbk)
ISBN 978-3-7258-4930-7 (PDF)
https://doi.org/10.3390/books978-3-7258-4930-7

Contents

Preface . . . . **vii**

Marien Alet Graham, Guillaume Matthys Kruger and Linda van Ryneveld
Developing a Novel Model for ICT Integration in South African Education: Insights from TIMSS
Reprinted from: *Educ. Sci.* **2024**, *14*, 865, https://doi.org/10.3390/educsci14080865 . . . . **1**

Sophie Grace King, Jonathan David Bostic, Toni A. May and Gregory E. Stone
A Usability Analysis and Consequences of Testing Exploration of the Problem-Solving Measures–Computer-Adaptive Test
Reprinted from: *Educ. Sci.* **2025**, *15*, 680, https://doi.org/10.3390/educsci15060680 . . . . **21**

Laura Moore and Robert N. Ronau
Interactive Homework: A Tool for Parent Engagement
Reprinted from: *Educ. Sci.* **2024**, *14*, 103, https://doi.org/10.3390/educsci14010103 . . . . **42**

Eabhnat Ní Fhloinn and Olivia Fitzmaurice
The Impact of the COVID-19 Pandemic upon Mathematics Assessment in Higher Education
Reprinted from: *Educ. Sci.* **2025**, *15*, 449, https://doi.org/10.3390/educsci15040449 . . . . **63**

Stefanie D. Livers, Kristin E. Harbour and Patrick L. Sullivan
(Up)Grading: A (Re)Humanizing Assessment Process with a Focus on Feedback
Reprinted from: *Educ. Sci.* **2024**, *14*, 646, https://doi.org/10.3390/educsci14060646 . . . . **83**

Kathleen Hoffman, Tory H. Williams and Kerrie Kephart
The Use of Guided Reflections in Learning Proof Writing
Reprinted from: *Educ. Sci.* **2024**, *14*, 1084, https://doi.org/10.3390/educsci14101084 . . . . **95**

Sherri K. Prosser and Stephen F. Bismarck
Concrete–Representational–Abstract (CRA) Instructional Approach in an Algebra I Inclusion Class: Knowledge Retention Versus Students' Perception
Reprinted from: *Educ. Sci.* **2023**, *13*, 1061, https://doi.org/10.3390/educsci13101061 . . . . **120**

Hisae Miyauchi and Robinson Thamburaj
Exploratory Study on Geometric Learning of Students with Blindness in Mainstream Classrooms: Teachers' Perspectives Using the Van Hiele Theory
Reprinted from: *Educ. Sci.* **2025**, *15*, 475, https://doi.org/10.3390/educsci15040475 . . . . **133**

Christopher R. Rakes, Angela Wesneski and Rebecca Laws
Building Mathematics Learning through Inquiry Using Student-Generated Data: Lessons Learned from Plan-Do-Study-Act Cycles
Reprinted from: *Educ. Sci.* **2023**, *13*, 919, https://doi.org/10.3390/educsci13090919 . . . . **144**

Preface

This Special Issue examines innovations in mathematics education that improve professional practice. Professional practice in mathematics education demands explicit attention to equity (National Council of Teachers of Mathematics (NCTM), 2018). Equitable, meaningful assessment is an indispensable component of classroom practice; for example, aligning expectations with assessments, providing multiple forms of assessment, delivering high-quality and consistent feedback, consciously attending to the influence of biases and assumptions about student ability, and ensuring the distinction of assessments from grades (Copur-Gencturk et al., 2020; Porter et al., 2007; Webb, 1997). Equitable classrooms also provide opportunities for students to take on leadership roles in their learning and the assessment of their learning. In this way, students gain ownership of their learning and that of their peers. Lessons that promote leadership also promote instructional vision, common goals, and collective collaboration.

In the classroom, professional practice focuses on developing robust mathematics lessons that open the conceptual space to all students (e.g., increased student communication, multiple representations, climate of respect; Sawada et al., 2002) and providing instructional supports to ensure the success of all students (e.g., additional time; NCTM, 2018).

In 'Developing a Novel Model for ICT Integration in South African Education: Insights from TIMSS', Graham, Kruger, and van Ryneveld begin this Special Issue with an examination of the opportunity for students in South Africa to use technology for learning mathematics.

Continuing the focus on the use of technology in mathematics education assessment, King, Bostic, May, and Stone present 'A Usability Analysis and Consequences of Testing Exploration of the Problem-Solving Measures–Computer-Adaptive Test'. In this study, they analyze testing validity of the PSM-CAT test and explore student perceptions of benefits and limitations of the exam.

'Interactive Homework: A Tool for Parent Engagement' presents an exploration by Moore and Ronau on the use of interactive homework for empowering parents to support their child's mathematics learning.

In 'The Impact of the COVID-19 Pandemic upon Mathematics Assessment in Higher Education', Fhloinn and Fitzmaurice move the Special Issue from a purely technological focus to examine how assessment practices were influenced by COVID-19, especially focusing on assessment formats, time limits, academic integrity, and satisfaction with the assessments.

Continuing this focus on assessment innovation, in '(Up)Grading: A (Re)Humanizing Assessment Process with a Focus on Feedback', Livers, Harbour, and Sullivan examine a novel grading approach that values growth in learning and provides students the opportunities to reflect on their learning experiences and give input on their course grade.

Later, in 'The Use of Guided Reflections in Learning Proof Writing', Hoffman, Williams, and Kephart examine the efficacy of student self-assessment for proof writing as a learning tool and focus on the growth of student metacognition regarding their proof writing skills.

In 'Concrete–Representational–Abstract (CRA) Instructional Approach in an Algebra I Inclusion Class: Knowledge Retention Versus Students' Perception', Prosser and Bismarck direct the Special Issue toward instructional practices to improve mathematics learning. They examine the use of manipulatives and a sequential instructional framework and its effect on conceptual understanding and knowledge retention.

Miyauchi and Thamburaj continue the focus on mathematics instruction in 'Exploratory Study on Geometric Learning of Students with Blindness in Mainstream Classrooms: Teachers' Perspectives

Using the Van Hiele Theory', as they examine geometric learning development trajectories for students with blindness.

Finally, in 'Building Mathematics Learning through Inquiry Using Student-Generated Data: Lessons Learned from Plan–Do–Study–Act Cycles', Rakes, Wesneski, and Laws conclude the Special Issue with classroom research conducted by a classroom teacher and teacher candidate. In this study, they describe how the professional development led them to try new instructional ideas in their classes and provide example lesson activities for engaging students in mathematics inquiry.

We are pleased to bring you this set of studies to advance the field of mathematics education. We hope these studies provide a foundation for continuing innovation in the mathematics classroom.

Christopher R. Rakes, Robert N. Ronau, and Jon Saderholm
Guest Editors

education sciences

MDPI

Article

Developing a Novel Model for ICT Integration in South African Education: Insights from TIMSS

Marien Alet Graham [1,*], Guillaume Matthys Kruger [2] and Linda van Ryneveld [3]

1 Department of Early Childhood Education, University of Pretoria, Pretoria 0028, South Africa
2 Department of Science, Mathematics and Technology Education, University of Pretoria, Pretoria 0028, South Africa
3 Comprehensive Online Education Services, University of Pretoria, Pretoria 0028, South Africa; linda.vanryneveld@up.ac.za
* Correspondence: marien.graham@up.ac.za

Abstract: ICT integration in the classroom is viewed as a panacea towards resolving education challenges. A quantitative approach using South African Trends in International Mathematics and Science Studies (TIMSS) 2019 Grades 5&9 data with a positivist philosophical stance was used to explore ICT use. For a long time, most school research took the form of small-scale qualitative studies, such as case studies or critical policy studies; however, research in education has witnessed an increasing demand for high-quality, large-scale quantitative studies such as the current study. TIMSS utilised a two-stage stratified cluster sampling design, sampling schools by size and selecting intact classes. This study focusses on South Africa where 297 schools, 294 mathematics teachers, and 11,903 students were sampled at Grade 5 level, and, at Grade 9 level, the sample consisted of 519 schools, 543 mathematics teachers and 20,829 students. More than 50% of students attend schools lacking computers/tablets, a figure that rises to nearly 90% concerning their availability in classrooms. Less than half of students attend schools utilising online learning systems or providing digital resources. Principals in approximately half the schools indicated shortages/inadequacies in technologically competent staff, and audio-visual and computer technology/software resources. Approximately 80% of teachers expressed interest in future technology integration training for both grades when surveyed. Over half of the students lacked home internet access; however, the majority had access to cell phones and computers/tablets at home. In tailoring this study to the South African context, a novel model for ICT integration emerged which draws upon the Dynamic Model of Educational Effectiveness and the TIMSS curriculum model. Recommendations for improving policy and practice in ICT implementation in schools are structured around the new model.

Keywords: information communication technology; mathematics teaching and learning; TIMSS

1. Introduction

Growing global apprehension surrounds the academic performance of South African mathematics students in schools. Initiatives such as the Trends in International Mathematics and Science Studies (TIMSS) aim to delve into various facets of mathematics achievement. TIMSS measures student performance at Grade 4 and Grade 8 level; however, South Africa participates at Grade 5 and Grade 9 levels due to its overall low performance in previous rounds of TIMSS [1]. Among the 64 countries participating in TIMSS 2019 at Grade 5 level, South Africa ranked amongst the lowest, with a score of 374, notably below the international benchmark of 500 points [2]. At Grade 9 level, among the 39 participating countries, South Africa ranked second to last with a score of 389 [3]. TIMSS sets a minimum benchmark of 400 points, indicating basic proficiency in mathematics. According to Reddy et al. [2,3], only 37% of South African Grade 5 students and 41% of South African Grade 9 students have reached this basic proficiency threshold, suggesting a substantial lack of fundamental mathematical knowledge for both grades. The concerning performance of South African students in mathematics underscores the urgent need for research into what can be improved within the South African educational system concerning mathematics teaching and learning (T&L). There is an expectation that integrating Information Communication Technologies (ICTs) into T&L will translate into better achievement of educational outcomes across schooling systems [4,5]. This research was initiated as a response to reports of significant expenditure on integrating ICT for T&L into South African schools [6]; however, evidence suggests that even investing substantial amounts into ICT integration in South African schools does not significantly improve student achievement [2,3]. This study hypothesizes that ICT investment has not improved student achievement because ICT has not been widely and effectively integrated into mathematics instruction. The research questions (RQs) are: RQ1: What ICTs are being used for mathematics T&L in South Africa? RQ2: Why are (or are not) certain ICTs being used for mathematics T&L in South Africa? RQ3: What models can be used to inform the implementation of effective ICT integration strategies within a South African context? Further investigation into the use of these ICTs in South African classrooms, and even outside the classrooms by the students, is paramount to discovering the problem areas, and, in this research, we used TIMSS 2019 data to explore this matter.

2. Literature Review

The present study builds on a literature foundation describing the benefits of ICT integration in the teaching and learning of mathematics. It also considers the extent to which technology has been made available in South African schools and how that technology has and has not been used. The review concludes with a conceptual framework.

2.1. Importance of ICT Integration in the T&L of Mathematics

Internationally, the benefits of integrating ICTs into the classroom have been empirically proven and established in recent research conducted in countries such as the United States [4], Italy [5], Israel [7], Indonesia [8] and Spain [9]. Engelbrecht and Borba [10] recently published an article on the new developments in using digital technology in mathematics education. In their article, they discuss various topics from redefined learning spaces (e.g., flipped classrooms where students are not introduced to new materials within mathematics lessons but, rather, are expected to work through materials before their lessons (usually made available online beforehand)), to the use of GeoGebra, student collaboration through virtual learning environments and social media, Artificial Intelligence (AI) and hyper-personalisation of learning, and multimodality (e.g., videos, virtual reality (VR), augmented reality (AR)). The authors highlight the benefits of all these new technologies but also list some concerns, such as the digital divide (some individuals (typically in low socio-economic areas) not having access to ICTs). The advantages of ICT integration in mathematics T&L have been documented by many authors (e.g., [10–12]); however, if one does not have access to technologies, how can one use them for the T&L of mathematics? Accordingly, the situation regarding technology diffusion in South African schools is considered next.

2.2. Technology Diffusion in South African Schools

Regarding technology diffusion, which in the context of the current study refers to the degree to which technology is present in South African schools, several national-level ICT initiatives have been implemented, such as the Teacher Laptop project, Sentech Ltd. and the Telkom Internet Project, which aimed to establish Supercentres in over 1300 schools equipped with computers, software, internet connections, and rent-free telephone lines. Initiatives like the eMindset Network and U-Tong portal have been launched to provide digital content resources via satellite television. eSchoolNet, South Africa’s primary educator ICT development programme, aims to empower teachers to integrate ICT into the curriculum confidently. Furthermore, initiatives like Intel Teach to the Future and Microsoft

Partners in Learning offer training programmes covering basic ICT skills, ICT integration, peer coaching, and ICT leadership for education managers. Collaborations between the government and the private sector have led to projects like the Khanya Project and Gauteng Online, providing ICT-based resources in specific provinces. Ongoing national-level projects include the development of strategies from the integrated ICT policy review process, implementation of the SANReN (SANReN stands for "South African National Research Network"; see [13] for details) and TENET (TENET stands for "Tertiary Education and Research Network" of South Africa; see [13] for details), the "Broadcasting Digital Migration Policy for South Africa" (for more details on "Broadcasting Digital Migration Policy for South Africa", see [14]), and SA Connect (SA Connect is a "National wide government broadband connectivity project aimed at connecting government facilities" [15] (para. 1]) [16]. Despite all these initiatives, recent research in South African schools still reports that many South African schools do not have the necessary ICTs; for example, in the recent study by Mokotjo and Mokhele [17] on the challenges of integrating GeoGebra in the T&L of mathematics in South African secondary schools, they reported that there were insufficient resources in the schools (mostly due to security issues—schools being robbed and vandalized), causing teachers to become demotivated and disadvantaging students' learning of mathematics.

South Africa is categorised as an upper–middle–income country with high levels of compulsory school enrollment and significantly higher annual government expenditure on education compared to many other nations [6]. The World Bank [6] provides a notable example: In Sri Lanka, a lower–middle–income country, the average expenditure per primary school-aged child from 2015 to 2019 was approximately PPP$615; PPP stands for purchasing power parity. Despite this comparatively modest investment, Sri Lanka achieved remarkable results, with a learning poverty rate of only 15%. In stark contrast, South Africa allocated nearly PPP$2400 per primary school-aged child during a similar timeframe. However, the learning poverty rate in South Africa stood at a staggering 79%. This figure is akin to that of much poorer Guinea, where the expenditure per child was a mere PPP$144. Thus, despite significant investment in education, South Africa's outcomes are extremely poor, and many students lack basic mathematics skills [2,3]. Accordingly, ICT use, inside and outside the classroom, by teachers and South African students, at primary and secondary levels, warrants investigation. While the majority of studies have focused on the secondary level, investigating ICT use at the primary school level is essential. Acquiring basic ICT skills at a young age lays the foundation for more advanced ICT literacy skills later in life. These competencies equip young people for future technological use and critical reasoning. Accordingly, the next section considers the uses of ICT in South African mathematics T&L at primary and secondary levels.

2.3. Uses of ICT in South African Mathematics T&L

In South Africa, studies involving primary schools are considered first, followed by studies in secondary schools. Mwapwele and colleagues [18] analysed the baseline data from the ICT4E initiative (ICT4E stands for "Information and Communications Technology for Education"; see [19] for more detail), which encompassed data from 197 teachers from 24 primary and secondary rural schools across seven of the nine provinces of South Africa. They found that, despite some financial, technical, and digital skills challenges at their schools, teachers were optimistic about the advantages that ICT integration into T&L could bring. Mahwai and Wotela [19] also used the ICT4E project data, but only those of rural schools in Seshego Circuit, and concluded that the promise of successful ICT integration through this project was unsuccessful as the aims and objectives of the ICT4E project had not been achieved. In the same year, Dlamini [20] published the results of a large quantitative study (837 respondents from 133 schools) undertaken in Gauteng and concluded that teachers' limited technological pedagogical knowledge and limited experience in integrating computers into the classroom has had a negative impact on ICT uptake; they used the analytical framework of the Second Information Technology in Education Study

(SITES) in their investigation. Ramafi [21] analysed the data from 59 questionnaires and five interviews collected from public school teachers and identified six factors influencing ICT use: (i) government support, (ii) security measures provided for the ICT tools, (iii) teacher efficacy, (iv) learner efficacy, (v) state of ICT tools, (vi) and the use of ICT tools. Graham and colleagues [22] explored the reasons behind why (or why not) South African primary and secondary school teachers integrate ICTs in their classrooms using the UTAUT as a theoretical lens. They concluded that teachers only viewed technology integration as beneficial when it increased productivity and social influence. Using the same data set, these authors published a quantitative study one year later that investigated which ICTs were being used most in South African mathematics classrooms [23]. The researchers discovered that laptops/computers were the most frequently utilised ICT, with data projectors following as the next most frequently utilised ICT. They advised that professional development initiatives should prioritise instructing teachers on how to incorporate ICTs into their classrooms in a way that requires fundamental pedagogical adjustments.

Some examples of studies that only considered the secondary school level in South Africa are considered next. Ojo and Adu [24] conducted a study in the Eastern Cape Province using self-developed questionnaires and data from 450 students and 150 teachers. It was determined that the most abundant ICT resources in every chosen school were mobile phones, and these were utilised by pupils to exchange ideas and information regarding their courses and download pertinent information. Chisango and colleagues [25] adopted a qualitative research approach to explore rural secondary school teachers' perceptions of the use of ICTs in T&L and found that although teachers had a positive attitude towards the adoption of ICTs and were ready to integrate ICTs in T&L, they lacked the requisite ICT skills. Filita and Jita [26] conducted a study on teachers' perspectives on ICT integration in the teaching of Sesotho (one of South Africa's official languages) by conducting interviews, using the Technological Pedagogical Content Knowledge (TPACK) framework, and concluded that teachers lacked technological knowledge, and that the lack of Sesotho content in ICT resources negatively affected ICT adoption. More recently, Zenda and Dlamini [27] examined the factors that influence teachers' adoption of ICTs in rural secondary schools using a survey and found that having ICT infrastructure and a training policy in place were some of the reasons why teachers adopted ICTs in T&L; the modified UTAUT was used to guide this investigation. In 2024, Mnisi and colleagues [28] conducted a study in Gauteng using interviews and open-ended questionnaires with ten teachers and one curriculum specialist, and concluded that most schools are improving ICT use, but the biggest factor still hindering ICT integration is a lack of internet access in classrooms, hindering teachers from making full use of ICTs.

Some examples of studies that only considered the primary school level in South Africa are considered next. Saal and colleagues [29,30] used Grade 5 TIMSS 2015 data to explore the use of ICT in T&L in mathematics and found that almost 90% of South African students were taught by teachers who did not even have computers in their mathematics classrooms. This is a devasting finding, because they also found students who were in mathematical classes with computers significantly outperformed those without computers available to them. These same authors published a qualitative case study at primary school level, using the UTAUT (UTAUT stands for "Unified Theory of Acceptance and Use of Technology"; see [31] for more details) as a theoretical lens, to investigate the elements facilitating and hindering the integration of educational technology in mathematics education in economically disadvantaged areas of South Africa, and found that facilitating conditions (such as adequate technological infrastructure and qualified information technology technicians), and social influence (such as other teachers using ICTs in their classrooms) had the greatest impact on actual ICT use in the classroom of all the UTAUT constructs [32]; the interested reader is referred to Saal and colleagues [32] for more details on the UTAUT. Kolobe and Mihai [33] conducted an investigation into how ICTs are used as an intervention tool for progressed learners in T&L of English First Additional Language in Gauteng and concluded that ICTs had the potential to reduce failure rates, minimizing the number of learners who need to be progressed without meeting promotional requirements; the TPACK was used as theoretical framework. In a more recent study, Mahlo and Waghid [34] explored ICT integration among teaching in township public primary schools using lesson observations and interviews and concluded that the influence of personal conversion factors (ICT skills obtained through a community of practice and university training), had created the capabilities for teachers to use ICTs for T&L purposes, although to a limited extent.

The above literature review considered RQ1: What ICTs are being used for mathematics T&L in South Africa? RQ2: Why are (or are not) certain ICTs being used for mathematics T&L in South Africa? RQ3: What models can be used to inform the implementation of effective ICT integration strategies within a South African context? For RQ1 and RQ2, studies from many researchers were considered on what ICTs are being used, why they are (or are not) used (the latter speaking to ICT integration challenges and barriers), whereas, for RQ3, some of the frameworks and models used to inform effective integration were mentioned (e.g., TPACK, UTAUT, SITES framework). Moreover, regarding RQ3, some current studies purely focus on formulating ICT integration frameworks that are effective within a South African context; for example, ref. [35] formulated an ICT integration framework (described by the authors as an extension of the Technology Acceptance Framework [TAM]), responsive to the challenges that led to low ICT integration and more effective ICT integration in Gauteng schools. The literature points to the questions being posed in the current study as to what, why (or why not) ICTs are being used and what models can be used to inform ICT integration in South African schools as topical research.

2.4. Conceptual Framework: Towards a Model for the Integration of ICT in School

In tailoring this study to the South African context, an adaptation of the Dynamic Model of Educational Effectiveness (DMEE) was utilised as the foundational conceptual framework [36]. (The DMEE's contextual factors were redefined to incorporate South African-specific educational policies and ICT provisions, which differ significantly from the original model's European context. We tailored the school-level factors to reflect the distinct challenges related to technology integration faced by South African schools, such as limited access to digital resources and a lack of technologically competent staff. Classroom-level factors were adjusted to account for the varied levels of ICT availability in South African classrooms and the impact of this on teaching and learning practices. The student-level factors were revised to consider the external influences affecting South African students, such as socioeconomic barriers to technology access at home) This choice was made because the delineated levels of educational effectiveness within the DMEE align closely with the categories outlined in the TIMSS curriculum model [37]. The DMEE endeavours to delineate the factors correlated with educational effectiveness across four interconnected levels: context-related factors (e.g., national and regional educational policies), school-related factors (e.g., teaching and learning policies within schools), factors pertinent to the classroom and educators, and those related to students themselves [38]. This framework, illustrated in Figure 1, elucidates the interconnectedness of the DMEE, demonstrating how each level exerts either a direct or indirect influence on the others within the model.

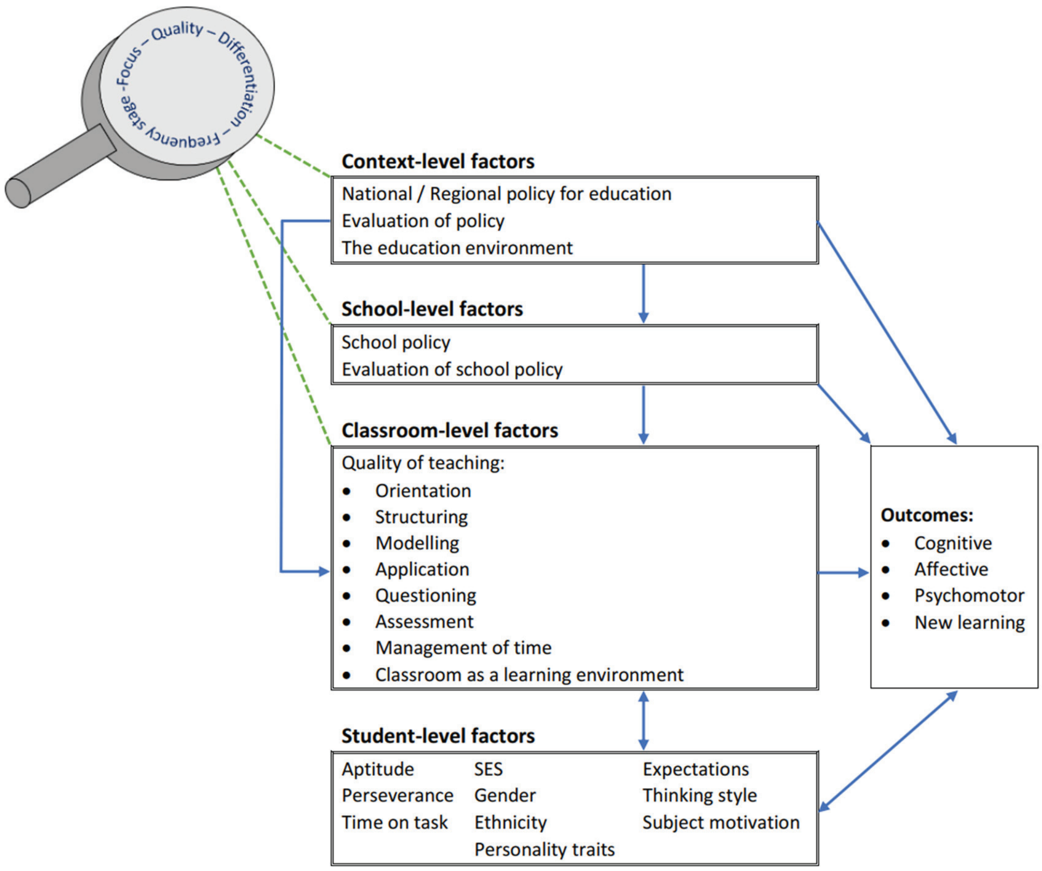


Figure 1. The DMEE [39], pp. 77, 150.

Upon reviewing the DMEE [36] alongside the IEA guidelines for researchers utilizing TIMSS data [40], it became evident that this study must also integrate the research areas outlined by the IEA when analyzing the TIMSS 2019 data. TIMSS studies are structured around a curriculum model that emphasises three key dimensions of teaching and learning: the intended curriculum, the implemented curriculum, and the attained curriculum [37], as depicted in Figure 2. This curriculum model serves as the cornerstone of TIMSS investigations and harmonises effectively with the levels of educational effectiveness delineated in the DMEE.

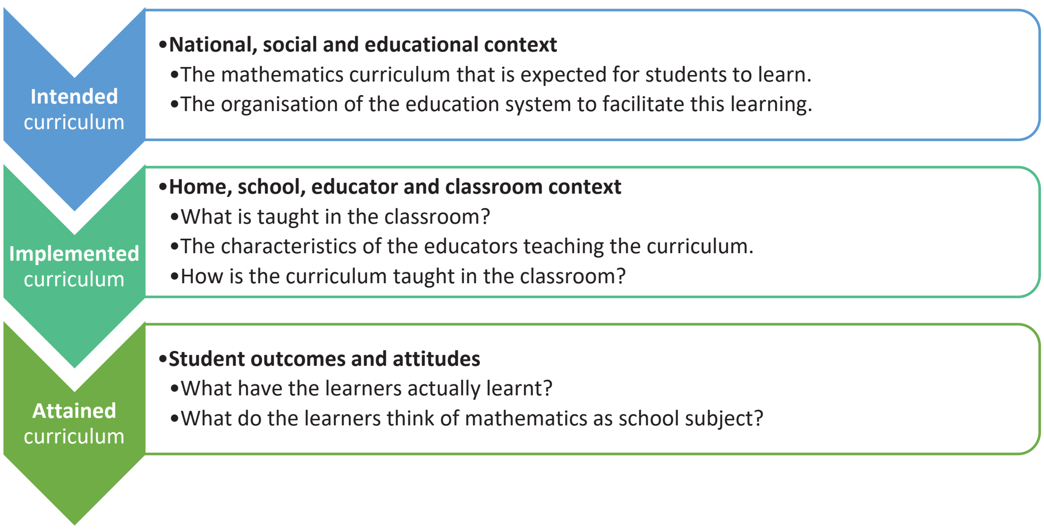


Figure 2. The TIMSS curriculum model. Note. Adapted from TIMSS framework of intended implemented and attained curriculum. Image, by [41]. CC BY-SA 4.0.

There are four distinct areas of research on which the TIMSS context questionnaires focus [42] and they are depicted in Table 1.

Table 1. TIMSS context questionnaires research areas.

TIMSS Context Questionnaire	Areas of Research
Mathematics curriculum questionnaire Country context	The mathematics curriculum as established by the Department of Education of the participating country.
ScQ School context	The educational environment in which both the student and instructor operate; this consists of elements like resource accessibility, the perception of safety on campus, and the support received from school administration.
TQ Educator and classroom context	The educator's background and the impact they have on the efficacy of teaching and learning in the classroom are factors to consider. This encompasses the educator's teaching methods, the practical implementation of acquired knowledge, and their educational credentials.
HQ Home context	Details concerning educational resources available at home, perspectives on the parents' highest level of education and employment circumstances, evaluations of their child's school, attendance record in preprimary education programmes, prioritisation of literacy and numeracy activities at home, and the parents' literacy and numeracy proficiency at the start of the academic year are all pertinent information.
StQ Student context	Student-specific information, including student-related context such as the student's home environment, academic motivation and application, and parental background and support availability, is encompassed within this category.

The relationship between the DMEE, the TIMSS curriculum model and the TIMSS context questionnaires is depicted in Table 2.

Table 2. Relationship between the DMEE, the TIMSS curriculum model and the TIMSS context questionnaires and assessment.

DMEE	TIMSS Curriculum Model	TIMSS Context Questionnaires and Assessment
Context-level factors (country and region)	Intended curriculum	Mathematics curriculum—Mathematics Curriculum Questionnaire
School-level factors	Implemented curriculum	School context—ScQ
Classroom-level factors	Implemented curriculum	Classroom and educator context—TQ
Home-level factors	Implemented curriculum	Home context—HQ
Student-level factors	Attained curriculum	Student achievement in TIMSS –Mathematics assessment Student context and background—StQ

While South African schools adhere to the uniform mathematics curriculum known as CAPS [43], the approach to implementation and the contextual factors vary greatly among schools nationwide. Given the diverse demographics of both schools and students, a simplistic perspective on mathematics achievement would be inadequate. Moreover, schools and students are embedded within social contexts, necessitating an examination of TIMSS results within broader community frameworks. The conceptual framework of this study aims to integrate these complex considerations, as illustrated in Figure 3.

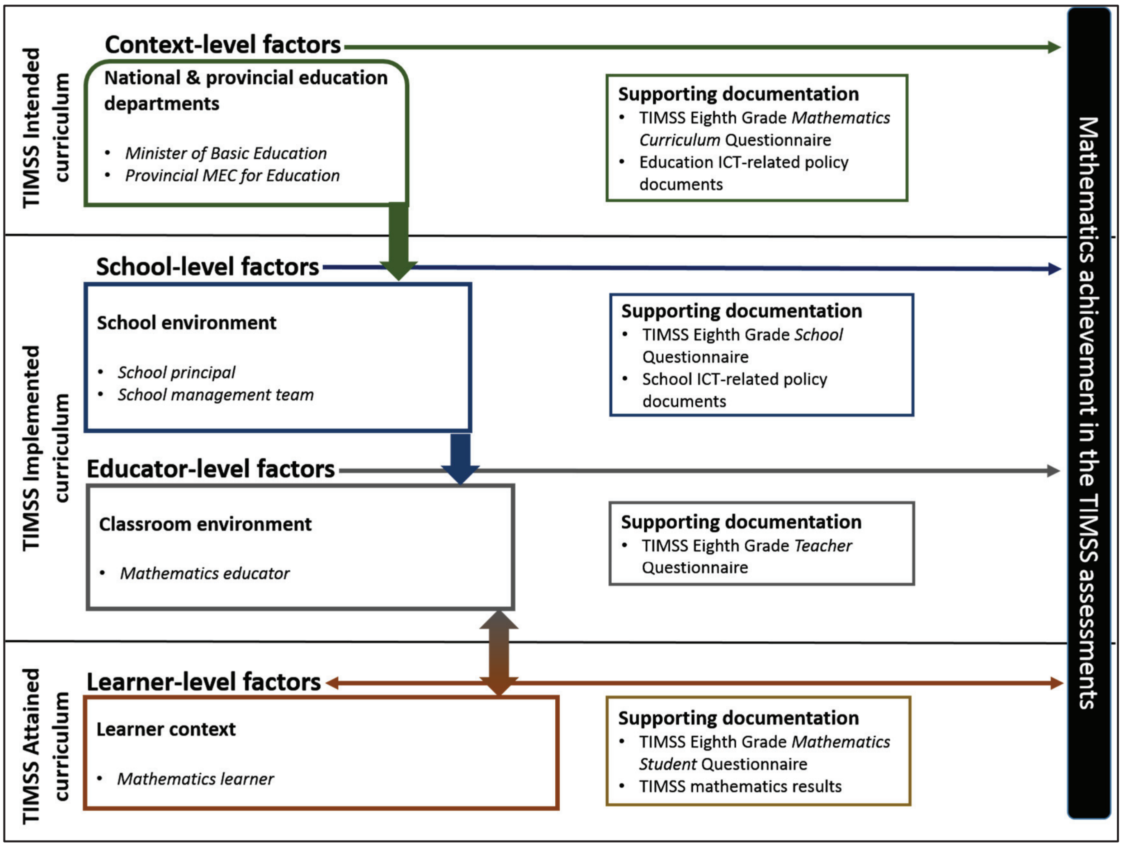


Figure 3. Conceptual framework of this study as adapted from [37,39].

3. Methodology

3.1. Research Approach and Design

A quantitative approach was followed with a positivist philosophical stance that relied on measurement and reason, and that knowledge was revealed from neutral and measurable (quantifiable) observations. Positivism builds on verifying a-priori hypotheses and experimentation by operationalizing variables and measures [44]. This study hypothesised that ICT investment has not improved South African students' mathematics achievement because ICT has not been widely and effectively integrated into mathematics instruction and, using the variables listed in Tables 3 and 4 in Section 4, we aimed to verify this a-priori hypothesis. The research design is a secondary data analysis [45], as this study used secondary data from TIMSS 2019. TIMSS data are cross-sectional in nature and not longitudinal, as TIMSS analyse data from different participants each cycle, i.e., they do not follow the same group of individuals over time [46]. As the focus of the current study is on South Africa, a purposive sampling technique [47] from within TIMSS 2019 was used by only selecting and working with the South African TIMSS 2019 mathematics data.

3.2. Participants

For South Africa, at Grade 5 level, the realised sample was 297 schools, 294 mathematics teachers, and 11,903 students [2]; whereas, at Grade 9 level, the sample consisted of 519 schools, 543 mathematics teachers, and 20,829 students [3]. The TIMSS 2019 employed a two-stage stratified cluster sampling methodology. In the first stage, schools were selected based on their size. In the second stage, one or more intact classes from the grade level of interest from each school that participated were chosen [48].

3.3. Instruments and Quality Assurance

TIMSS 2019 included a series of context surveys for various stakeholders, which were used in the present study to investigate the research questions [42]. Specifically, the principals answered a school questionnaire (ScQ), the teachers answered a teacher questionnaire (TQ), and the students answered the student questionnaire (StQ). Selected items from these context questionnaires relating to ICTs were analysed, and the items are displayed in Tables 3 and 4 in Section 4. All three authors of the manuscript analysed the context questionnaires, reaching a consensus on which items to use, specifically those related to ICT use. It should be noted that the selected questions are predominantly inventory-type (e.g., the number of computers available in the schools) and frequency–of–use questions (e.g., how often ICTs are used in class), rather than attitude-type questions. However, there are some opinion-based questions; for example, principals were asked to indicate how much the school's capacity to provide instruction is affected by a shortage or inadequacy of ICTs. This question is not purely inventory-type (i.e., whether the ICT is available or not), but rather relies on the subjective opinions of principals regarding the negative impact of potential ICT shortages. These questionnaires were designed within the framework of the TIMSS curriculum model [37]. In terms of quality assurance, TIMSS 2019 implemented various measures to ensure the reliability and validity of the assessment [48,49].

3.4. Data Analysis

The IEA IDB analyzer was used, supported by SPSS, to conduct the statistical analyses, which included descriptive statistics such as percentages, measures of location (mean, median), and measures of spread (standard deviation, interquartile range). Questions were analysed from the ScQ, TQ, and StQ. All variables that were considered, and their responses at Grade 5 and Grade 9 level, are shown in Tables 3 and 4 of Section 4. Multiple imputation was used to address missing values because it is widely regarded as the most valid method of addressing missingness, even when the data are not missing at random [50]. A last note is, when interpreting the results, it is important to note that, in TIMSS studies, the student is the unit of analysis; thus, say we interpret the responses of the principals (ScQ), it would be interpreted as the percentage of students attending the school, and not the number of principals—the same applies to the TQ answered by the teachers.

4. Results

Tables 3 and 4 display the TIMSS variables considered in the study, along with the responses at both Grade 5 and Grade 9 level, respectively. Table 3 shows the variables relating to RQ1: What ICTs are being used for mathematics T&L in South Africa?, whereas Table 4 shows the variables relating to RQ2: Why are (or are not) certain ICTs being used for mathematics T&L in South Africa? In Tables 3 and 4, variables with four response options were simplified to two response options for easier interpretation. The downward arrow ↘ indicates the grouping of the first two response options into one category, while the upward arrow ↗ indicates the grouping of the last two response options into another category.

Table 3. The TIMSS questions, variable names, and responses relating to RQ1.

TIMSS Question and Variable Name		Grade 5			Grade 9		
TQ (Grade 5: Answered by 294 Mathematics Teachers; Grade 9: Answered by 543 Mathematics Teachers)							
"If yes to having access to a computer or tablet in class, how often do you do activities on computers during mathematics lessons to support learning for":	"Whole class" Grade 5: ATBM04CA Grade 9: BTBM17CA	"Never or almost never" (17.6%) "Once or twice a month" (34.1%) "Once or twice a week" (48.3%) "Every or almost every day" (0.0%)	↘ ↗ ↘ ↗	Never to 1–2 pm (51.7%) 1–2 pw to always (48.3%)	"Never or almost never" (48.4%) "Once or twice a month" (24.9%) "Once or twice a week" (18.4%) "Every or almost every day" (7.9%)	↘ ↗ ↘ ↗	Never to 1–2 pm (73.7%) 1–2 pw to always (26.3%)

Table 3. *Cont.*

TIMSS Question and Variable Name		Grade 5			Grade 9		
“If yes to having access to a computer or tablet in class, how often do you do activities on computers during mathematics lessons to support learning for”:	“Low-performing students” Grade 5: ATBM04CB Grade 9: BTBM17CB	“Never or almost never” (29.8%) “Once or twice a month” (31.0%) “Once or twice a week” (36.3%) “Every or almost every day” (2.9%)	↘ ↗ ↘ ↗	Never to 1–2 pm (60.8%) 1–2 pw to always (39.2%)	“Never or almost never” (56.3%) “Once or twice a month” (23.8%) “Once or twice a week” (7.7%) “Every or almost every day” (12.2%)	↘ ↗ ↘ ↗	Never to 1–2 pm (80.1%) 1–2 pw to always (19.9%)
	“High-performing students” Grade 5: ATBM04CC Grade 9: BTBM17CC	“Never or almost never” (28.4%) “Once or twice a month (21.7%) “Once or twice a week” (44.6%) “Every or almost every day” (5.3%)	↘ ↗ ↘ ↗	Never to 1–2 pm (50.1%) 1–2 pw to always (49.9%)	“Never or almost never” (52.6%) “Once or twice a month” (25.0%) “Once or twice a week” (14.4%) “Every or almost every day” (8.1%)	↘ ↗ ↘ ↗	Never to 1–2 pm (77.5%) 1–2 pw to always (22.5%)
	“Students with special needs” Grade 5: ATBM04CD Grade 9: BTBM17CD	“Never or almost never” (38.2%) “Once or twice a month” (22.6%) “Once or twice a week” (31.0%) “Every or almost every day” (8.2%)	↘ ↗ ↘ ↗	Never to 1–2 pm (60.8%) 1–2 pw to always (39.2%)	“Never or almost never” (57.0%) “Once or twice a month” (20.9%) “Once or twice a week” (12.6%) “Every or almost every day” (9.5%)	↘ ↗ ↘ ↗	Never to 1–2 pm (77.9%) 1–2 pw to always (22.1%)

Note: Never to 1–2 pm = Never to once or twice per month; 1–2 pw to always = Once to twice per week to always. All direct quotes are from the TIMSS questionnaires [51,52].

Table 1 shows the variables related to RQ1: What ICTs are being used for mathematics T&L in South Africa? Teachers (using TQ) reported using computers or tablets for T&L for the whole class more at Grade 5 level (approximately half reported using it “1–2 pw to always”) as opposed to Grade 9 level where only about a quarter reported using it “1–2 pw to always”. In Section 3.4, we noted that the student is the unit of analysis, so it is more accurate to say that, for Grade 5, almost half of the students attended classes where the teacher used computers or tablets one or two times per week to always. In contrast, only about a quarter of Grade 9 students were taught in classes where this occurred one or two times per week to always. Using computers and tablets for low-performing students again showed a higher percentage of use at Grade 5 level than Grade 9, and a similar pattern is seen when they reported on the use for high-performing students. When asked about the use of computers and tablets for students with special needs, again, the Grade 5 percentage was higher (approximately 40%) when reporting using it “1–2 pw to always” compared to Grade 9 teachers (approximately 20%).

Table 4. The TIMSS questions, variable names and responses relating to RQ2.

TIMSS Question and Variable Name	Grade 5	Grade 9
ScQ (Grade 5: Answered by 297 Principals; Grade 9: Answered by 519 Principals)		
“How many computers (including tablets and iPads) does your school have for use by Grade 5/9 students?” Grade 5: ACBG07 Grade 9: BCBG07	Mean = 12.26 SD = 20.42 Median = 0.00 * Interquartile range = 20.00	Mean = 21.79 SD = 42.45 Median = 0.00 ** Interquartile range = 30.00
“Does your school use an online learning management system to support learning (e.g., educator –student communication, management of grades, student access to course materials)?” Grade 5: ACBG09 Grade 9: BCBG09	Yes (12.6%) No (87.4%)	Yes (25.5%) No (74.5%)
“Does your school provide students with access to digital learning resources (e.g., books, videos)?” Grade 5: ACBG12 Grade 9: BCBG12	Yes (39.9%) No (60.1%)	Yes (49.7%) No (50.3%)

Table 4. *Cont.*

TIMSS Question and Variable Name		Grade 5		Grade 9	
"How much is your school's capacity to provide instruction affected by a shortage or inadequacy of":	"Technologically competent staff" Grade 5: ACBG13AF Grade 9: BCBG13AF	"Not at all" (13.0%) "A little" (27.6%) "Some" (40.7%) "A lot" (18.7%)	None to a little (40.6%) Some to a lot (59.4%)	"Not at all" (13.8%) "A little" (32.1%) "Some" (38.8%) "A lot" (15.2%)	None to a little (46.0%) Some to a lot (54.0%)
	"Audiovisual resources for delivery of instruction (e.g., interactive white boards, digital projectors)" Grade 5: ACBG13AG Grade 9: BCBG13AG	"Not at all" (29.7%) "A little" (20.5%) "Some" (16.3%) "A lot" (33.5%)	None to a little (50.2%) Some to a lot (49.8%)	"Not at all" (20.8%) "A little" (27.4%) "Some" (29.7%) "A lot" (22.1%)	None to a little (48.2%) Some to a lot (51.8%)
	"Computer technology for teaching and learning (e.g., computers or tablets for student use)" Grade 5: ACBG13AH Grade 9: BCBG13AH	"Not at all" (29.7%) "A little" (19.5%) "Some" (14.0%) "A lot" (36.8%)	None to a little (49.2%) Some to a lot (50.8%)	"Not at all" (25.9%) "A little" (24.7%) "Some" (25.4%) "A lot" (24.0%)	None to a little (50.6%) Some to a lot (49.4%)
	"Computer software/applications for mathematics instruction" Grade 5: ACBG13BB Grade 9: BCBG13BB	"Not at all" (29.0%) "A little" (24.1%) "Some" (18.8%) "A lot" (28.1%)	None to a little (53.1%) Some to a lot (46.9%)	"Not at all" (25.2%) "A little" (25.5%) "Some" (29.0%) "A lot" (20.2%)	None to a little (50.8%) Some to a lot (49.2%)
TQ (Grade 5: answered by 294 mathematics teachers; Grade 9: answered by 543 mathematics teachers)					
"Students in this class have computers (including tablets) available to use during their mathematics lessons," Grade 5: ATBM04A Grade 9: BTBM17A		Yes (9.1%) No (90.9%)		Yes (12.3%) No (87.7%)	
"If yes to having access to a computer or tablet in class, what access do they have":	"Each student has a computer" Grade 5: ATBM04BA Grade 9: BTBM17BA	Yes (4.5%) No (95.5%)		Yes (21.3%) No (78.7%)	
	"The class has computers that students can share" Grade 5: ATBM04BB Grade 9: BTBM17BB	Yes (38.3%) No (61.7%)		Yes (12.4%) No (87.6%)	
	"The school has computers that the class can use sometimes" Grade 5: ATBM04BC Grade 9: BTBM17BC	Yes (83.6%) No (16.4%)		Yes (53.7%) No (46.3%)	
"In the past two years, have you participated in professional development in integrating technology into mathematics instruction?" Grade 5: ATBM09AD Grade 9: BTBM22AD		Yes (44.8%) No (55.2%)		Yes (50.6%) No (49.4%)	
"Do you need future professional development in integrating technology into mathematics instruction?" Grade 5: ATBM09BD Grade 9: BTBM22BD		Yes (86.1%) No (13.9%)		Yes (85.0%) No (15.0%)	
StQ (Grade 5: answered by 22,903 students; Grade 9: answered by 20,829 students)					
"Do you have any of these things at your home?"	"A computer or tablet" Grade 5: ASBG05A Grade 9: BSBG05A	Yes (56.9%) No (43.1%)		Yes (52.2%) No (47.8%)	
	"Internet connection" Grade 5: ASBG05D Grade 9: BSBG05D	Yes (36.2%) No (63.8%)		Yes (43.0%) No (57.0%)	
	"Your own cell phone" Grade 5: ASBG05E Grade 9: BSBG05E	Yes (67.8%) No (32.2%)		Yes (79.1%) No (20.9%)	
	"Electricity" Grade 5: ASBG05G Grade 9: BSBG05G	Yes (83.5%) No (15.8%)		Yes (94.0%) No (6.0%)	

Note: Never to 1–2 pm = Never to once or twice per month; 1–2 pw to always = Once to twice per week to always. All direct quotes are from the TIMSS questionnaires [53–56]. * More than half (59.9%) of responses were zero. ** More than half (52.4%) of responses were zero.

Table 2 shows the variables relating to RQ2: Why are (or are not) certain ICTs being used for mathematics T&L in South Africa? Many of these questions are inventory-type questions about whether the ICTs are available in the first place, because a reason for not using ICTs could be not having access to them. There are also some questions about teacher professional development because if a teacher does not know how to use ICTs, they most probably will not use them. For school-level (TQ [Table 3] and ScQ [Table 4]), the ScQ will be considered first and it can be seen that the average number of computers (including tablets and iPads) available to Grade 5 students (mean = 12.26) is significantly lower than for Grade 9 students (mean = 21.79); however, the median is the same for both grades with the median of zero indicating that, for both Grade 5 and Grade 9, more than half of the responses were that there are zero computers available to students. Recall that, in Section 3.4, we mentioned that the student is the unit of analysis, so, using the percentages, the more accurate way of reporting these results would be to say that 59.9% of Grade 5 students attended classes where the principals reported there were no computers available to students, whereas this percentage is 52.4% for the Grade 9 learners. Regarding the use of online learning management systems to support learning, Grade 9 (25.5%) indicated higher usage compared to Grade 5 (12.6%). Recall that, in Section 3.4, we mentioned that the student is the unit of analysis, so the more accurate way of reporting these results would be to say that about one quarter of Grade 9 students attended class where the principals reported that the schools are using online learning management systems, compared to only 12.6% of Grade 5 students. When asked whether the school provides students with digital learning resources, almost 40% of Grade 5 students attended schools where this is the case, whereas, for Grade 9, this percentage was approximately 50%. When principals were questioned about how much their schools' capacity to provide instruction is affected by a shortage of different things related to technology, for both grades, the ratios were about 50–50 for all technology-related concepts (technologically competent staff, audio-visual and computer technology/software resources) indicating about a 50–50 split between them affecting instruction and not affecting instruction. Next, the responses to the TQ are considered. For Grade 5, approximately 90% of students attended mathematics lessons with no computers (including tablets) available in the mathematics classrooms, and this percentage was also approximately 90% for Grade 9. When asked whether teachers participated in professional development in integrating technology into mathematics instruction in the past two years, the responses were roughly 50–50 for both grades; however, this ratio changed dramatically when teachers were asked whether they would like to go for future professional development on this topic, where the ratio is approximately 20–80 for no–yes for both grades. For the student-level (StQ [Table 4]), for both grades, the majority of students had access to electricity at home. Regarding internet access at home, for both grades, more than half of the students indicated that they did not have it. When asked whether they owned a computer (or tablet) and their own cell phone, the percentage for cell phones was higher for both grades than for a computer/tablet, with all the percentage yes responses being above 50%, meaning more than half of Grade 5 and Grade 9 students had a computer/tablet and cell phone at home.

For RQ3, "What models can be used to inform the implementation of effective ICT integration strategies within a South African context?," there were no TIMSS questions that addressed this question; however, different models were considered in Section 2.1 using the available literature, and, during our study, a novel model for incorporating ICT into schools emerged, which are considered and discussed in detail in Sections 2.2 and 6.

5. Discussion

For RQ1, "What ICTs are being used for mathematics T&L in South Africa?", the results showed that computers and tablets were being used more often at Grade 5 level as opposed to Grade 9 level. These results could be attributed to the fact that many studies have shown the benefit of ICT integration in T&L for younger learners [57,58]. Unfortunately, the percentage use of other ICTs (e.g., interactive whiteboards) can not be discussed, as the TIMSS instruments do not go into that level of detail regarding ICT use.

For RQ2, "Why are (or are not) certain ICTs being used for mathematics T&L in South Africa?", the reasons seem to be three-fold. Many South African schools do not have access to ICTs (Reason 1), many students do not have access to ICTs at home (Reason 2), and teachers do not know how to integrate ICTs into T&L properly (Reason 3). Reason 1 was derived from the principals' responses for both Grade 5 and Grade 9, which made it evident that more than half of South African students are attending schools where there are no computers or tablets available. These results are concerning, given the substantial investment in ICT procurement, as they point to the fact that funds have not been fully utilised in their intended way which warrants investigation into how these funds were spent. When reporting on the use of online learning management systems to support learning and whether the school provided students with digital learning resources, very low percentages were reported by principals. Again, these low percentages, considering the financial investments made [6,16], are troubling. For Reason 2, although South African students have access to electricity at home, other ICT-related issues came to light; for example, regarding internet access at home, it is concerning that, for both grades, more than half of the students indicated that they do not have it. It should be noted that electricity was included in the analysis due to South Africa's ongoing electricity issues, which negatively affects the educational milieu [59]; one cannot use technology without electricity. For Reason 3, it came to light that many teachers expressed the desire to attend more professional development programmes on integrating ICTs in mathematics instruction and that approximately only half of teachers had undergone recent training in it. Similar results of South African teachers needing more professional development in ICT integration have been found by other researchers [22,28]. These conditions should be alarming to all stakeholders because they highlight significant gaps in the effective use of ICTs for mathematics teaching and learning.

For RQ3, "What models can be used to inform the implementation of effective ICT integration strategies within a South African context?," there were no relevant questions in the TIMSS dataset. However, various models and frameworks (e.g., TPACK, UTAUT, SITES framework) were reviewed in Section 2.1 based on the existing literature. Additionally, during our study, a new model for incorporating ICT into schools emerged, which draws upon the Dynamic Model of Educational Effectiveness and the TIMSS curriculum model. This novel model is thoroughly examined and discussed in Sections 2.2 and 6. We believe this model advances the theoretical framework for ICT integration in South African schools, as no single model has proven to be entirely effective. Organised around the Four Zones Model, recommendations emphasise the need for tailored support and continuous professional development at all levels of the education system. National and provincial education departments must provide substantial support and contextualised resources to schools, while school management should ensure equitable access to ICT resources and implement designated time slots for their use. Educators require targeted professional development programs focusing on fundamental computer literacy and practical aspects of ICT integration, while students should have access to high-quality educational software and equitable opportunities for engagement. By implementing these recommendations, South African schools can effectively harness the potential of ICT to enhance T&L outcomes, preparing students for success in a technology-driven society.

6. Improving the Integration of ICT in Schools to Show an Increased Educational Return on Investment

Upon analysing the research findings within the framework of this study, a novel model for incorporating ICT into schools emerged. This model draws upon the Dynamic Model of Educational Effectiveness (DMEE) by [36] and the TIMSS curriculum model [37]. Its objective is to guide the integration of ICT for T&L in schools by delineating distinct zones of impact, key stakeholders, and curriculum expectations throughout the ICT implementation process. This proposed model, termed the "Four Zones Model for the Integration of ICT in Schools" (Four Zones Model), is illustrated in Figure 4.

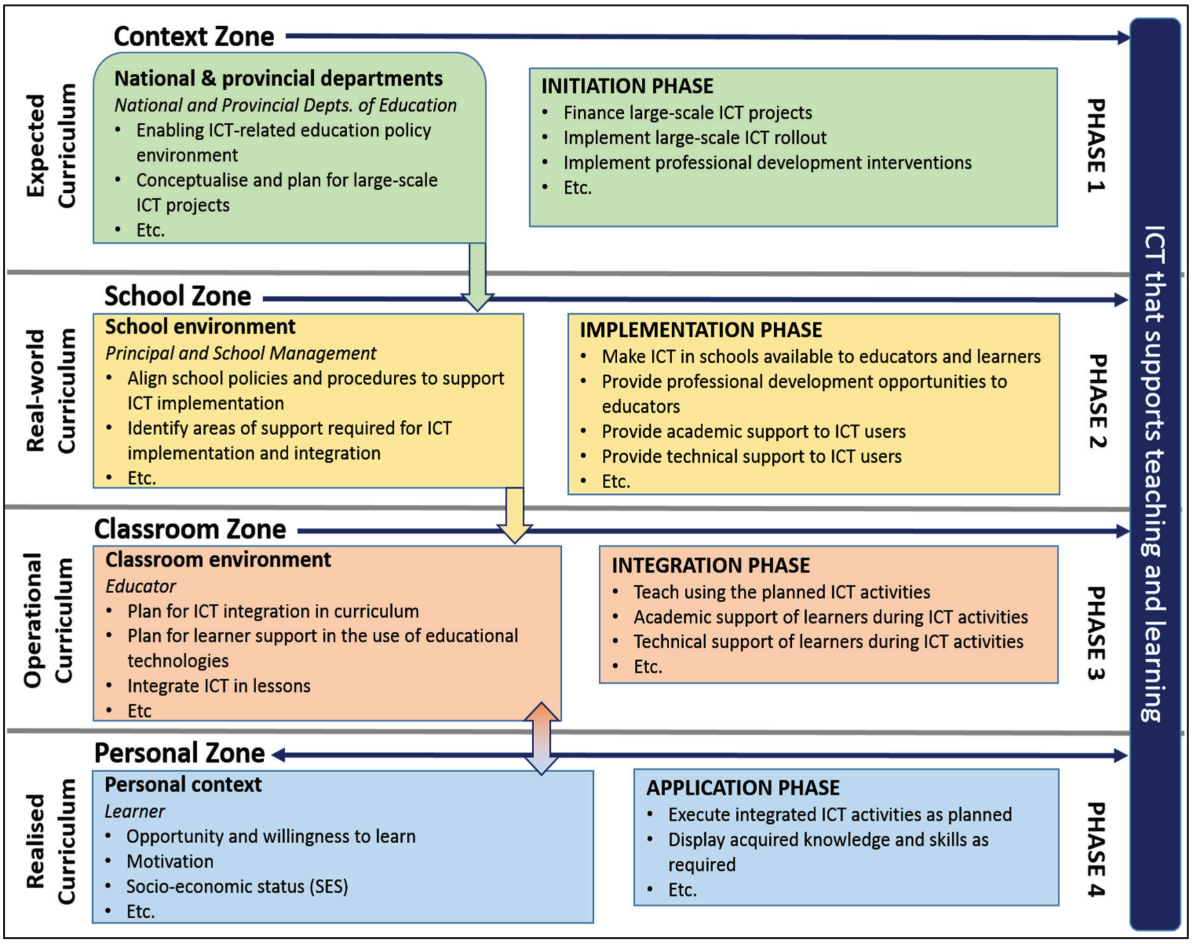


Figure 4. The Four Zones Model.

Structured around four zones of impact akin to those identified by Creemers and Kyriakides [39], the Four Zones Model comprises the Context Zone, School Zone, Classroom Zone, and Personal Zone. Each zone is associated with specific curriculum expectations mirroring those outlined in the TIMSS curriculum by the IEA [37], namely the Expected Curriculum, Real-world Curriculum, Operational Curriculum, and Realised Curriculum. Additionally, the model delineates phases essential for successful ICT integration in teaching and learning environments, as identified through an examination of the roles and responsibilities within each impact zone. ICT integration and implementation, outlined in the Four Zones Model, include the Initiation Phase, Implementation Phase, Integration Phase, and Application Phase.

The initial stage of the Four Zones Model is termed the Initiation Phase, situated within the Context Zone. This phase encompasses the functions and duties of national and provincial Departments of Education. These departments play a pivotal role in establishing a conducive policy and legal framework to facilitate the integration of ICT within schools. Given the limited financial resources of many schools to acquire ICT resources for educational purposes, this study underscores the imperative for Departments of Education to spearhead the distribution of ICT infrastructure to schools. Additionally, within the Context Zone lies the delineation of the Expected Curriculum, which embodies the official curriculum sanctioned by the Departments of Education.

The second stage in the Four Zones Model is termed the Implementation Phase, situated within the School Zone, delineating the responsibilities of school principals and management. Within this phase, school management holds the pivotal role of fostering an environment conducive to ICT integration through the formulation of school policies and protocols that facilitate its effective utilisation. Findings from this study suggest instances where, despite the presence of ICT infrastructure within schools, its integration into T&L processes remained limited. Many educators, despite undergoing professional development interventions, seemed unable to apply acquired knowledge in practice. Establishing an enabling environment for ICT integration may involve seemingly straightforward measures such as ensuring equitable access to computer facilities by scheduling dedicated time for each class. Moreover, school management bears the responsibility of arranging necessary support for ICT users, which could be as simple as assigning an enthusiastic staff member to oversee ICT assistance. Additionally, an often-overlooked necessity is the provision of additional time for educators to plan and implement ICT integration into their lessons. School management could address this need by recognising ICT planning and implementation as a designated extracurricular activity, allowing educators the requisite time. Within the School Zone, the Real-world Curriculum is outlined, reflecting the curriculum's implementation guided by school-specific policies, procedures, and supportive structures.

The third stage within the Four Zones Model, known as the Integration Phase, holds paramount significance as it marks the operationalisation of ICT integration. Nestled within the Classroom Zone, this phase delineates the duties and obligations of educators within their classrooms. The incorporation of ICT into the school curriculum represents a novel undertaking for most educational institutions, necessitating educators to revise their existing lesson plans to accommodate ICT integration seamlessly. Following the planning stage, educators assume the responsibility of delivering lessons utilising newly devised ICT activities. Moreover, educators are tasked with providing technical support to students encountering challenges while utilising ICT to fulfill assigned tasks. Findings from this study indicate that, although many educators exhibit enthusiasm towards ICT integration in T&L, and participate in professional development initiatives, such interventions often fall short in adequately preparing educators for the practical realities of integrating ICT into their pedagogical practices. Within the Classroom Zone, the Operational Curriculum is outlined, reflecting the curriculum as implemented by educators within the school setting.

The final stage in the Four Zones Model is the Application Phase, situated within the Personal Zone, elucidating the roles and obligations of students within and outside the classroom environment. With ICT integration, the dynamics of T&L shift from being centered on educators to becoming centered on students. Consequently, students are entrusted with the responsibility of taking charge of their own educational journey. Within the Personal Zone, the Realised Curriculum is delineated, representing the curriculum as grasped and achieved by students. It serves as the culmination of the Expected Curriculum, embodying the ultimate outcome of the educational process.

The Four Zones Model endeavours to address the prevalent issue of ambiguity surrounding the roles and responsibilities of various stakeholders involved in the integration of ICT within schools. While originally conceptualised within the context of South African education, the model possesses a level of generality that renders it adaptable to diverse educational settings beyond South Africa. Hence, it is recommended that the Four Zones Model be regarded as a guiding framework for forthcoming ICT integration projects within schools.

6.1. Reliability of the Four Zones Model

The researchers have engaged in discussions with domain experts in educational technology and policy-making to critique and refine the model. This expert feedback is being systematically incorporated to strengthen the model's consistency and application potential. A series of hypothetical applications of the model to past ICT integration projects are being undertaken. By examining how the model would have functioned in these well-documented instances, we aim to assess its reliability in various educational settings.

6.2. Validity of the Four Zones Model

We have more deeply grounded the model in the existing literature and theories of educational technology adoption, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT), to establish its face and content validity. The model has also been subjected to scrutiny by a panel of experts in ICT in education, who have provided insights and recommendations to ensure that it adequately represents the complex dynamics of ICT integration in schools.

7. Recommendations for Improved Implementation of ICT in Schools

Recommendations for improved policy and practice in the implementation of ICT in schools are organised according to the four zones of impact as identified in the proposed "Four Zones Model for the Integration of ICT in Schools". The zones identified are the Context Zone, the School Zone, the Classroom Zone and the Personal Zone.

7.1. The Context Zone (National and Provincial Departments of Education)

It is evident that schools require substantial support from both national and provincial education departments when implementing ICT hardware and software. The selection of educational software must be tailored to meet the specific educational needs of students in each school, rather than adopting a uniform solution for all schools within a province. Furthermore, contextualising educational software is essential to ensure that students can relate to the content, language, and assessment methods employed. For large-scale ICT projects in education, continuous technical and academic support should be provided to schools and educators to facilitate the seamless integration of educational software into regular classroom practices.

7.2. The School Zone (Principals and School Management)

According to the conclusions drawn from this study, it appears that school principals require distinct professional development initiatives for effectively integrating ICT into the school curriculum, differing from those tailored for educators. Many principals demonstrate a lack of clarity regarding their role in implementing ICT integration within the curriculum and struggle to recognise the potential positive impacts of ICT on T&L within their schools. Additionally, their leadership in promoting the integration and utilisation of ICT in T&L appears inadequate when they themselves lack a comprehensive understanding of ICT integration. Considering the substantial financial investment required for ICT implementation, it becomes imperative for school management to assume responsibility for the effective utilisation of ICT resources. It is incumbent upon school management to administer the school's ICT resources in a manner that ensures equitable access for all educators and classes. Implementing a weekly designated time slot within the formal school timetable could be a practical solution to ensure each class receives fair access to the school's ICT equipment.

7.3. The Classroom Zone (Educators)

Initiating professional development programs for educators concerning ICT integration within the school curriculum should commence with addressing fundamental computer literacy and skills. It is crucial for educators to feel at ease with computer usage for personal tasks, as lacking this confidence may impede their ability to effectively integrate ICT into their teaching practices. There is a pressing need to enhance opportunities for professional development among educators concerning ICT integration within the school curriculum. The observed number of educators who reported non-attendance at professional development activities underscores this urgency. Centralised management of these professional development initiatives by national or provincial education departments could ensure equitable access for all educators. Professional development interventions for educators must prioritise practical aspects of ICT integration within the school curriculum and incorporate workplace-based support.

7.4. The Personal Zone (Students)

Regular use of ICT is essential for students to cultivate non-academic skills essential for active engagement in the knowledge-driven society awaiting them beyond school. Nevertheless, many students lack access to computers at home, underscoring the responsibility of education departments and schools to provide access to computers and software during school hours. Educational software must be tailored to the educational level of students within each school, and students should have sufficient time allocated for engaging with high-quality educational software. The more time students spend with high-quality educational software, the greater the likelihood of meaningful learning and improved academic achievement. Thus, it is incumbent upon school management to ensure equitable access to ICT resources for all students within the school environment.

8. Limitations

A limitation of the current study is that a secondary data analysis was conducted. Conducting a secondary data analysis presents several limitations that researchers must consider. Firstly, the original data may not have been collected with the specific research questions in mind, leading to potential gaps in the dataset that may hinder comprehensive analysis. Additionally, the quality of the data may vary, as it relies on the accuracy and reliability of the original data collection methods and procedures. Researchers may encounter issues with missing or incomplete data, inconsistencies in data coding, or inaccuracies in measurements, all of which can compromise the reliability and validity of the findings. Furthermore, secondary data analysis may limit researchers' ability to control for confounding variables or explore alternative explanations for observed phenomena, as they have no control over the data collection process. These limitations were mitigated by thoroughly studying the TIMSS booklets and familiarising ourselves with all the steps and procedures followed by the TIMSS researchers.

9. Conclusions

While our study is grounded in the specific context of South African schools, the findings and implications carry broader significance for several reasons. The obstacles and successes identified in the South African context often mirror those in other emerging economies and even in under-resourced areas of developed countries. The strategies and models we propose can be informative for similar contexts where educational technology integration is a work in progress. Furthermore, the Four Zones Model, though developed within the South African framework, is designed with adaptability in mind. It is based on universal principles of ICT integration that are relevant to diverse educational settings. We anticipate that the model can be adjusted to suit different regional and cultural contexts. Finally, the trends and patterns in ICT use we have identified contribute to the global discourse on educational technology. Our research adds to the understanding of how ICT can influence educational outcomes, which is a subject of international concern.

Author Contributions: Conceptualization: M.A.G., G.M.K. and L.v.R.; methodology: G.M.K. and L.v.R., formal analysis: M.A.G., writing—original draft, review and editing: M.A.G., G.M.K. and L.v.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethics approval obtained from the University of Pretoria (SM 16/06/02).

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

Data Availability Statement: The TIMSS datasets are available on the TIMSS website (https://timssandpirls.bc.edu/timss2019/ (accessed on 11 June 2021)).

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

References

1. Reddy, V.; Zuze, T.L.; Visser, M.; Winnaar, L.; Juan, A.; Prinsloo, C.H.; Arends, F.; Rogers, S. *Beyond Benchmarks: What Twenty Years of TIMSS Data Tell Us about South African Education*; Human Sciences Research Council: Pretoria, South Africa, 2015.
2. Reddy, V.; Winnaar, L.; Juan, A.; Arends, F.; Harvey, J.; Hannan, S.; Namome, C.; Zulu, N. *TIMSS 2019 Highlights of South African Grade 5 Results in Mathematics and Science*; Department of Basic Education: Pretoria, South Africa, 2020.
3. Reddy, V.; Winnaar, L.; Juan, A.; Arends, F.; Harvey, J.; Hannan, S.; Namome, C.; Sekhejane, P.; Zulu, N. *TIMSS 2019: Highlights of South African Grade 9 Results in Mathematics and Science*; Department of Basic Education: Pretoria, South Africa, 2020.
4. Dahshan, M.; Galanti, T. Teachers in the loop: Integrating computational thinking and mathematics to build early place value understanding. *Educ. Sci.* **2024**, *14*, 201. [CrossRef]
5. Schiavo, F.; Campitiello, L.; Todino, M.D.; Di Tore, P.A. Educational robots, emotion recognition and ASD: New horizon in special education. *Educ. Sci.* **2024**, *14*, 258. [CrossRef]
6. World Bank. *The Adequacy of Public Expenditure on Education and the Needs Post-COVID-19*; World Bank Group: Bretton Woods, NH, USA, 2023.
7. Rayan, B.; Watted, A. Enhancing education in elementary schools through gamified learning: Exploring the impact of Kahoot! on the learning process. *Educ. Sci.* **2024**, *14*, 277. [CrossRef]
8. Rasdiana Wiyono, B.B.; Imron Al Rahma, L.; Arifah, N.; Azhari, R.; Elfira; Sibula, I.; Maharmawan, M.A. Elevating teachers' professional digital competence: Synergies of principals' instructional e-supervision, technology leadership and digital culture for educational excellence in digital-savvy era. *Educ. Sci.* **2024**, *14*, 266. [CrossRef]
9. Elecalde, R.G.; Garcia, J.C.; Martos AL, B.; Arnáez, B.S. Digital and social-civic skills in future primary education teachers: A study from the didactics of social sciences for the improvement of teacher training in competences. *Educ. Sci.* **2024**, *14*, 211. [CrossRef]
10. Engelbrecht, J.; Borba, M.C. Recent developments in using digital technology in mathematics education. *ZDM–Math. Educ.* **2023**, *56*, 281–292. [CrossRef]
11. Attard, C.; Holmes, K. *Technology-Enabled Mathematics Education: Optimising Student Engagement*; Routledge: London, UK, 2019.
12. Hansson, S.O. Technology and mathematics. *Philos. Technol.* **2020**, *33*, 117–139. [CrossRef]
13. South African National Research Network. The South African NREN. SANReN. 2024. Available online: https://www.sanren.ac.za/south-african-nren/ (accessed on 20 February 2024).
14. Department of Communications. *Broadcasting Digital Migration Policy for South Africa*; Republic of South Africa: Pretoria, South Africa, 2008.
15. Infrastructure South Africa. SA Connect. Republic of South Africa. 2022. Available online: https://infrastructuresa.org/sip-projects/sa-connect-phase-2/ (accessed on 10 December 2023).
16. Barakabitze, A.A.; Lazaro, A.W.-A.; Ainea, N.; Mkwizu, M.H.; Maziku, H.; Matofali, A.X.; Iddi, A.; Sanga, C. Transforming African education systems in science, technology, engineering, and mathematics (STEM) using ICTs: Challenges and opportunities. *Educ. Res. Int.* **2019**, *2019*, 6946809. [CrossRef]
17. Mokotjo, L.; Mokhele, M.L. Challenges of integrating GeoGebra in the teaching of mathematics in South African high schools. *Univers. J. Educ. Res.* **2021**, *9*, 963–973. [CrossRef]
18. Mwapwele, S.D.; Marais, M.; Dlamini, S.; Van Biljon, J. Teachers' ICT adoption in South African rural schools: A study of technology readiness and implications for the South Africa connect broadband policy. *Afr. J. Inf. Commun.* **2019**, *24*, 1–21.
19. Mahwai, N.J.; Wotela, K. Integrating technology in teaching and learning: Have Seshego Circuit rural schools escaped the challenges? *J. Public Adm. Dev. Altern.* **2022**, *7*, 55–67. [CrossRef]
20. Dlamini, R. Factors constraining teacher integration of ICT in Gauteng schools. *Indep. J. Teach. Learn.* **2022**, *17*, 28–43.
21. Ramafi, P. Investigating the barriers of ICT use in teaching and learning at public schools in South Africa. In Proceedings of the 2022 International Conference on Intelligent and Innovative Computing Applications, Balaclava, Mauritius, 8–9 December 2022; pp. 92–102. [CrossRef]
22. Graham, M.A.; Stols, G.; Kapp, R. Teacher practice and integration of ICT: Why are or aren't South African teachers using ICTs in their classrooms. *Int. J. Instr.* **2020**, *13*, 749–766. [CrossRef]
23. Graham, M.A.; Stols, G.H.; Kapp, R. Integrating classroom technology: South African mathematics teachers. *Comput. Sch.* **2021**, *38*, 189–213. [CrossRef]
24. Ojo, O.A.; Adu, E.O. The effectiveness of Information and Communication Technologies (ICTs) in teaching and learning in high schools in Eastern Cape Province. *S. Afr. J. Educ.* **2018**, *38*, 1483. [CrossRef]
25. Chisango, G.; Marongwe, N.; Mtsi, N.; Matyedi, T.E. Teachers' perceptions of adopting information and communication technologies in teaching and learning at rural secondary schools in eastern cape. *S. Africa. Afr. Educ. Rev.* **2020**, *17*, 1–19. [CrossRef]
26. Filita, N.; Jita, T. Teachers' perspectives on the use of ICT in the teaching of a South African home language, Sesotho. *J. Lang. Teach.* **2021**, *55*, 219–241. [CrossRef]
27. Zenda, R.; Dlamini, R. Examining factors that influence teachers to adopt information and communication technology in rural secondary schools: An empirical study. *Educ. Inf. Technol.* **2023**, *28*, 815–832. [CrossRef]
28. Mnisi, B.R.; Mtshali, T.I.; Moses, M. Moving beyond the challenges of learning through technologies: The current status of ICT integration in South African schools. *J. Educ. e-Learn. Res.* **2024**, *11*, 125–134. [CrossRef]

29. Saal, P.E.; Van Ryneveld, L.; Graham, M.A. The relationship between using information and communication technology in education at the mathematics achievement of students. *Int. J. Instr.* **2019**, *12*, 405–424. [CrossRef]
30. Saal, P.E.; Van Ryneveld, L.; Graham, M.A. Comparing the relationship between using educational technology in mathematics and student achievement in South Africa and Germany. *J. Res. Technol. Educ.* **2022**, *54*, 581–598. [CrossRef]
31. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
32. Saal, P.E.; Graham, M.A.; Van Ryneveld, L. Integrating educational technology in mathematics education in economically disadvantaged areas in South Africa. *Comput. Sch.* **2020**, *37*, 253–268. [CrossRef]
33. Kolobe, L.; Mihai, M. The integration of technology in supporting progressed learners in English First Additional Language comprehension. *Perspect. Educ.* **2021**, *39*, 303–323. [CrossRef]
34. Mahlo, L.; Waghid, Z. Exploring information and communication technology integration among teachers in township public primary schools. *S. Afr. J. Educ.* **2023**, *43*, 2160. [CrossRef]
35. Ntsobi, M.; Nyamkure, B. Responsive ICT integration framework to enhance teaching and learning in Gauteng schools. In Proceedings of the 2024 8th World Conference on Qualitative Research, Azores (Portugal), Johannesburg (South Africa), and online, 23–25 January 2024.
36. Creemers, B.P.; Kyriakides, L. School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scand. J. Educ. Res.* **2010**, *54*, 263–294. [CrossRef]
37. Mullis, I.V.S. Introduction. In *TIMSS 2019 Assessment Frameworks*; Mullis, I.V.S., Martin, M.O., Eds.; Boston College, TIMSS & PIRLS International Study Center: Chestnut Hill, MA, USA, 2017; pp. 3–10. Available online: https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/context-questionnaire-framework/ (accessed on 27 April 2024).
38. Antoniou, P.; Kyriakides, L.; Creemers, B.P.M. Investigating the effectiveness of a dynamic integrated approach to teacher professional development. *CEPS J. Cent. Educ. Policy Stud. J.* **2011**, *1*, 13–41. [CrossRef]
39. Creemers, B.P.; Kyriakides, L. *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice, and Theory in Contemporary Schools*; Routledge: London, UK, 2007.
40. Fishbein, B.; Foy, P.; Yin, L. *TIMSS 2019 User Guide for the International Database*, 2nd ed.; TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement: Chestnut Hill, MA, USA, 2021.
41. Hossain, M.Z. Implementation of Grade 8 science curriculum 2012 in Bangladesh: Challenges and way forward. *J. Educ. Res.* **2019**, *9*, 93–117. [CrossRef]
42. Hooper, M.; Mullis, I.V.S.; Martin, M.O.; Fishbein, B. TIMSS 2019 context questionnaire framework. In *TIMSS 2019 Assessment Frameworks*; Mullis, I.V.S., Martin, M.O., Eds.; Boston College, TIMSS & PIRLS International Study Center: Chestnut Hill, MA, USA, 2017; pp. 59–78. Available online: https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/context-questionnaire-framework/ (accessed on 10 August 2020).
43. Department of Basic Education. *Curriculum Assessment Policy Statements (CAPS)*; Government Printers: Pretoria, South Africa, 2011.
44. Park, Y.S.; Konge, L.; Artino, A.R., Jr. The positivism paradigm of research. *Acad. Med.* **2020**, *95*, 690–694. [CrossRef]
45. Mouton, J. *How to Succeed in Your Master's and Doctoral Studies: A South African Guide and Resource Book*; Van Schaik: Pretoria, South Africa, 2001.
46. Wang, X.; Cheng, Z. Cross-sectional studies: Strengths, weaknesses, and recommendations. *Chest* **2020**, *158*, S65–S71. [CrossRef]
47. Andrade, C. The inconvenient truth about convenience and purposive samples. *Indian J. Psychol. Med.* **2021**, *43*, 86–88. [CrossRef]
48. LaRoche, S.; Joncas, M.; Foy, P. Sample design in TIMSS 2019. In *Methods and Procedures: TIMSS 2019 Technical Report*; Martin, M.O., von Davier, M., Mullis, I.V.S., Eds.; TIMSS & PIRLS International Study Center: Amsterdam, The Netherlands, 2020; pp. 51–83.
49. Cotter, K.E.; Centurino, V.A.S.; Mullis, I.V.S. Developing the TIMSS 2019 mathematics and science achievement instruments. In *Methods and Procedures: TIMSS 2019 Technical Report*; Martin, M.O., von Davier, M., Mullis, I.V.S., Eds.; TIMSS & PIRLS International Study Center: Amsterdam, The Netherlands, 2020; pp. 6–41.
50. Van Ginkel, J.R.; Linting, M.; Rippe, R.C.A.; Van der Voort, A. Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *J. Personal. Assess.* **2020**, *102*, 297–308. [CrossRef] [PubMed]
51. Trends in International Mathematics and Science Study. Teacher Questionnaire: Grade 5. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).
52. Trends in International Mathematics and Science Study. Teacher Questionnaire: Grade 9. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).
53. Trends in International Mathematics and Science Study. School Questionnaire: Grade 5. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).

54. Trends in International Mathematics and Science Study. School Questionnaire: Grade 9. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).
55. Trends in International Mathematics and Science Study. Student Questionnaire: Grade 5. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).
56. Trends in International Mathematics and Science Study. Student Questionnaire: Grade 9. Boston College, TIMSS & PIRLS International Study Center. 2018. Available online: https://timssandpirls.bc.edu/timss2019/questionnaires/index.html (accessed on 10 August 2020).
57. Siew, P.H. Pedagogical change in mathematics learning: Harnessing the power of digital game-based learning. *Educ. Technol. Soc.* **2018**, *21*, 259–276.
58. Janković, A.; Maričić, M.; Cvjetićanin, S. Comparing science success of primary school students in the gamified learning environment via Kahoot and Quizizz. *J. Comput. Educ.* **2024**, *11*, 471–494. [CrossRef]
59. Pillay, N.S.; Okonkwo, C.W.; Anele, A. The impact of loadshedding on student academic performance: A data analytics approach. In Proceedings of the 2023 International Conference of Teaching, Assessment and Learning in the Digital Age, Cape Town, South Africa, 6–7 December 2023; pp. 329–341.

education sciences

MDPI

Article

A Usability Analysis and Consequences of Testing Exploration of the Problem-Solving Measures–Computer-Adaptive Test

Sophie Grace King [1], Jonathan David Bostic [1,*], Toni A. May [2] and Gregory E. Stone [3]

[1] School of Inclusive Teacher Education, Bowling Green State University, Bowling Green, OH 43403, USA; sgking@bgsu.edu
[2] College of Community and Public Affairs, Binghamton University, Binghamton, NY 13902, USA; tmay3@binghamton.edu
[3] Clarity Assessment Systems Ltd., Oswego, NY 13126, USA; gregory@clarityassessmentsystems.com
* Correspondence: bosticj@bgsu.edu

Abstract: Testing is a part of education around the world; however, there are concerns that consequences of testing is underexplored within current educational scholarship. Moreover, usability studies are rare within education. One aim of the present study was to explore the usability of a mathematics problem-solving test called the Problem Solving Measures–Computer-Adaptive Test (PSM-CAT) designed for grades six to eight students (ages 11–14). The second aim of this mixed-methods research was to unpack consequences of testing validity evidence related to the results and test interpretations, leveraging the voices of participants. A purposeful, representative sample of over 1000 students from rural, suburban, and urban districts across the USA were administered PSM-CAT followed by a survey. Approximately 100 of those students were interviewed following test administration. Findings indicated that (1) participants engaged in the PSM-CAT as desired and found it highly usable (e.g., most respondents were able to use and find the calculator and several students commented that they engaged with the test as desired) and (2) the benefits from testing largely outweighed any negative outcomes (e.g., 92% of students interviewed had positive attitudes towards the testing experiences), which in turn supports consequences from testing validity evidence for PSM-CAT. This study provides an example of a usability study for educational testing and builds upon previous calls for greater consequences of testing research.

Keywords: assessment; computer-adaptive test; problem solving; test; usability; validity

1. Introduction

Educators need access to comprehensive, valid information about their students' mathematics learning. In turn, educators should make data-based decisions as a result of using valid results from high-quality assessments (Lawson & Bostic, 2024; Fennell et al., 2023). Assessment in this study refers to the activities that teachers and others use to gather student data, as well as the activities that provide teachers with feedback for modifying their teaching and student learning (Fennell et al., 2023). In some capacities, in-the-moment questioning can draw out students' knowledge during classroom instruction, which informs teachers about what and how students are learning (Fennell et al., 2023). More formal assessments such as quizzes and tests have potential to provide necessary information about students' learning, too (Fennell et al., 2023). A distinction is sometimes made between the terms test and assessment because a test can encompass broader sources of information than a single instrument (AERA et al., 2014); however, in this manuscript, both terms will be used interchangeably to promote readability for a broad audience. Readers interested in the language differences and nuances are encouraged to consult the *Standards for Educational and Psychological Testing* ([*Standards*], AERA et al., 2014).

Schools around the world use tests to gather information about students' mathematics performance. However, a test's usability—the degree to which a respondent engages with it in an intended manner—can implicate a respondent's results (Estrada-Molina et al., 2022). Further, negative consequences from engaging with a test may also unnecessarily impact the respondent and produce greater test score variance (AERA et al., 2014; Lane, 2020).

The present study explores the consequences of testing related to a mathematics problem-solving test as well as its usability with the intended population: grades six to eight (ages 11–14) students. A purposeful, representative sample (Creswell, 2014) of students from the population participated in surveys and interviews immediately following test administration. One goal of this study was to disseminate findings related to the consequences of testing and usability related to a mathematics problem-solving test that is grounded in classroom standards. A second goal was to provide a study that could serve as a model for educational researchers with an example of how to conduct a usability study.

2. Literature Review

2.1. Educational Policies: An Overview

The No Child Left Behind Act of 2001 (NCLB) was designed to improve the performance of students enrolled in public schools in the United States of America (USA). One key feature of NCLB was accountability for student achievement, which initially required each state to develop standardized tests in reading, mathematics, and later in science, to be administered annually in grades three–eight, and once in high school. This act was reflective of USA educational and standardized test data use policies that went into effect. While standardized testing is commonplace in the USA, other countries also implement standardized tests: Japan's National Assessment of Academic Ability (Hino & Ginshima, 2019), South Africa's Annual National Assessment (Maphalala & Khumalo, 2018), Germany's Abitur examination (Bruder, 2021), Mexico's Plan Nacional para la Evaluación de los Aprendizajes (Céspedes-González et al., 2023), and Israel's Bagrut Matriculation Exams (Naveh, 2004). Outside of standardized tests, another typical classroom assessment is an end-of-unit exam typically administered by teachers at the end of a unit or course to measure the degree to which students have mastered the material taught through instruction. These are two types of summative assessments. Summative assessments are tools to gather data about how much has been learned and/or whether an individual has reached a desired level of proficiency (A. H. Schoenfeld, 2015). Formative assessment includes "all those activities undertaken by teachers and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black & Wiliam, 2010, p. 7). One type of formative classroom assessment is a progress monitoring test. These progress monitoring tests provide teachers with information that communicates what students know and may provide teachers with information they can use to plan further instruction. Taken collectively, these types of assessments can be framed around notions of summative assessment and formative assessment.

All assessments have consequences and effects on teachers and students (McGatha & Bush, 2013). Positive consequences from tests for formative and summative assessments include but are not limited to improvements in teacher and student motivation and effort; better content and format of assessments; and advancements in the use and nature of test preparation activities (Lane & Stone, 2002). There can also be negative consequences: narrowing of curricula and instruction; use of unethical test preparation materials; unfair test score use; reassignment of teachers or students based on a single data point; and decreases in student and teacher confidence, motivation, and/or self-esteem (Lane & Stone, 2002). Evidence related to consequences of testing should be explored with any test—including both formative and summative assessments (AERA et al., 2014; Bostic, 2023; Kane, 2006; Lane, 2014).

2.2. Computer-Adaptive Tests (CATs): The Transition from PSM to PSM-CAT

The PSM test measures middle school (i.e., grades six through eight; ages 11–14) students' mathematical problem-solving performance in ways that leverage their understanding of grade-level mathematics content, as derived from classroom standards. Classroom content standards may differ to some degree across states within the USA; however, 36 states maintain features originally found in the Common Core State Standards for Mathematics (CCSSM) that were implemented in 2011 (CCSSI, 2010). This PSM test functions primarily as a formative assessment. It may be used to gather information about students' problem-solving performance while drawing on their mathematics knowledge related to classroom content standards, and in turn inform teachers' future mathematics instruction.

Computer-adapted tests (CATs) are designed to improve measurement efficiency using a smaller number of ability-targeted items than traditional paper-pencil tests or other static tests (Martin & Lazendic, 2018). Static tests require all students to engage with a defined set of identical or nearly identical items (Wainer & Lewis, 1990). In order to effectively measure all students with a fixed set of items, test makers generally include a range of items, running from less to more difficult. Because the items are consistent across all students, regardless of ability, the desired level of measurement error (precision) is only achievable after a large number of items are administered. Items equally as difficult as a student is able have the capacity to reduce measurement error quickly and are considered "well-targeted". Items that are easier than a student is able, or conversely, items that are more difficult than a student is able, reduce measurement error more slowly and are considered "less well-targeted". A substantial number of items on static tests are required—not necessarily to ensure content coverage, but rather, psychometrically, to ensure that no matter what ability a student may possess, the test administration can generate an accurate measurement of student content mastery (Martin & Lazendic, 2018).

CATs accomplish the task of reaching measurement precision with far fewer items than traditional static paper-pencil tests because items are typically well-targeted to a student's specific level of ability (Martin & Lazendic, 2018). In a CAT environment, students are first administered an item of moderate difficulty. Subsequent items are then selected based on a student's response pattern (Davey, 2011). If a student answers correctly, then the next item delivered is more difficult (see Figure 1). On the other hand, when a student answers incorrectly, then the next item delivered is less difficult. Therefore, CATs have the capacity to efficiently zero in on (target) a student's particular capacity (ability) in a content area by delivering items that offer increased information about each student (Lane, 2020) if effectively developed.

CATs have been used in mathematics contexts since the 1990s (Davey, 2011). They provide a mechanism to better target students' mathematics abilities with shorter testing durations compared to static tests. There are numerous examples of CATs used in mathematics contexts with many coming between 1990 and 2010 during a period of rapid development. We share a recent example of mathematics-focused CAT use to demonstrate its continued appropriateness during a period of time when artificial intelligence, machine learning, and natural language processing are becoming more popular. Uko et al. (2024) created a mathematics and science CAT for Nigerian secondary students. One conclusion was that their CAT for mathematics and science contexts provided an accurate and efficient means to assess secondary students' abilities. Thus, they recommend that "CAT should be introduced in assessment of learning. . .of the students [and] to be tested with sufficient accuracy" (p. 85).

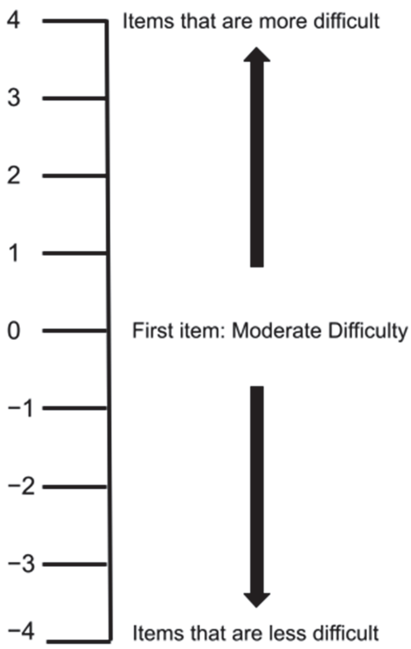


Figure 1. Example of Item Difficulty Measures for CATs. Numbers in this example represent item difficulties as measured using Rasch analysis, grounded in logits.

In the case of the PSM-CAT, delivery was made via a web-based browser directly in the classroom. Students were able to use an online calculator and formula sheet, which were located on the testing platform. In addition, they could use scratch paper and a writing utensil. The test had a time limit (30 min), and students could work at their own pace with no maximum number of items to complete. Thus, there was some freedom in the testing experience that might differ from a fixed-length test. Classroom tests like the PSM-CAT test are no different from other educational tests in the sense that it is essential that test scores are interpreted and used in an appropriate way (AERA et al., 2014), which implicates test developers, users, and administrators. Much like other CATs, respondents are provided more difficult items as they answer items correctly. Conversely, they see less difficult items after answering a present item incorrectly. It is plausible, though rare in practice, for a student engaged with PSM-CAT to see an item from a different grade level based upon their response pattern.

2.3. Validity

Validity ensures that an assessment accurately captures the intended construct or phenomenon being studied (Bostic, 2023; Kane, 2013). Validity is defined as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores" (Messick, 1989, p. 13). Validation "can be viewed as a process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (AERA et al., 2014, p. 11). It is a process by which scholars gather evidence related to an assessment to better assist others in evaluating the degree to which it measures what it purports (Bostic, 2023; Carney et al., 2022; Kane, 2013).

Five sources of validity evidence are described in the *Standards* (AERA et al., 2014). The five sources of validity evidence include test content, response processes, internal structure, relations to other variables, and consequences of testing (AERA et al., 2014). Figure 2 provides a description of each source of validity evidence. Using the metaphor of a rope with intertwined braids helps to demonstrate the idea that validity represents a unitary concept and the sources work together, and the strands represent the different sources of evidence working together to ground a validity argument within the construct. Past research (e.g., Krupa et al., 2024) has demonstrated that the consequences of testing is underexplored and rarely reported, especially in mathematics education assessment research. Stated more concretely, Krupa et al.'s (2024) research synthesis found that of the papers describing mathematics assessments between 2000 and 2020, consequences of testing validity evidence was described in 61 articles out of the total reviewed sample of 1206 articles, which equates to consequences of testing being described less than 2% of the time.

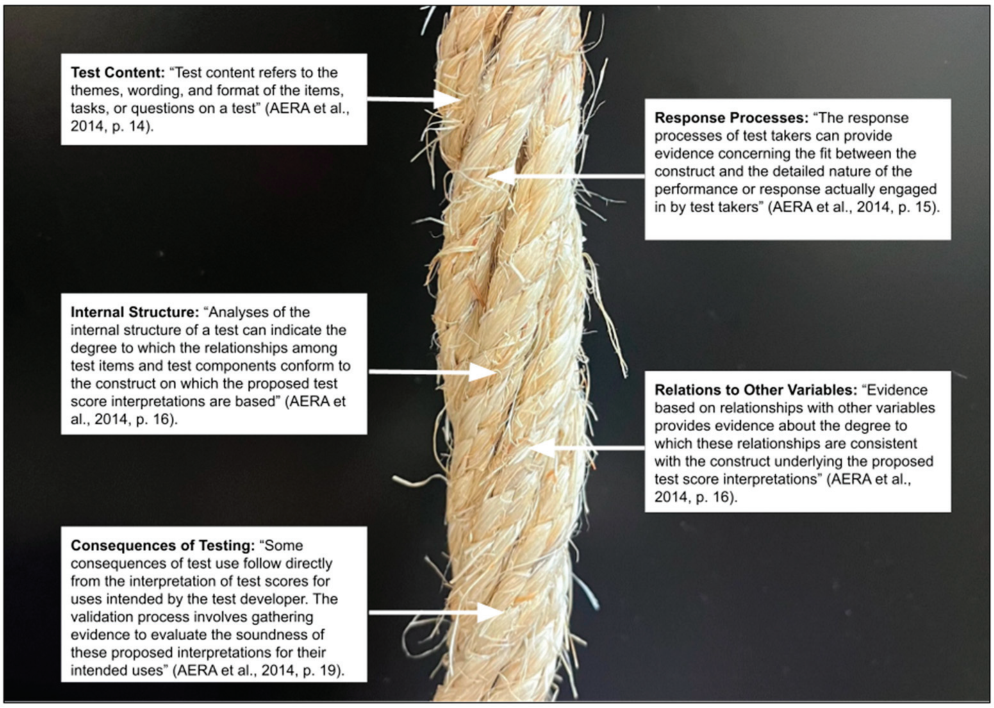


Figure 2. Metaphorical Knotted Rope Demonstrating Sources of Validity Evidence within a Concept. For more information, please reference (AERA et al., 2014).

2.4. Consequences of Testing

Consequences of testing includes both positive and negative outcomes and should be purposefully evaluated for any test (Lane, 2020; Sireci & Benitez, 2023). Some positive consequences include but are not restricted to student, teacher, and administrator motivation and efforts; the use of assessment results on teacher instruction and student learning; improvement or change to school courses (Lane, 2020). There are also potentially negative effects of an assessment that may arise from giving an assessment including narrowing of curricula, unfair question use, as well as decreased confidence and/or affect (Lane, 2020). Taken collectively, consequences of testing validity evidence help others to understand potential outcomes that may arise from administering a test and using its results.

Consequences of a test can include substantive outcomes such as student placement into a class or advancement to a grade level (AERA et al., 2014; Lee, 2020). Lee (2020) focused on the test consequences of an English reading course placement test (e.g., advanced, remedial, general) that was administered at a large university. Lee talked to students about the effects this assessment had on their learning and attitudes. One-on-one interviews were conducted to draw out perceptions about the test and respondents' experiences. The results showed that there was a near split amount of positive and negative perceptions towards the test. There were good attributes related to the test, but nearly half of the respondents felt there were negative attributes to the test's results and ensuing interpretations. Similarly, three-fourths of the respondents communicated that the test questions and reading passages were too complex, conveying that a consequence of using the result may be linked with students' negative perceptions of their testing experience. This is a high number of respondents concerned with negative consequences. An outcome from Lee's work was a further study of that test to guarantee that the results and their interpretations are justified in robust consequences of testing validity evidence. Lee's study also provides a guide for data collection with the use of interviews. A second outcome is that Lee's work demonstrates the importance of investigating the degree to which a test has negative and positive outcomes. More positive outcomes than negative outcomes, or at least a balance, is desirable (AERA et al., 2014). Greater negative outcomes than positive outcomes warrant concern and should be considered before test administration (AERA et al., 2014).

Consequences of a test can also be explored with content-focused tests administered by teachers. Heissel et al. (2021) designed a study with third- to eighth-grade students where they measured cortisol levels at various times during the day, including time during a test (Heissel et al., 2021). Cortisol is a hormone produced by the human body released in response to stress, like during a test. Samples of cortisol were collected to gauge the physiological stress response during test periods, and how it correlated with their test performance (Heissel et al., 2021). Researchers found that during the test, students had higher cortisol levels than during the rest of the school day. Students' negative test consequences experiences were linked with their test scores. That is, higher cortisol levels, like those during the test, were associated with lower test performance. If a goal is to effectively and fairly assess students' knowledge, then it is critical that testing situations limit anxiety or stress that might negatively contribute as variance (error) to a test score. In addition to consequences of testing, the need for evaluating assessment usability features is critically important.

2.5. Assessment Usability Features

Any K-12 assessment designed for classroom use should be usable by students and teachers. While that might sound simple, a usability study can explore the degree to which respondents understand the questions on the assessment, the ways respondents engage with the test, and whether test administrators (e.g., teachers and staff) perceive that the test can produce robust information usable for data-based decisions (AERA et al., 2014; Estrada-Molina et al., 2022). Interpretation and use statements for quantitative instruments are helpful when considering an assessment (Carney et al., 2022; Kane, 2013). Carney et al. (2022) explain that usability features are centrally important for test administrators and developers because they communicate necessary information. The degree to which users perceive features of the test as easy to locate and understand ultimately impacts test usability. For example, if respondents are supposed to use a calculator embedded in a CAT but they cannot easily locate it on the website, then there may be usability issues that negatively impact student performance and/or consequences of testing. Understanding the usability features of a test before wide administration also provides structure and support to decide whether the test has the potential to represent student knowledge accurately. Additionally, usability testing helps to identify potential areas of needed assessment or delivery modification. Accordingly, exploring assessment usability during pilot testing is essential to minimize the impact of conditions that may contribute to the validity of test results (AERA et al., 2014).

Research related to usability is often conducted in the healthcare field (e.g., Denecke et al., 2021; Hudson et al., 2012; Saad et al., 2022; Thielemans et al., 2018) but far less frequently within educational research. After conducting a thorough literature review on usability, our results showed that there were none in mathematics education and a limited number in education related to the usability of educational tests. This outcome led our team to draw from the healthcare literature where there was an abundant amount of usability studies. One healthcare usability study created a mobile mental health chatbot for regulating emotions. This chatbot was operated by multiple users, and researchers conducted a usability test (i.e., User Experience Questionnaire) to study users' experiences with it (Denecke et al., 2021). A 26-question survey with closed-end and open-ended items was administered to gather participants' experiences with the chatbot. Denecke and colleagues found that participants confirmed the chatbot as understandable, easy to learn, and clear. However, attractiveness, novelty, and dependability were scored as below average. Usability results, like those from a survey, allow for judgment, comparisons, and evidence-informed modifications to be made to the tool under study.

Another healthcare assessment usability study employed direct observation, focus groups, and questionnaires to understand a test's usability (Thielemans et al., 2018). Thielemans and colleagues' work outlines how to study the usability of a healthcare assessment through focus group discussions (FGDs) and a self-administered questionnaire. Participants used a device and were asked to critique features. They engaged in a usability assessment with a mixed-methods approach including observations, surveys, and focus groups. The present study employed a convergent mixed-methods approach (Creswell & Plano Clark, 2018) similar to Thielemans et al. (2018), and builds upon usability studies harnessing interviews and surveys (Denecke et al., 2021).

3. Materials and Methods

3.1. The Present Survey

Usability studies can be conducted to explore how students understand the test's directions and questions, and how students perceive the features of the test to be easy to use and foster positive outcomes after testing (AERA et al., 2014; Estrada-Molina et al., 2022). If students cannot easily access the resources, or the questions they saw were not applicable to them, then the assessment's results will not accurately characterize students' performance. Similarly, it is necessary to evaluate consequences of testing validity evidence to confirm that the benefits outweigh the negative outcomes.

The original version of PSM-CAT is paper-pencil and delivered in a static format (see Bostic & Sondergeld, 2015, 2018; Bostic et al., 2017); however, the version investigated for this study is delivered in a computer-adaptive format (i.e., PSM-CAT; Bostic et al., 2024). The paper-pencil version of the PSM test has been ongoing since 2019; meanwhile, the PSM-CAT development started in 2021 and the first administration was in 2024.

The PSM-CAT is grounded in mathematical problems and problem-solving frameworks. The items align with A. Schoenfeld's (2011) framework characterizing mathematical problems as tasks such that the number of solutions is uncertain, the pathway to a solution (i.e., strategy) is unclear, and there are multiple pathways to a solution. These are word problems; therefore, we also address Verschaffel et al.'s (1999) framework for mathematical word problems. That is, PSM-CAT items are also grounded as being open, complex, and realistic. Open tasks can be solved in multiple developmentally appropriate strategies. Complex tasks engage problem solvers in ways that cause them to think, pause, and reflect. Realistic tasks draw upon experienced or believable situational knowledge as a key part of the task. Finally, the PSM-CAT is grounded in Lesh and Zawojewski's (2007) framing of mathematical problem solving. Taken collectively, these frameworks help to ground the PSM and PSM-CAT.

A concern with any test is its content validity evidence. The PSM-CAT uses the CCSSM's Standards for Mathematics Content as a content blueprint. Content domains within the grades six–eight regarding Standards for Mathematics Content vary; albeit, they include domains such as Geometry, Number Sense, Expressions and Equations, and Statistics and Probability. Some PSM-CAT items have a primary and secondary standard. Readers interested in test content validity evidence should consult Bostic et al. (2024). Three sample items have been released from the test, which are shared in Figure 3 below to assist readers. The grade six item aligns to Ratio and Proportions and Expressions and Equations content standards. The grade seven item aligns to an Expressions and Equations content standard, whereas the grade eight item aligns to Number Sense content standards. In summary, all items were deemed by multiple expert panels to align with the desired Standards for Mathematics Content, and additionally, each item connected to one or more Standards for Mathematical Practice (see (CCSSI, 2010) for more information).

Grade 6:
A group of 150 tourists were waiting for a shuttle to take them from a parking lot to a theme park's entrance. The only way they could reach the park's entrance was by taking this shuttle. The shuttle can carry 18 tourists at a time. After one hour, everyone in the group of 150 tourists reached the theme park's entrance. What is the fewest number of times that the shuttle picked tourists up from the parking lot?

Grade 7:
A water tower contains 16,880 gallons of water. Each day half of the water in the tank is used and not replaced. This process continues for multiple days. How many gallons of water are in the tower at the end of the fourth day?

Grade 8:
A chess board is made of eight rows with eight squares in each row. Each square has an area of 3 $inches^2$. What is the exact length for one edge of the chess board?

Figure 3. Sample of PSM-CAT items.

This study used a convergent mixed-methods research approach to collect data about PSM-CAT, specifically (a) its usability among potential users and (b) validity evidence related to consequences of testing. One intended study outcome is to broadly distribute findings for potential test users and administrators (e.g., school personnel, researchers, and evaluators) to consider when selecting a mathematics problem-solving test. A second outcome is to provide readers with a model of a study that combines exploring usability and consequences from testing for a K-12 test. The research questions for this study are

RQ1: *To what degree is the PSM-CAT usable for students?*

RQ2: *What consequences of testing evidence exist for the PSM-CAT?*

3.2. Research Design

This convergent mixed-methods research (MMR) design (Creswell & Plano Clark, 2018) utilized a survey and 1–1 interviews (QUAN + QUAL). There are aspects from the survey and interviews that address each of the two research questions. In an MMR study, both quantitative and qualitative data are collected and then analyzed (Creswell & Plano Clark, 2018). We compared the quantitative and qualitative results to build a common understanding (Creamer, 2017; Creswell, 2014). Data were gathered immediately following test completion to accurately measure students' experiences with the test. Students were more likely to recall their experiences with just minutes between test completion, survey, and interview.

The present study is part of a larger grant-funded project (*National Science Foundation*—2100988; 2101026) that works with representatively and purposefully sampled school districts across the USA. These participating districts were purposefully sampled to represent different regions and contexts of the USA, including one urban district in the Pacific region, one large suburban district in the Mountain West region, and multiple school districts inclusive of suburban and rural contexts in the Midwest. This study met exempt IRB status; students who wished not to participate in data collection did not participate. All names used in this study are pseudonyms that participants selected. We describe methods for the survey followed by methods for the interviews.

3.3. Survey Methods

3.3.1. Participants

A survey was administered to 1010 students in grades six through eight to capture their perceptions of the usability and consequences of taking the PSM-CAT. Participant demographic information is shown in Table 1. We used purposeful, representative sampling (Creswell, 2014) because our team sought to generate findings that may reflect the diversity of middle school students across three regions of the USA. Participant selection was performed with a goal to have a broad pool of students with respect to sex, racial/ethnic diversity, and geographic location (both region and context). Students self-identified their sex and race/ethnicity. As in prior work from this project, participants chose (a) male, (b) female, or (c) prefer not to say for gender. This approach was used after students in earlier studies with this project recommended a third option (see Bostic et al., 2024). Students were offered multiple options for race/ethnicity that followed the USA census' approach.

Table 1. Demographics of Participants for Survey and Interviews.

Demographic Type	Demographic Information	Participants in the Survey	Participants in the Interviews
Sex	Male	467 (46%)	60 (49.6%)
	Female	515 (51%)	59 (48.7%)
	Prefer not to say	28 (3%)	2 (1.7%)
Race/Ethnicity	White/Caucasian	606 (60%)	75 (62%)
	Hispanic/Latino	140 (13.8%)	16 (13.2%)
	Asian/Pacific Islander/Hawaiian	71 (7%)	7 (5.8%)
	Black/African American	54 (5.4%)	4 (3.3%)
	Mixed/Biracial	117 (11.6%)	14 (11.6%)
	Other	22 (2.2%)	5 (4.1%)
Grade Level	Sixth Grade	264 (26.1%)	25 (20.6%)
	Seventh Grade	351 (34.8%)	52 (43%)
	Eighth Grade	395 (39.1%)	44 (36.4%)
Location	Pacific	282 (28%)	44 (36.4%)
	Mountain West	477 (47.2%)	38 (31.4%)
	Midwest	251 (24.8%)	39 (32.2%)

3.3.2. Instrument and Data Collection

The survey was constructed with the intent of integrating and gathering data towards both research questions. It was further modeled after prior usability studies as well as surveys about consequences of testing (Denecke et al., 2021; Thielemans et al., 2018). In Denecke and coauthors' study, survey questions resulted in binary responses, such as "confusing/clear", "not understandable/understandable", and "cluttered/organized". Thielemans et al. (2018) survey related to a tool's ease of use, and readability/comprehension, which was mirrored in the current study focusing on the usability of tools associated with the test (i.e., calculator and formula sheet), readability of items and test directions, and overall test usability. Survey questions are presented in Figure 4.

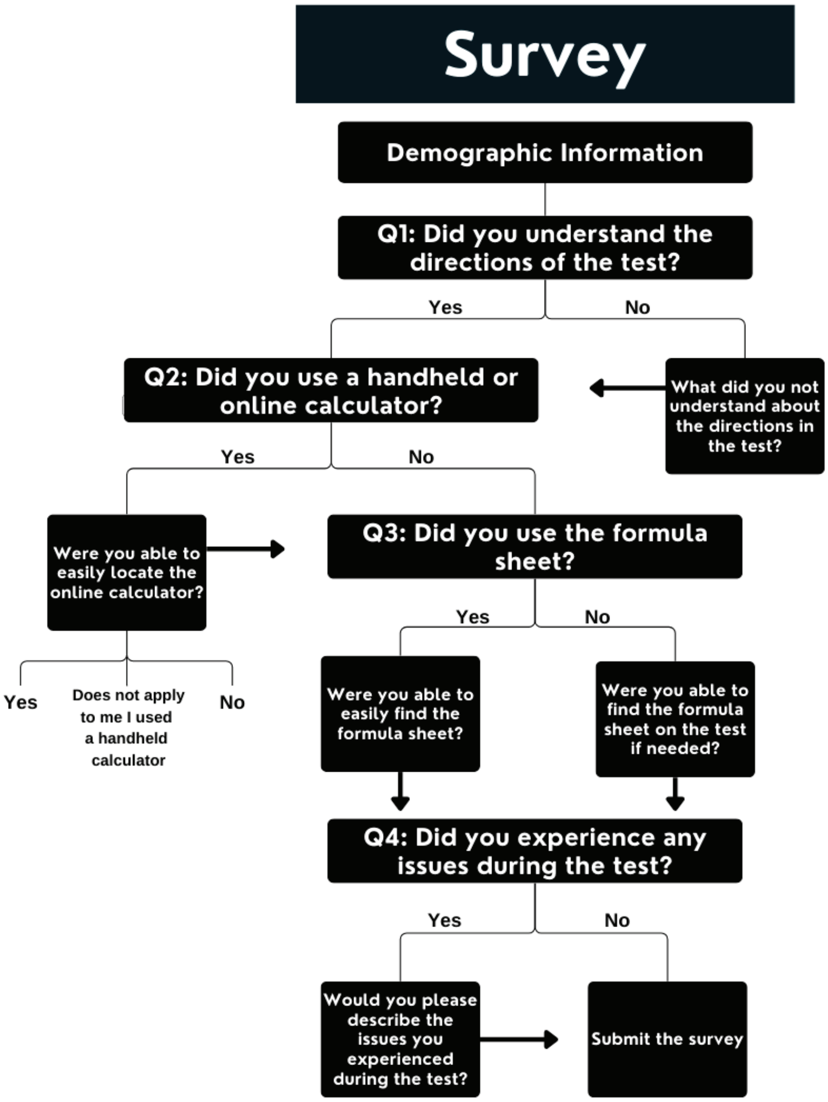


Figure 4. Survey Questions and Skip Logic.

Classroom teachers distributed the survey to students immediately following the completion of the PSM-CAT through a hyperlink. Students were asked to provide demographic information. Then, they began responding to survey questions related to the test. In total, the survey had four focal questions. The four focal survey questions focused on whether they understood the test and could find appropriate materials (i.e., calculator and formula sheet). Those focal questions were (1) Did you understand the test? (2) Did you use a handheld or online calculator? (3) Did you use the formula sheet? (4) Did you experience any issues during the test? The survey consisted of 'skip logic', which parallels past research (e.g., Ifeachor et al., 2016; O'Regan et al., 2020). An additional five questions branched from the focal questions depending on student response. For example, if students answered 'no' to a question, then they would be given a different question than if they responded 'yes' (see Figure 4 for more information). Most students completed the survey within four minutes.

3.3.3. Data Analysis

Descriptive statistical analyses were performed on the closed-ended survey data to examine how students responded to each binary choice (yes/no). Open-ended survey questions were analyzed through an inductive, multi-stage thematic analysis approach (Creswell, 2014; Miles et al., 2014). This five-segment approach included multiple steps (see Figure 5). A goal of this approach was to have multiple opportunities to review the qualitative data, create observations, take notes, and generate themes from the data. Collaboration across two researchers (i.e., King and Bostic) was also included in this process to provide opportunities to evaluate and critique the processes and outcomes that were generated. In addition, results during qualitative data analysis were shared with co-authors (i.e., May and Stone) to promote triangulation (Creamer, 2017; Miles et al., 2014).

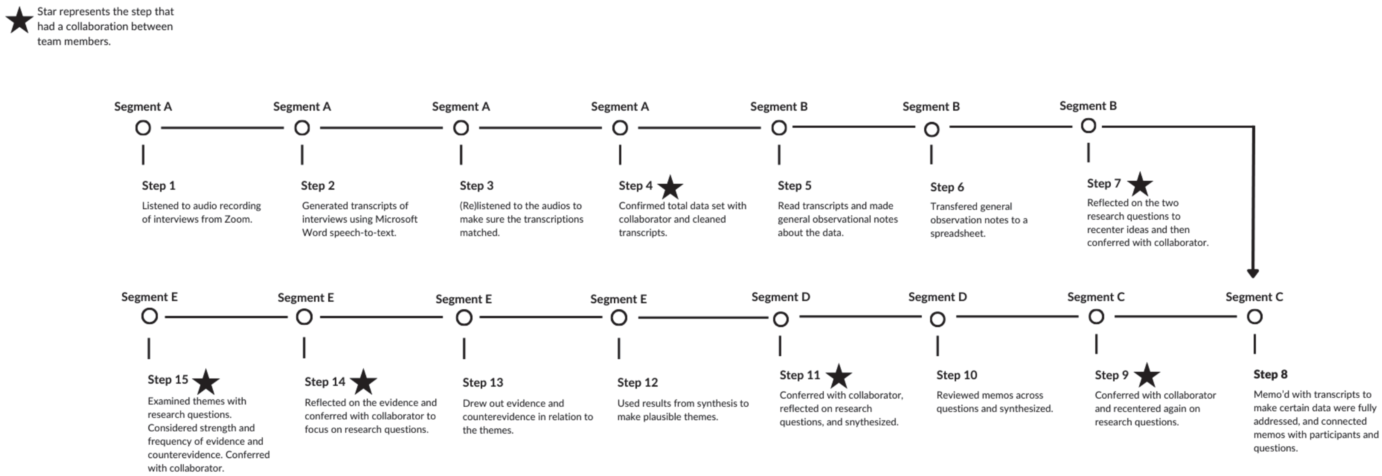


Figure 5. Multistage Qualitative Analysis Process.

Each segment consisted of a series of steps, which are shown in five segments (A through E). Two researchers conferred at the end of each segment (see Figure 5). Segment one was used to become intimately familiar with the data and prepare data for analysis. Segment two consisted of making generalizable notes on a spreadsheet. Segment three was to draw together the notes into memos. Segment four was used through synthesizing the created memos into codes. The fifth and final segment consisted of making plausible themes from the codes in relation to the research questions.

3.4. Interviews: Data Collection and Analysis

Students were purposefully and representatively chosen for an interview and asked to confirm willingness to participate as volunteers. An aim for sampling interviewees was to reflect the demographics of the surveyed participants. In total, 121 students participated in 1–1 interviews with an intent to gather data on students' perceptions of the PSM-CAT (see Table 1). The purpose of this interview was to gather data for both research questions. Lee's interview questions (Lee, 2020) were used as a model for the interview protocol in this study, which focused on aspects related to testing consequences and usability. Interview questions are found in the Appendix A. Data collection started when two researchers (i.e., King and Bostic) asked teachers whether any students volunteered for interviews. Teachers provided a list of volunteers. As students were selected, they were asked to confirm participation in the interview on their way to the interview space. Students were escorted one-at-a-time to a quiet space for the interview.

The researchers used a structured protocol that asked questions about students' perceptions and experiences with the PSM-CAT. First, they were asked to confirm their participation in the interview. After confirming their participation, they were asked to provide information related to their demographic data and to choose a pseudonym. Next, researchers provided an overview of the interview. Participants were handed a paper with the purpose of the PSM-CAT and interview questions (see Appendix A). A researcher confirmed whether the participant understood the purpose of the PSM-CAT and could read the questions. The interview started after students communicated their understanding. If students had any questions, then they were answered. Researchers redirected participants when necessary. Interviews took approximately five minutes.

Similar to the interview, a thematic data analysis approach (Creswell, 2014; Miles et al., 2014) was used to analyze the data from 1–1 interviews (see Figure 5). The same process described earlier was used with the interview data. We connect students' responses to interview questions and communicate when students' ideas were prompted by the final question, "Is there anything that you want to share with me about your experience during the test?" We frame responses to that final question as 'unprompted' because they are not necessarily resulting from a targeted interview question.

4. Results

Data analysis led to themes and ideas for RQ1 and RQ2.

RQ1: *To what degree is the PSM-CAT usable for students?*

RQ2: *What consequences of testing evidence exist for the PSM-CAT?*

To summarize the findings, the theme for RQ1 was that students perceived the PSM-CAT with a high degree of usability. There were two ways that usability was framed: resources and test items. Data informing RQ2 led to two themes. The first theme was that students experienced positive outcomes from taking the PSM-CAT. These positive outcomes are grounded in two ways: student learning and student attitudes. The second theme was that students believed their teachers might gain information as a consequence of the PSM-CAT. This second theme was grounded in two ways. First, students felt that teachers might learn what students understand mathematically from the test results. Second, students felt their teachers might be better equipped to help students mathematically grow. We display these themes and ideas behind the themes in Figure 6. Quotations are intentionally selected to demonstrate consistency and coherence across grade levels and are shared in greater detail through sections below. As ideas are discussed, the pseudonyms that participants chose are used. However, an individual's demographic information (e.g., sex and race/ethnicity) are not shared to (a) protect anonymity and (b) to follow current Executive Order 13985 https://www.whitehouse.gov/presidential-actions/2025/01/ending-radical-and-wasteful-government-dei-programs-and-preferencing/ (accessed on 23 February 2025). This executive order directs the federal government to eliminate diversity, equity, and inclusive programs and policies that promote discrimination (White House, 2025), which includes scholarships stemming from federally grant-funded research.

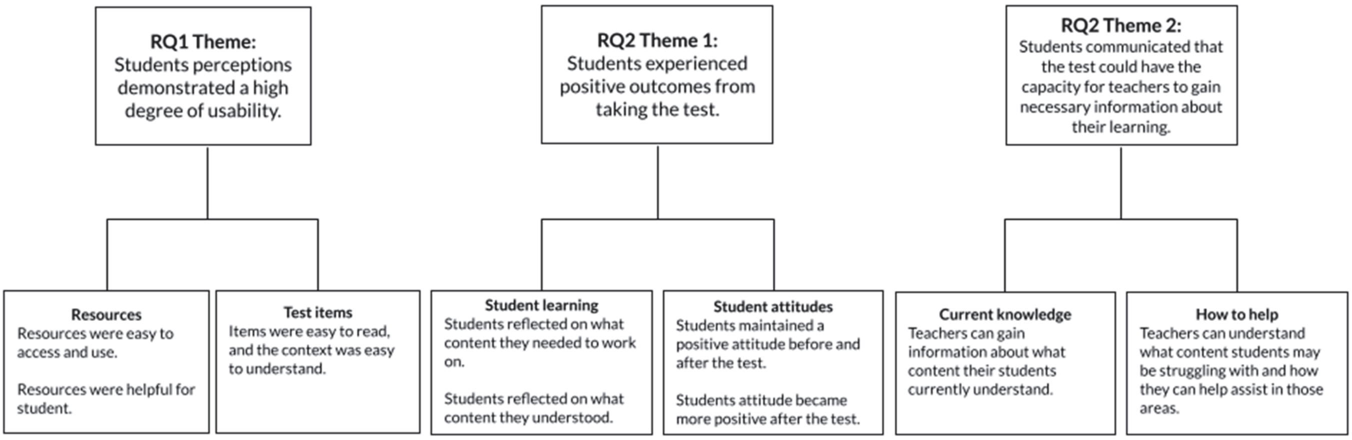


Figure 6. Themes Responding to Research Questions.

4.1. RQ1: Usability

Data analysis led to a single theme for RQ1: Students' perceptions of the PSM-CAT demonstrated a high degree of usability. This theme was seen through two ideas: (a) resources and (b) items. Results for RQ1 were found through quantitative survey data related to usability features, and integrated with qualitative data about usability (e.g., students could have responded to the open-ended interview questions related to usability or consequences of testing). Test usability evidence also came across in a broad sense from several students. Cornetheus, a seventh-grade student communicated that he preferred PSM-CAT over other tests he has taken recently, "I definitely do prefer that [CAT] test over our usual paper tests, because, you can see that it definitely makes an impact based on the way that it was set up and the way the time limit is, and the questions on the screen worked well. I mean, like, I knew what to do." Test usability was also stated in a similar sense by Michael, an eighth-grade student who shares his thoughts about the usability of the time limit, "I didn't get many questions done, but that is okay because the time that I needed was enough for every question. It was actually content that I should know and knew. That's different from other tests where I feel I have to rush to answer every problem". Kennedy, a sixth-grade student, added comments about the directions and flow of the test: "the test [directions] explained most stuff and there were pretty good explanations of the problems. . .it was easy to get through the test and I knew what to do to get to the next problem." In summary, students conveyed a high degree of test usability from the directions to the navigation to the timer and time limit.

4.1.1. Resources

A finding for RQ1 was that resources (i.e., online calculator and formula sheet) provided in the PSM-CAT system were accessible and easy to use. This is also evidenced through students who perceived the resources on the PSM-CAT as helpful. With regards to accessibility, the majority of students easily found the online calculator and formula sheet, and could use them (see Figure 7). In total, 856 of the 1010 students used a calculator. There were 662 students who used the online calculator and 194 students who used a handheld calculator. Of those 662 students that used the online calculator, 97% (n = 642) of survey respondents indicated they were able to locate it. Similarly, there were 694 of the 1010 students who wanted to use the formula sheet. Of those 694 students, 82% (n = 569) were able to access it with ease. Interview data complemented this finding of resources being easily accessible and usable. John, a seventh-grade student, was one of many students who shared that he thought there were more benefits to the test during an unprompted response: "I liked how when I used the calculator on the test, it didn't take up the whole screen. I could see the question while also seeing the calculator instead of having to memorize what I need to type in and then go back to the calculator. I also think that the resources were just organized well because I could find them". Data were consistent across interviews; however, there was a small group of students who wanted to use resources but could not locate them (1.6%, n = 16). The survey data also included information about whether students could understand the directions as a usability feature, and if they had issues with the testing system. Results showed that 98% of students surveyed could clearly read and understand what the test directions were, and less than 3% of students experienced issues with the testing system. An example of the biggest issue (i.e., 1.7%) was that students could not locate the resources while taking PSM-CAT.

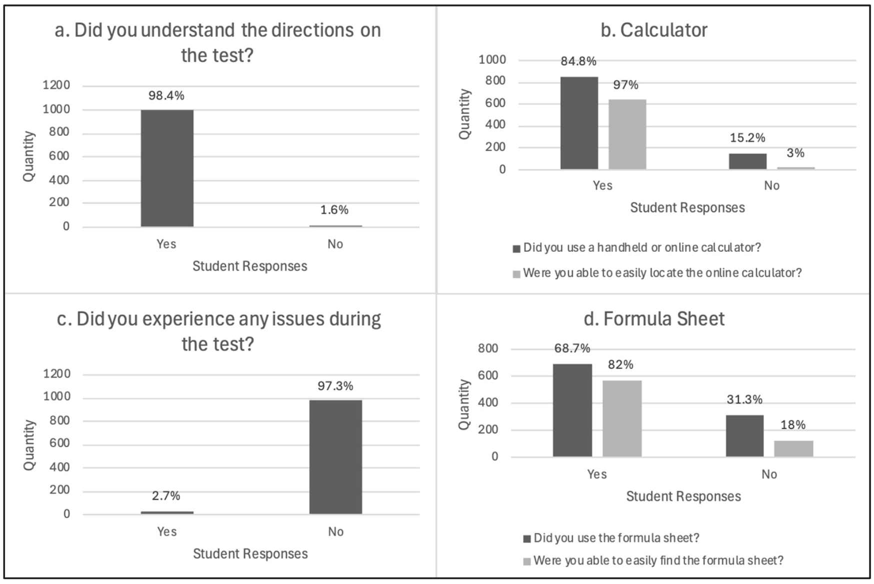


Figure 7. Survey Responses (**a**) results for understanding the directions; (**b**) results for using and finding the calculator; (**c**) results for experiencing issues; (**d**) results for using and finding the formula sheet.

The finding that the PSM-CAT was usable was also described by students who expressed that resources were helpful to them while test-taking, often at the end of the interview. We found seven comments from interviewed students, which were connected to the other data from the survey. Four of these comments were captured from the interview questions themselves, and three of these comments about the resources were found in the data when asked if they had anything else they would like to share. Lucy, a sixth-grade student, conveyed this feeling when asked if she had anything else to share about the test overall, "There were good resources that you could use while taking the test which, helped me while problem solving". Similarly, Peter, a seventh-grade student, said this quote when asked if he had anything else to share, "On the test I got to use a calculator. It was helpful to focus on the actual question because I had a calculator there. It would be more work if you did not have a calculator". Peter and others were able to focus on the test items because of the provided calculator. In summary, students expressed that the resources were usable because they were easy to access and use, and they were valuable tools needed to support students while testing (i.e., 6% of students from interviews, and 97% (calculator ease), 82% (formula sheet ease), were found from survey data).

4.1.2. Test Items

Our finding for the test being highly usable was grounded through a second way related to the test items itself rather than the resources. Students' perceived PSM-CAT items as readable and the item contexts were relatable. Charlie, a seventh-grade student, offered the following at the end of the interview for any final thoughts, "The problems were just fun and they were about fun things like driving and finding the cost of things. I understood the questions because they were about actual things you could use in the world". Charlie and another participant, John, emphasized that the items' contexts connected with their real-world experience or familiarity with a context. John, a sixth-grade student, said in his interview when asked about the perceived benefits of the test (i.e., interview question one) that "The questions were clear and easy to understand, it was very simple. I find that other test's questions are hard to understand or find the information". Faith, an eighth-grade student, shared this sense when asked the same question as John, "I think this test had benefits because the questions were worded in a way that I was able to understand". This idea of the test items being easy to read and understandable was expressed consistently across participants. Interview questions were open-ended, and therefore, led students to discuss information unprompted about the usability of the test as well, leading to these results. This consistency led to the finding that supported a high degree of usability by intended respondents.

4.2. RQ2: Test Consequences

The first theme related to test consequences was that students perceived positive outcomes from the PSM-CAT. A second theme was that students believe their teachers might gain information about their mathematics proficiency from the PSM-CAT. We unpack theme one, then theme two.

4.2.1. Outcomes: Student Learning

Our first theme for RQ2 was that there were positive outcomes from taking the PSM-CAT. Theme one was reified through two ideas. The first idea was that students felt they were able to demonstrate their mathematics learning through PSM-CAT completion. That is, they perceived their test score as reflective of their mathematics learning. Students viewed the test as a coherent map of their mathematics learning. While they did not specifically reference mathematical content in their ideas, their responses were a result of thinking about the mathematics content that they came across in the testing system. Roman, a seventh-grade student, shared, like many others, that the PSM-CAT gave an opportunity to reflect on what mathematical content he knew and areas for growth: "The good from this test is that you experience more things in math, and you see more what you need help with so you can do better". In Roman's case, he was asked about whether the test had more benefits or negatives. His response is evidence of the first idea of theme one that students were able to demonstrate mathematics learning. As a reminder, the PSM-CAT seeks to measure students' abilities accurately and efficiently, which requires administering items that may be somewhat beyond what they have learned. Brielle, a seventh-grade student, shared an experience with the test when also asked about the benefits: "This test has more benefits than issues because you could see what you already know how to do in math". Her comments unpack the idea that the test can remind students what mathematics they know how to do, also evidence of the first idea. Saje, an eighth-grade student, supported this idea: "Good came from the test because it helped me to reflect on how we (Saje and peers) figure out math problems and if I am able to solve them correctly". Statements like these were consistent across the sample and portrayed how students perceived the PSM-CAT to have positive outcomes, such as gaining information about their current content knowledge and content that they may need to work on. There were rare instances where students had negative experiences beyond the control of the PSM-CAT, (1.1%, n = 11). Those experiences include the test not loading or an item failing to open after answering another item. Such experiences may have been due to the school's internet or computer quality rather than the test. This seemed to be limited to a paucity of students across the entire testing sample. Moreover, most students experienced positive outcomes from taking the PSM-CAT, and this was grounded through evidence related to students reflecting on what content they understood and what content they needed to work on.

4.2.2. Outcomes: Student Attitudes

A second idea related to this theme (i.e., students experienced positive outcomes from taking the test) was that the PSM-CAT led to maintaining or promoting positive attitudes. Participating students either maintained a positive attitude from start to finish with the PSM-CAT, or their attitude became more positive (i.e., increasingly positive) compared to when they started. Of the 121 interviewed students, twenty-six (21%) felt positive about the testing experience both before and after the PSM-CAT, whereas eighty-six students (71%) described a positive increase in their attitude towards the PSM-CAT after completing it. Taken collectively, 92% of students interviewed had positive attitudes towards the testing experiences. This finding was a result of interview questions about how students felt before and after the testing experience. Students shared a variety of information characterizing why they felt positive. A seventh-grade student, Jane, told the interviewer: "I felt positive towards the test because as you answered a question, they got harder and harder to match your level. I liked that and it made me feel good". This student felt positive about the experience because the test was computer-adaptive; she was able to answer the PSM-CAT items she could, and some items were beyond her current abilities. Alternately, Trinity, a sixth-grade student, simply expressed: "I got a lot of questions done, and I felt good after the test". Overall, most students expressed a positive attitude after completing the PSM-CAT.

4.2.3. Teachers Gain Knowledge: What Students Know

RQ2 had a second theme. Students believed their teachers might gain information about their mathematics proficiency as a result from the PSM-CAT. This theme was grounded through the first idea reflecting that teachers could gain information about what content their students currently understand. Bianca, an eighth-grade student told an interviewer when asked about the benefits of the test, "What the test is trying to do is help teachers understand what's going on. You know—through a student's head, and how they (students) think. Personally, I think that's it's important for my teacher to know". Participants expressed that teachers may gain information about their students' thinking as a result from completing the PSM-CAT. Harper, a sixth-grade student also answered this question, "There are benefits and positives to this test because your teacher can find out how you learn and what your math level is at. I want my teacher to know what I know". Interviewed students consistently communicated a strong desire for their teachers to gain information about what content they knew and understood.

4.2.4. Teachers Gain Knowledge: How to Help Students

The second theme to RQ2 had a second idea that acted as supporting evidence for this theme. Students believed the PSM-CAT would result in information for teachers about helping with student mathematics learning. That is, help with content that students learned previously. Students were reflecting more on how their teachers can take the results from the PSM-CAT and help them with the content that they may not have understood in that moment. Collin, a seventh-grade student, communicated that "This test helps my teacher realize what I'm struggling in (with math). If they (my teacher) can understand that, then they can see and figure out what I need help with. I hope my teacher uses the test results to help me where I need it". Rachel, an eighth-grade student said something similar: "The test will help teachers, like mine, figure out what to help students, like me, with math I'm learning, and that they (me and my peers) need help with some things—like those problems

I got wrong”. Neither of these students were asked interview questions specifically related to their teachers gaining knowledge; however unprompted, these students and others responded in such ways in relation to the benefits of testing or how they felt during the test. In summary, the findings from interview data characterized positive consequences, such as the PSM-CAT allowing their teachers to learn what students can do in mathematics and areas where they may need help.

5. Discussion

One goal of this manuscript was to explore usability outcomes related to the PSM-CAT. A second goal was to share validity evidence related to the consequences of testing for the PSM-CAT. Reporting on these two goals provides potential users—researchers, evaluators, and school personnel—the necessary information to make informed choices related to a CAT measure of mathematical problem-solving. Evidence from this mixed-methods study provides a strong foundation for findings related to the two goals. That is, (1) the PSM-CAT had a high degree of usability by potential users; (2) the positive outcomes related to consequences of testing outweigh negative outcomes related to the PSM-CAT.

5.1. Connecting to Consequences of Testing

Accountability and computer-based testing are ubiquitous phenomena in the modern education landscape. Students take tests that have outcomes; hence, it is important that validity evidence is examined from a broad perspective, as described in the *Standards* (AERA et al., 2014). Test results and their interpretations that lack robust validity evidence can lead to spurious findings or worse, harmful outcomes (Bostic, 2023). Krupa et al. (2024) have shown that consequences of testing information related to mathematics tests is rarely shared in the literature, which makes this study helpful as a contributor to the discussion about consequences of testing. This study also extends prior research on consequences of testing (e.g., Lee, 2020). The findings from the present study indicated that there was an overall positive experience from testing and outcomes from the PSM-CAT. Students communicated that they developed or maintained a positive attitude while testing, and that the test has the capacity for teachers to gain information about their learning. Previous research findings suggested substantive negative consequences from testing are possible (see Lee, 2020) and should be sufficiently explored. Lee’s (2020) students conveyed concerns about how test results might be used, and roughly half felt the consequences were positive. Lee’s findings contrast our findings in which less than 3% of students communicated any negative experiences or outcomes from the test. Tests should have more positive consequences than negative ones.

Narratives from test respondents provide potential test users with greater confidence that the benefits of testing with PSM-CAT are greater than the negatives associated with it (e.g., issues while testing). There were several unprompted responses from participants during the interview regarding consequences, suggesting that consequences of testing was something they considered and felt important enough to convey to interviewers. If teachers are expected to administer a test, then the time taken during testing should be buttressed with reasonable evidence that the testing time is worth it. Conclusions like this one are called for in research (e.g., Carney et al., 2022) and warranted to support robust validity claims about consequences of testing that have been rare in literature (Krupa et al., 2024) or highlight greater positives than negatives unlike Lee (2020). Findings from this study provide readers and potential PSM-CAT users with confidence that the results from these tests are being used appropriately. This study also provided an example of researching consequences of testing for test results and, in turn, it is a case where the positive consequences from a test were greater than the negative consequences.

5.2. Usability Studies in Education Research

The present study extends past usability studies from other areas into education. Denecke et al.'s (2021) work, as well as Thielemans et al.'s (2018) usability study, provided a foundation for the present study. The *Standards* (AERA et al., 2014) indicate that test developers should explore usability. Scholars (e.g., Carney et al., 2022; Estrada-Molina et al., 2022) have expanded on these guidelines and recommend that studies clearly communicate usability-related information about tests. One outcome from the present study is an example of a convergent mixed-methods project that highlighted areas of strength and areas for improvement related to a mathematical problem-solving test. A second, and important outcome, is that this study serves as an example for other usability studies within educational testing research. Usability investigations from medicine and technology can be reasonably applied with some modifications to educational testing situations, and this study may serve as one way to translate from other research areas. Educational scholars might find the present study useful for conducting their own usability studies. Past usability studies described acceptable usability features, but also characterized potential improvements (Thielemans et al., 2018). Our study also found that the PSM-CAT had a high degree of usability due to the nature of the resource accessibility and test design; however, some students (less than 2%) expressed struggles locating the formula sheet. This finding led the development team to move the formula sheet's location and convey its location more clearly in the directions. In effect, usability studies present a form of design research around validation work that can shape better outcomes for future users.

5.3. Limitations and Future Directions

We share some limitations of the study, which could be improved with future work. First, the survey did not include statements leading to quantitative data for consequences of testing. At the same time, the interviews generated qualitative data for test usability. This study drew upon both data sources (i.e., interviews and surveys) to convey a broader narrative around the test. A limitation is a lack of quantitative data related to consequences of testing and a lack of qualitative data related to usability. Our research team made those methodological choices by drawing from past research. A future study might develop a survey with items related to consequences of testing and interview questions related to usability. Second, the data set was drawn from three diverse districts (i.e., rural, suburban, and urban). There are some areas where the data could be strengthened, including greater breadth in student backgrounds such as increasing numbers of students from urban and rural populations. A future investigation might better study outcomes with these populations and explore the degree to which the findings from the present investigation extend. A third limitation was related to the usability results. Our findings discuss that the test demonstrated a high degree of usability with the features that were studied. To improve results, additional research including more usability features could have the capacity to improve the PSM-CAT. More usability studies need to be performed in education, specifically and especially in CAT environments. Our hope is that this study might serve as a model for more usability studies in education.

Author Contributions: Conceptualization, S.G.K., J.D.B., T.A.M. and G.E.S.; Methodology, S.G.K. and J.D.B.; Validation, J.D.B.; Formal analysis, S.G.K. and J.D.B.; Investigation, S.G.K. and J.D.B.; Resources, S.G.K., J.D.B. and T.A.M.; Data curation, S.G.K.; Writing—original draft, S.G.K. and J.D.B.; Writing—review & editing, J.D.B., T.A.M. and G.E.S.; Visualization, S.G.K.; Supervision, J.D.B.; Project administration, J.D.B.; Funding acquisition, J.D.B., T.A.M. and G.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science Foundation grant numbers [2100988; 2101026; 1920621; 1920619]. The APC was funded by [1920621].

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Bowling Green State University (protocol code #1749616-7 on 11 October 2024). The Bowling Green State University Institutional Review Board determined this research is exempt according to federal regulation AND the research met the principles outlined in the Belmont Report.

Informed Consent Statement: Consent/assent from participants was waived due to the research meeting exempt status according to federal regulations. All participants were provided information about the study prior and had the option for their data not to be used as part of the study.

Data Availability Statement: Data are not publicly available.

Conflicts of Interest: Author Gregory E. Stone was employed by the company Clarity Assessment Systems Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAT	Computer-adaptive test
FGD	Focus group discussions
IRB	Institutional review board
MMR	Mixed methods research
USA	United States of America

Appendix A

Purpose Statement and 1–1 Interview Questions

Interview Protocol:

Purpose Statement (read by students)

A purpose of this test is to give you and your teachers information about your math problem solving. This test may help your teacher learn about your ability to solve math problems. It may also help your teacher design instruction to support you and your peers. Scores from it will never be used for a grade.

Overview Protocol (read by a researcher to students)

In this interview, you will be answering questions based on your experiences during the test, such as if you were stressed, frustrated, etc. The questions will mostly be yes or no, and then you may be asked to explain your response more in-depth. On my computer here I have an app open that will capture our audio, but it will not be video recording. On the tab next to the audio, I will be transcribing what we are saying on a word document. The researchers are the only ones who will listen to the audio so we can remember what was said and match it to the transcription. Do you understand the process, and do you have any questions before we start?

Interview questions (read by a researcher to students)

- Do you think the benefits from taking this test are greater than any negative effects that you experienced?
- How did you feel before you started the test?
- How did you feel after completing the test?
- Did you have any challenges during the test that were more difficult than what you usually experience during a math test?
    - If yes: What were those challenges?
        - How did you work through those challenges?
- Did you feel frustrated during the test, more than you usually do during a math test?
    - If yes: When did you get frustrated?
        - What do you think caused that feeling?
- Did you feel any anxiety during the test, more than you usually do during a math test?
    - If yes: What do you think caused that anxiety?
- Did you feel any stress during the test, more than you usually do during a math test?
    - If yes: What do you think caused this stress?
- Is there anything that you want to share with me about your experience during the test?

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (ECME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. [CrossRef]

Bostic, J. (2023). Engaging hearts and minds in assessment research. *School Science and Mathematics Journal*, *123*(6), 217–219. [CrossRef]

Bostic, J., May, T., Matney, G., Koskey, K., Stone, G., & Folger, T. (2024, March 6–8). *Computer adaptive mathematical problem-solving measure: A brief validation report*. 51st Annual Meeting of the Research Council on Mathematics Learning (pp. 102–110), Columbia, SC, USA.

Bostic, J., & Sondergeld, T. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics Common Core. *School Science and Mathematics Journal*, *115*, 281–291. [CrossRef]

Bostic, J., & Sondergeld, T. (2018). Validating and vertically equating problem-solving measures. In D. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe* (pp. 139–155). Springer.

Bostic, J., Sondergeld, T., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement*, *18*(2), 151–162.

Bruder, R. (2021). Comparison of the Abitur examination in mathematics in Germany before and after reunification in 1990. *ZDM Mathematics Education*, *53*, 1515–1527. [CrossRef]

Carney, M., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, *53*(4), 334–340. [CrossRef]

Céspedes-González, Y., Otero Escobar, A. D., Ricárdez Jiménez, J. D., & Molero Castillo, G. (2023, November 6–10). *Academic achievement in mathematics of higher-middle education students in veracruz: An approach based on computational intelligence*. 11th International Conference in Software Engineering Research and Innovation (CONISOFT) (pp. 177–185), León, Mexico. [CrossRef]

Common Core State Standards Initiative (CCSSI). (2010). *Common core standards for mathematics*. Available online: http://www.corestandards.org/Math/ (accessed on 5 September 2024).

Creamer, E. G. (2017). *An introduction to fully integrated mixed methods research*. SAGE Publications.

Creswell, J. W. (2014). *Research design: Quantitative, qualitative, and mixed method approaches* (4th ed.). SAGE Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Davey, T. (2011). *A guide to computer adaptive testing systems*. Council of Chief State School Officers.

Denecke, K., Vaaheesan, S., & Arulnathan, A. (2021). A mental health chatbot for regulating emotions (SERMO)—Concept and usability test. *IEEE Transactions on Emerging Topics in Computing*, *9*(3), 1170–1182. [CrossRef]

Estrada-Molina, O., Fuentes-Cancell, D. R., & Morales, A. A. (2022). The assessment of the usability of digital educational resources: An interdisciplinary analysis from two systematic reviews. *Education and Information Technologies*, *27*, 4037–4063. [CrossRef]

Fennell, F., Kobett, B., & Wray, J. (2023). *The formative 5 in action, grades K-12. Updated and expanded from the formative 5: Everyday assessment techniques for every math classroom*. Corwin.

Heissel, J. A., Adam, E. K., Doleac, J. L., Figlio, D. N., & Meer, J. (2021). Testing, stress, and performance: How students respond physiologically to high-stakes testing. *Education Finance and Policy*, *16*(2), 183–208. [CrossRef]

Hino, K., & Ginshima, F. (2019). Incorporating national assessment into curriculum design and instruction: An approach in Japan. In C. P. Vistro-Yu, C. P. Vistro-Yu, T. L. Toh, & T. L. Toh (Eds.), *School mathematics curricula* (pp. 81–103). Springer Singapore Pte. Limited. [CrossRef]

Hudson, J., Nguku, S. M., Sleiman, J., Karlen, W., Dumont, G. A., Petersen, C. L., Warriner, C. B., & Ansermino, J. M. (2012). Usability testing of a prototype phone oximeter with healthcare providers in high-and low-medical resource environments. *Anaesthesia*, *67*(9), 957–967. [CrossRef]

Ifeachor, A. P., Ramsey, D. C., Kania, D. S., & White, C. A. (2016). Survey of pharmacy schools to determine methods of preparation and promotion of postgraduate residency training. *Currents in Pharmacy Teaching and Learning*, *8*(1), 24–30. [CrossRef]

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education/Praeger.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. [CrossRef]

Krupa, E., Bostic, J., Folger, T., & Burkett, K. (2024, November 7–10). *Introducing a repository of quantitative measures used in mathematics education*. 46th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 55–64), Cleveland, OH, USA.

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, *26*(1), 127–135. [CrossRef]

Lane, S. (2020). Test-based accountability systems: The importance of paying attention to consequences. *ETS Research Report Series*, *2020*(1), 1–22. [CrossRef]

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, *21*(1), 23–30. [CrossRef]

Lawson, B., & Bostic, J. (2024). An investigation into two mathematics score reports: What do K-12 teachers and staff want? *Mid-Western Educational Researcher*, *36*(1), 12. Available online: https://scholarworks.bgsu.edu/mwer/vol36/iss1/12 (accessed on 18 December 2024).

Lee, E. (2020). Evaluating test consequences based on ESL students' perceptions: An appraisal analysis. *Columbia University Libraries and Applied Linguistics and TESOL at Teachers College*, *20*(1), 1–22. [CrossRef]

Lesh, R., & Zawojewski, J. S. (2007). Problem solving and modeling. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763–804). Information Age Publishing.

Maphalala, M. C., & Khumalo, N. (2018). *Standardised testing in South Africa: The annual national assessments under the microscope*. ResearchGate. Available online: https://www.researchgate.net/publication/321951815 (accessed on 12 May 2025).

Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of Educational Psychology*, *110*(1), 27. [CrossRef]

McGatha, M. B., & Bush, W. S. (2013). Classroom assessment in mathematics. In *SAGE handbook of research on classroom assessment* (pp. 448–460). SAGE.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.

Miles, M., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Naveh, K. (2004). *Matriculation in a new millennium: Analysis of a constructivist educational reform in Israeli high-schools* [Doctoral dissertation, University of Leicester and ProQuest Dissertations Publishing]. Available online: https://www.proquest.com/docview/U190040 (accessed on 2 December 2024).

O'Regan, S., Molloy, E., Watterson, L., & Nestel, D. (2020). 'It is a different type of learning'. A survey-based study on how simulation educators see and construct observer roles. *BMJ Simulation & Technology Enhanced Learning*, *7*(4), 230–238. [CrossRef] [PubMed] [PubMed Central]

Saad, M., Zia, A., Raza, M., Kundi, M., & Haleem, M. (2022). A comprehensive analysis of healthcare websites usability features, testing techniques and issues. *Institute of Electrical and Electronics Engineers*, *10*, 97701–97718. [CrossRef]

Schoenfeld, A. (2011). *How we think: A theory of goal-oriented decision making and its educational applications*. Routledge.

Schoenfeld, A. H. (2015). Summative and formative assessments in mathematics supporting the goals of the Common Core Standards. *Theory Into Practice*, *54*(3), 183–194. [CrossRef]

Sireci, S., & Benitez, I. (2023). Evidence for test validation: A guide for practitioners. *Psicotherma*, *35*(3), 217–226. [CrossRef] [PubMed]

Thielemans, L., Hashmi, A., Priscilla, D. D., Paw, M. K., Pimolsorntong, T., Ngerseng, T., Overmeire, B. V., Proux, S., Nosten, F., McGready, R., Carrara, V. I., & Bancone, G. (2018). Laboratory validation and field usability assessment of a point-of-care test for serum bilirubin levels in neonates in a tropical setting. *Wellcome Open Research*, *3*, 110. [CrossRef]

Uko, M. P., Eluwa, I., & Uko, P. J. (2024). Assessing the potentials of compurized adaptive testing to enhance mathematics and science student't achievement in secondary schools. *European Journal of Theoretical and Applied Sciences*, *2*(4), 85–100. [CrossRef] [PubMed]

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, *1*, 195–229. [CrossRef]

Wainer, H., & Lewis, C. (1990). Toward a psychometrics for testlets. *Journal of Educational Measurement*, *27*(1), 1–14. [CrossRef]

White House. (2025). *Ending radical and wasteful government DEI programs and preferencing [Presidential executive order]*. The White House.

education sciences

MDPI

Article

Interactive Homework: A Tool for Parent Engagement

Laura Moore [1,*] and Robert N. Ronau [2,*]

[1] School of Education, Johns Hopkins University, 2800 N Charles St, Baltimore, MD 21218, USA
[2] Department of Elementary, Middle & Secondary Teacher Education, College of Education and Human Development, University of Louisville, 2301 South 3rd Street, Louisville, KY 40292, USA
[*] Correspondence: laura_moore@hcpss.org (L.M.); bob@louisville.edu (R.N.R.)

Abstract: Families have largely been excluded from the implementation of the Common Core State Standards in Mathematics (CCSSM), reducing their ability to extend their child's mathematics learning. CCSSM emphasizes different instructional elements (e.g., pictorial representations, problem solving, multiple strategies for solving) that may differ greatly from how parents learned mathematics. In addition, many school officials have ineffectively engaged parents in the changes, further diminishing their capacity to participate in their child's learning. This case study examined parent mathematics self-efficacy and parent mathematics knowledge for teaching, factors that influence the effectiveness of their engagement in their child's mathematics learning. This study was also implemented to identify elements that the parent participant found helpful for their child's mathematics learning. A thematic analysis was performed on the data sources, the interactive homework assignments, a survey, observations, a researcher's journal, and an interview to conclude that the interactive homework assignments improved parent mathematics self-efficacy and parent mathematics knowledge for teaching. The parent participant also identified the assignments' side-by-side examples, additional practice, and the easy access of the assignments as features of the intervention that enhanced her ability to support her child.

Keywords: interactive homework assignments; mathematics knowledge for teaching; mathematics self-efficacy; parent engagement

1. Introduction

Mathematics achievement is considered critical for social mobility, career opportunities, and full enjoyment of the world [1–6]. As a discipline, mathematics is also an important economic driver for countries [7–9]. Despite the recognized importance of learning mathematics to their children's future, parents in school settings struggled to support them in learning mathematics [10–12]. The content, as well as the ways in which elementary mathematics was taught, changed with the implementation of the Common Core State Standards for Mathematics (CCSSM) [13]. For example, CCSSM focused on helping students develop a conceptual understanding of mathematics through physical and pictorial representations while also ramping up the mathematics content at each grade level. Furthermore, teachers with limited mathematics understanding required additional learning to use conceptual strategies; however, that notion was not always embraced by teachers, resulting in sometimes dated and ineffective mathematics explanations and instruction [14,15].

This study addresses how parents faced multiple barriers to helping their children with mathematics such as: the onslaught of the COVID-19 pandemic, the accompanying closing of schools, and the hasty implementation of remote instruction, which required more home support than ever before. This study was designed to support learning at home by helping parents move past traditional mathematics strategies such as algorithms and rote learning into science-based teaching methods that promote conceptual understanding of mathematics constructs. Based on Hatano and Inagaki's [16] framework on conceptual development, conceptual strategies promote a deeper understanding that is transferable and generalizable in which students can reason about mathematics by identifying relationships and patterns. In contrast, those wedded to rote memorization and traditional strategies often struggle to apply different approaches when solving novel problems and can be challenged to judge the reasonableness of their answers [17]. This study also examines ways to effectively support families experiencing barriers associated with the pandemic and remote instruction (e.g., technological challenges, delivering lessons intended for face-to-face instruction, and reduction in teacher capacity and accessibility). Additionally, this study aims to help parents engage in dialogue that more closely resembles classroom discussion that supports their child's critical thinking about mathematics.

There were two parts to the study: the needs assessment and the intervention. Prior to the development of the intervention study, a needs assessment was conducted to better understand what parents needed to help their children in learning mathematics. This needs assessment served as a preliminary tool to determine whether the school population's needs were consistent with the parent needs highlighted by Goldman and Booker [10], Jackson and Remillard [11], and Remillard and Jackson [12]. The assessment was distributed to 67 parents of third-, fourth-, and fifth-grade students from an affluent and diverse suburban elementary school. Approximately 93% of the online parent participants had a bachelor's degree (n = 21), a master's degree (n = 24), or a doctorate (n = 12) as their highest level of educational attainment.

The online survey of open-ended questions (n = 7) and close-ended Likert-scale items (n = 19) and follow-up interviews with six in-person interviews indicated that parents had lower levels of mathematics self-efficacy regarding supporting their students' mathematics learning of conceptual strategies and desired more instructional support in the form of textbooks and more homework that they could use with their children. Mathematics self-efficacy is the belief that one will be successful in mathematics and in performing mathematics tasks in general [18], and low mathematics self-efficacy is linked to low self-concept and high mathematics anxiety [19]. Parent responses also suggested misunderstandings about pictorial mathematics representations and their purpose. For example, one parent stated how he wished students learned useful mathematics such as percentages. This parent did not know that percentages are introduced in later grades after they have learned pictorial representations that support their understanding of percentages. Other parent participants emphasized the sole importance of arriving at the correct answer without understanding that pictorial representations help students develop a conceptual understanding so they can more accurately and more meaningfully use abstract algorithms. Additionally, parents were unaware that algorithms are taught after students developed an understanding of the meaning of the operations of addition, subtraction, division, and multiplication through physical and pictorial strategies. For example, in kindergarten, students are introduced to addition and subtraction through sets of physical objects (e.g., base-10 blocks, counters, and snap cubes) and pictures of various objects. Through a series of experiences comparing differences between two groups of these objects (e.g., visually, counting, and using balance beams), students begin to develop an understanding of addition and subtraction. The same materials serve to help students develop the meaning of place value. Understanding the concept of place value and the ideas behind the mathematical operations of addition and subtraction provides students with the understanding to use them in new settings and manage algorithms in subsequent grades.

An interactive homework intervention was shown to be effective for improving parent mathematics self-efficacy and parent mathematics knowledge for teaching. Interventions that resulted in improved parent mathematics self-efficacy included self-guiding tools [20–22] and emotional supports [23,24]. Interventions that included collaborative learning [25,26], engaging mathematics tasks [27,28], and guidance on the development of mathematics knowledge of content and mathematics knowledge of teaching [29,30] were also effective in developing parent mathematics knowledge for teaching. From the needs assessment and extant research, the following research questions served as the guide for the study: "In what ways does a homework intervention change perceived parent participant mathematics self-efficacy?"; In what ways does a homework intervention change perceived parent participant mathematics knowledge for teaching?"; and "What components of an interactive homework assignment program do parent participants identify as useful in helping them support their children with mathematics learning at home?" The interactive homework intervention addressed these questions through a focus on parent mathematics self-efficacy and parent mathematics knowledge for teaching by addressing misunderstandings and improving their understanding of the purpose of pictorial representations to support conceptual understandings.

2. Study Design

Due to the COVID-19 pandemic, the sample consisted of a parent and child partnership, which required a design change from a quasi-experimental concurrent mixed methods study to a descriptive, single-case holistic case study. This specific case study design is appropriate for examining one unit of analysis (e.g., parent participants) that is influenced by real world contexts [30]. This six-week case study began when parent participants were recruited with surveys about their experiences and backgrounds with traditional and conceptual mathematics strategies. Only one parent, Linda, was interested in participating in this project with her daughter, Laura Jean, a nine-year-old enrolled in a fourth-grade mathematics class. Linda and Laura Jean completed the required consent and assent documents. Then from February to March 2021, each week, Linda was emailed one interactive homework assignment, the first containing addition and subtraction problems, for completion at the end of each week. Six interactive homework assignments were sent in total. Zoom observations of the homework sessions, five in total and approximately 30 min in length, were conducted with the two participants using the interactive homework assignments. The researcher's observations of the participants occurred after the first week to clarify homework assignment directions and provide guidance when necessary. All observed sessions were recorded. A final Zoom session was used to interview Linda about her reported experiences with the interactive homework assignments. In addition, a reflective journal was kept by the researcher to capture the changes over time in the parent and child's interactions using the interactive homework assignments during the Zoom sessions. The researcher primarily served as a passive observer to gain an understanding of the assignment's effectiveness in guiding participant interactions and parent learning. She occasionally intervened to maintain fidelity by encouraging the parent participant's faithful adherence to assignment directions. All data sources—parent comments on the survey and interactive homework assignments, transcripts of observations, the researcher's journal, and the interview transcript—were analyzed. A thematic analysis was performed in which all data sources were reviewed to identify patterns which were then grouped, coded, and analyzed. Member-checking occurred with the final parent interview, and peer debriefing between the researcher and her advisor occurred at various stages of the study: homework assignment modifications, review of participant transcripts, and coding and analysis processes. Lastly, the researcher's weekly observations of the participants discussing and collaboratively solving their problems over Zoom further created a comprehensive understanding of parent mathematics self-efficacy and mathematics knowledge for teaching.

2.1. Participants

Linda is a White parent of two children, and Laura Jean, one of her daughters, was enrolled in a fourth-grade-level mathematics class in an affluent, suburban elementary school in Maryland. Linda has a master's degree in social work, and her mathematics background is extensive, having minored in mathematics in college. She was well-versed in traditional mathematics strategies focused on procedural knowledge but struggled with the conceptually focused activities her children brought home. She chose to participate in this study to learn conceptual mathematics strategies as the parent who primarily helps her children with mathematics at home. School district initiatives led to the rapid transition from in-class instruction to remote instruction due to the pandemic. These changes resulted in additional challenges (e.g., limitations to differentiating classroom instruction, limitations in student collaboration, and technical difficulties) in supporting Laura Jean's learning.

2.2. Researcher Identity

This study was conducted by Laura Moore, a National Board Certified educator with eight years of teaching experience. She is African American from a southern family and is third-generation college-educated. She would be considered middle class. Although Laura differs in race from the participants, similarities in social class and formal education reduced power imbalances between the researcher and parent participant. Given these factors, trust was quickly established between the researcher and the parent participant.

2.3. Interactive Homework

Interactive homework activities were used to address parents' mathematics knowledge for teaching and self-efficacy. These two constructs are important in helping parents with their children's homework. Ball et al.'s [31] mathematics knowledge for teaching (MKT) theory describes the mathematics understanding educators need for school instruction. Mathematics competence influences the role parents play in reinforcing mathematics instruction, and De Corte et al. [32] identified five necessary elements of mathematics competence: (a) positive mathematic beliefs, (b) specific mathematical knowledge, (c) heuristic methods, (d) metacognition, and (e) self-regulatory skills. De Corte et al. [32] reported that positive mathematics beliefs include high levels of self-concept and self-efficacy. Specific mathematical knowledge involves an understanding of mathematical constructs (e.g., symbols, procedures, and concepts) and the ability to make generalizations and use heuristic methods to make problem solving more meaningful.

Mathematics knowledge for teaching is divided between subject matter knowledge and pedagogical content knowledge, two dimensions originating from Ball et al.'s [31] MKT framework, which is based on Shulman's [33] pedagogical content knowledge (PCK) construct, which emphasizes the connection between knowledge of a subject and the ability to teach that subject. Within these dimensions are six domains: (a) common content knowledge (CCK), (b) horizon content knowledge (HCK), (c) specialized content knowledge (SCK), (d) knowledge of content and students (KCS), (e) knowledge of content and curriculum (KCC), and (f) knowledge of content and teaching (KCT). CCK, HCK, and SCK fall under subject matter knowledge, and KCS, KCC, and KCT fall under pedagogical content knowledge.

Self-efficacy plays a significant role in guiding one's motivation and engagement in specific behaviors described by Bandura [34]. Specifically, those with high levels of mathematics self-efficacy are more inclined to (a) create challenging goals, (b) view challenges as learning opportunities, (c) increase the effort required to master goals, and (d) associate failure with insufficient effort [28]. Thus, mathematics self-efficacy's connection to overcoming educational obstacles [19] likely explains why high levels of mathematics self-efficacy correlate with higher mathematics achievement [28].

Bandura's [34] triadic reciprocal determinism theory, coupled with his self-efficacy theory, demonstrates how environmental factors, personal factors, and behavioral factors interact to influence an individual's life and was used to construct the themes in the self-efficacy data. The interactive homework assignments (environmental factors), collaboration (a behavioral factor), and mathematics knowledge for teaching (a personal factor) appeared to mutually reinforce each other to improve parent mathematics self-efficacy.

The interactive homework assignments (Appendix A) consisted of four components: the problem-solving section, the discussion section, the writing section, and the parent feedback section. The first three sections (e.g., problem solving, discussion, and writing) served to promote parent mathematics knowledge for teaching and parent mathematics self-efficacy for teaching. The feedback section promoted parent mathematics self-efficacy for teaching by allowing the parent participant to direct the content of the next session to further support their learning of the strategies. Each section was created for the express purpose of promoting dialogue and joint problem solving between parent and child. These assignments were designed as a template for parents to guide instruction without the researcher's regular input.

The problem-solving section contained (a) a word problem or prompt, (b) side-by-side examples of strategies, (c) a list of problems to solve using the strategies, and (d) two areas designated for the parent and student to problem solve using their strategies. Each participant had to solve at least two problems, and this section's three probing questions were used to help the parent and child observe the strategies in greater detail in preparation for deciding which strategies would be more suitable for solving specific problems (Figure 1). Figure 1 contains side-by-side examples of two mathematics representations of subtraction: base-10 blocks and the traditional algorithm. Figure 1 demonstrates the process of regrouping with subtraction using base-10 blocks, which are physical objects arranged in singular pieces and in groups of 10, 100, and 1000 to represent the base-10 numerical system. This section provided participants with images of the strategies employed, demonstrated connections between those strategies and word problems (and prompts), and highlighted how different strategies could be employed to solve problems. Choosing a problem and defending their reasoning for strategy choice served to boost interactions, and the separate writing sections designated for the parent and child participants also served to promote their interactions through joint problem solving on the same paper.

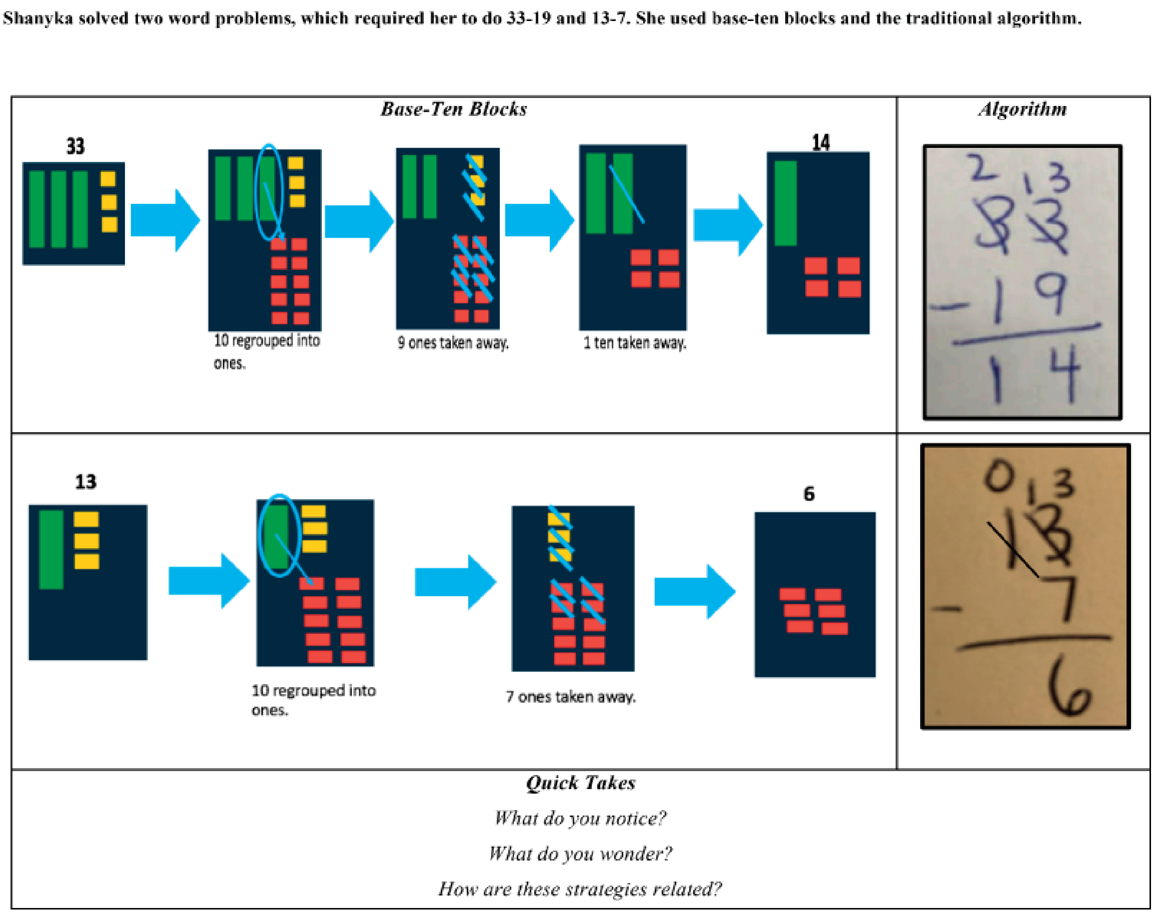


Figure 1. Problem-solving section: Side-by-side examples of subtraction.

The discussion and writing sections encouraged participants to reflect on and discuss their engagement with the problems to deepen their mathematics knowledge about the strategies and to reflect on the effectiveness of their partnering. The discussion section contained two questions that the participants answered aloud to gain more insight into the relationship between their strategies and problems to determine which strategies were more effective for specific problems (Table 1).

Table 1. Discussion section questions.

Question 1	Compare the strategies you used. Which strategy worked best for solving each problem? Please defend your response.
Question 2	How did your partner help you learn about the strategies (e.g., encouragement, providing a great explanation)?

Additionally, the writing section contained questions (two to three) that encouraged the adult participant to compare strategies and form predictions about which strategies work best with different problems, strengthening their understanding of conceptual strategies' purposes and advantages (Table 2). The number of questions would vary depending on whether the homework assignment contained a word problem or a prompt (Appendix A). If a homework assignment contained a prompt, Questions 1 and 2, were used. If a homework assignment contained a word problem Questions 1, 2, and 3 were used. The parent participant recorded her thinking in this section.

Table 2. Writing section questions.

Question 1	What were the advantages of each strategy?
Question 2	Can you predict which strategies would work best with these problems by inspection? Why or why not? Please explain.
Question 3	How did the strategies reflect the word problem?

The feedback section also contained three statements in which the adult participant could describe student progress on the assignment. This section also encouraged the adult to provide recommendations for future assignments. Goldman and Booker's [10] and Jackson and Remillard's [11] studies aligned with this study's needs assessment responses, which indicated that parents wanted more instructional resources on these strategies. Specifically, parent participants from the needs assessment desired supports that were prescriptive, such as textbook resources containing examples of how to use the strategies. The feedback section was created to empower the parent participant to request additional support on strategies to further aid their understanding of conceptual mathematics strategies. The researcher, also a fourth-grade teacher, was familiar with the strategies that the child participant had learned. Moreover, the researcher's observations of parent–child interactions guided her construction of subsequent homework activities as she inserted and changed the pictorial representations to support the parent's growing understanding of the strategies.

3. Results

This study was developed to examine the effectiveness of an interactive homework program for elementary mathematics students and their parents. Three research questions were designed to guide the study. Specifically, two research questions were designed to capture the intervention's impact on Linda's perceived mathematics self-efficacy and mathematics knowledge for teaching. The third research question involved her identifying helpful interactive homework program components. Many study results were unanticipated, and the emerging themes may prove helpful in guiding future studies regarding parent and child mathematics engagement.

3.1. Research Question One

Research Question One is "In what ways does the homework intervention change perceived parent participant mathematics self-efficacy?" Linda appeared to demonstrate greater confidence in teaching Laura Jean conceptual mathematics strategies, as indicated by the survey, interactive homework assignments, observations of the homework sessions, journal entries, and the interview, which yielded four themes: beliefs, autonomy, modification, and motivation. Beliefs is a theme in this study, referring to the parent participant's reactions to new teaching methods and how her beliefs evolved throughout the intervention. Autonomy represents the participants' capacity to direct the problem-solving process and collaboratively problem solve. Modification represents how the interactive homework sheets were changed based on participant interactions during the problem-solving process and parent requests. Motivation is the participants' willingness to engage in pictorial mathematics strategies and extend their learning outside the interactive homework sessions.

With respect to the theme of beliefs, Linda was initially frustrated by conceptual strategies (Interactive Homework One; Interactive Homework Two; Sessions One through Three; Survey). During the first two observed sessions, Linda's frustration appeared to stem from how she struggled to help Laura Jean realize success in mathematics when she had minored in mathematics (Survey and Session Two) and had extensive mathematics experiences with her own parents (Session Three). Linda shared, "Yeah, we [family] get pretty confused. My dad's an engineer, my mom's an accountant. So. . . numbers, we know them" (Session Three). Linda also believed that conceptual strategies were more complicated than necessary, as indicated by her response to the use of number lines, a strategy used to demonstrate the difference between whole numbers serving as endpoints:

> I don't like number lines. I really. . . I'm not asking you to make me do this again. But some of them seem, some problems are way harder to do with the number line because it's much more complicated math. And I feel like they should just leave that behind now [and] move on. (Session Three)

Upon debriefing with Linda at the end of the first observed session featuring Interactive Homework Two, her disapproval of conceptual strategies appeared to be another source of concern, evidenced by her discussion of adjustment, a compensation strategy, in which the minuend, the top number, and subtrahend, the second number, in a subtraction problem, are modified by an equal amount. Figure 2, from the first interactive homework assignment, shows both reduced by 1. Linda expressed relief upon reviewing Interactive Homework Two, which did not contain a side-by-side example of the adjustment strategy, "Yeah, I don't think I like adjustment. When I saw that. . . she [Laura Jean] started doing that, I'm like, I just. . .don't want to change the number. I want to work [with] the number I have" (Session One). Her dislike of the conceptual strategies also led to her resistance to using them. For example, as referenced earlier, Linda's comment regarding number lines, "I don't like number lines. . . I'm not asking you to make me do this again" (Interactive Homework Four; Session Three). Because she did not feel comfortable using the conceptual strategies, she expressed satisfaction when she checked her answers using the traditional algorithm, "And that's when you can use the algorithm and check against like. . .Oh, look at it" (Session Three).

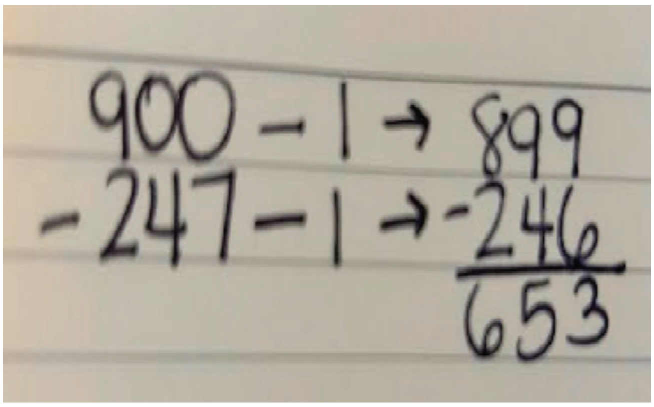


Figure 2. The adjustment strategy from the first assignment.

By the fourth session, Linda's concerns about the conceptual strategies began to rapidly subside as she embraced conceptual strategies to help Laura Jean understand fractions, concepts Laura Jean struggled with more than whole numbers. As Laura Jean began to arrive at the correct answers and grasp adding fractions using number lines and area models, a region that is partitioned into equal areas, Linda praised the number line strategy for adding fractions, "Hey, Laura Jean. That makes sense on the number line" (Session Four). Linda also appeared to value how the area model helped Laura Jean visualize adding fractions, "Like having this the area model of your picture. She could do that and see what it is" (Session Four).

Linda began to demonstrate greater autonomy as the sessions continued (Researcher's Journal, p. 4), and she became hopeful for what the assignments would bring, "Hopefully it'll help me. . . make me more confident with my younger one [her youngest daughter] when she's doing this stuff [conceptual strategies]" (Session Four). In the first session, Laura Jean selected the two problems (500-345, 400-289) that would have been appropriate for solving with the adjustment strategy. As a result, she left Linda with the remaining problems that were too difficult to solve with this strategy. As Linda conveyed in the assignment's writing section, she could not complete her last problem. This incident reflected her limited authority in redirecting Laura Jean's selection of problems as she did in later sessions, three through five, when Linda became more confident using conceptual strategies.

By the fourth observed session, when Laura Jean struggled to generate a number line for adding $\frac{3}{10}$ and $\frac{5}{10}$, Linda skillfully directed her to the interactive homework's pictorial representations and modeled creating a number line for the new problem. Linda's reliance on the conceptual models to rectify Laura Jean's confusion was apparent in her (a) highlighting features of the conceptual strategies; (b) creating her own model as a demonstration; and (c) asking follow-up questions to deepen and assess Laura Jean's understanding of computing with fractions: "Okay. So, you know that one of these. We only need one for the tenths, and we cut [it] into 10 pieces. Three of the 10 pieces, right? Okay, but where [what] does this one represent? Do you still need these?" Linda's techniques were based firmly on pictorial representations (Figure 3), demonstrating greater confidence in using conceptual strategies (Researcher's Journal, p. 5)

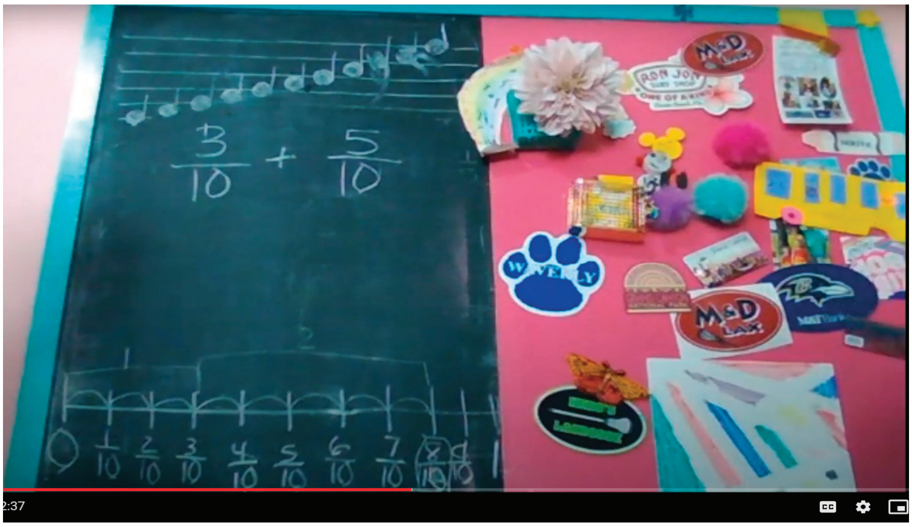


Figure 3. Video screenshot of parent work during session four using interactive homework five.

Each interactive homework assignment underwent modifications based on researcher observations and parent feedback on the assignments, effectively conforming the assignments to participant need to improve their confidence using the conceptual strategies. Linda was disincentivized from using conceptual strategies, believing that they were less useful after Laura Jean selected all the problems for solving that were appropriate for the adjustment strategy on the first assignment, leaving Linda with problems that were only suitable for solving with the traditional algorithm. Thus, the directions on the second interactive homework were modified to ensure that participants alternated turns when choosing problems to facilitate discussion and improve their opportunities for selecting strategies that correspond to the problems. The second assignment was also modified to include three strategies instead of two, as in Interactive Homework Assignment One, to facilitate Linda's understanding and use of conceptual strategies (Figures 4 and 5). Figure 4 shows side-by-side examples of the adjustment strategy (discussed above) and the traditional subtraction algorithm. Figure 5 illustrates three strategies. Removal is a number line strategy in which the total quantity is positioned at the right endpoint of the number line and the second number is removed through repeated subtraction in segments; the answer is the most left or last point on the number line. The partial difference strategy involves subtracting the minuend incrementally, often by place value. At the end of Session Three, Linda requested help with teaching fractions. In response, Interactive Homework Assignments Five and Six were modified to include ways to conceptually develop an understanding of fractions and fraction operations. Her request reflects improved confidence in using conceptual strategies to support Laura Jean's understanding of the subtraction of whole numbers (Researcher's Journal, p. 4).

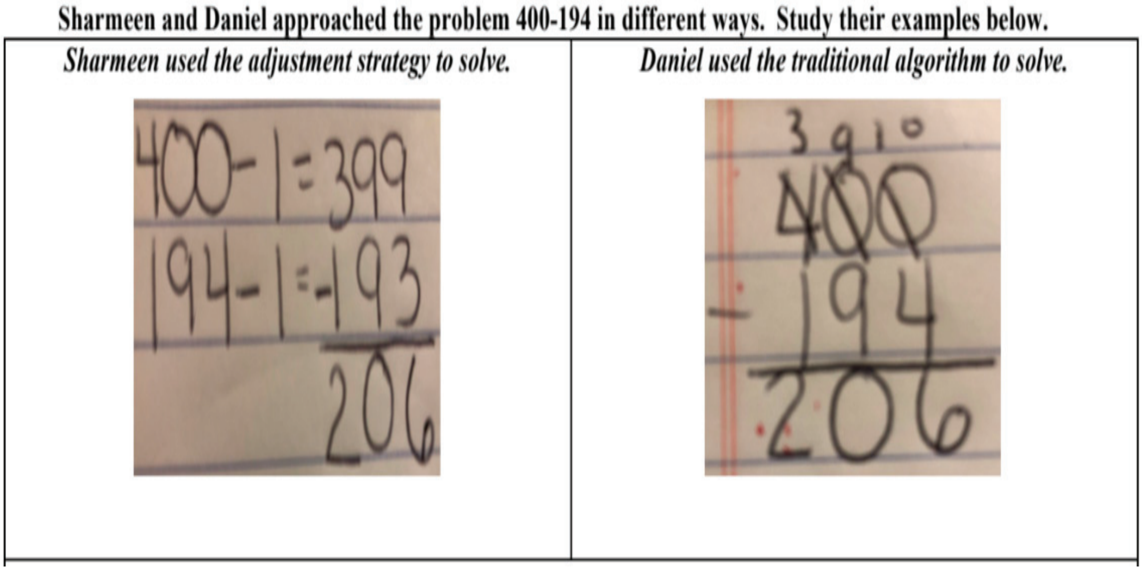


Figure 4. Interactive homework assignment one.

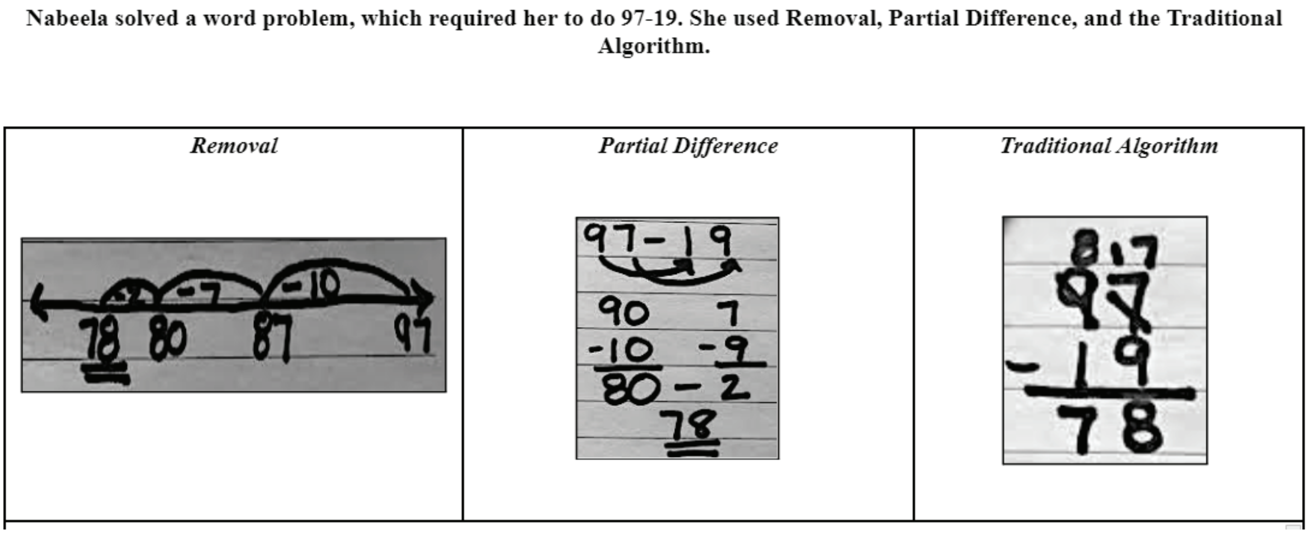


Figure 5. Interactive homework assignment two.

The interactive homework assignments contributed to Laura Jean's motivation to engage in mathematics activities. According to Linda, Laura Jean began to enjoy mathematics because she could practice and master concepts for which she had previously struggled (Interview; Sessions Three through Five). The assignments were tailored to her areas of development. As a result, Laura Jean experienced even greater success on more challenging problems, which also improved her motivation to participate in the intervention (Researcher's Journal, p. 4). Her enjoyment of the interactive homework assignments and mathematics was evident when she opted to solve most of the problems instead of sharing the responsibility equally with her mother (Researcher's Journal, pp. 3–5; Sessions Three through Five).

Linda credited the interactive homework's collaborative nature to Laura Jean's newly found interest in mathematics, which in turn motivated Linda to find additional problems for Laura Jean to practice:

> But it was helpful for me to work with Laura Jean and for her to have somebody else. . . we have the assignment to go along, but we had to do [it] together and she actually enjoyed it. So, I think seeing her getting interested to study math, I think it helped, and gave us a reason to do more math. And we just said, hey, let's just do some extra, like homework or sheets I found online. And so, it kind of gave us [a] purpose to do more math and work on some of the stuff that she's been working on. (Interview)

3.2. Research Question Two

Research Question Two is "In what ways does the homework intervention change perceived parent participant mathematics knowledge for teaching?" An examination of participant comments on the survey and assignments, observations, researcher's journal entries, and the interview indicated Linda's greater proficiency in identifying and using conceptual strategies to support Laura Jean's mathematics learning. For the second research question, understanding and functionality were identified themes that characterized improved parent mathematics knowledge for teaching. Understanding was defined as the participant's growing knowledge of the pictorial representations or conceptual mathematics strategies and how and when to use them, and functionality was the parent participant's growing understanding of the purposes of each set of strategies.

Linda's limited understanding of conceptual mathematics strategies contributed to her challenges with helping Laura Jean. Due to these past struggles, she revealed in her interview that she joined the study to learn more about them. Because Linda did not know the conceptual strategies for computing with fractions, helping her daughter with fractions was particularly difficult:

> . . . I don't actually remember how to, like, do them, like subtract fractions and everything. . . . I can visually and I know what the number is like, and I could do it. But I'm like actually showing her. . . the traditional algorithm kind of way. I'm like, I don't know that I know it [a conceptual strategy]. (Session Three)

Linda's unfamiliarity with these mathematics concepts also influenced her ability to identify connections between conceptual strategies and the traditional algorithm, "I don't know. Do you see? . . . The algorithm next to the example . . . looks totally different to me. Like it doesn't. I don't see [the] relation" (Session Three). Yet, as the sessions progressed, Linda used the support from the side-by-side examples of how to solve the problem to connect her mathematics background to what Laura Jean was learning in school. In her interview she elaborated:

> So, I guess it [the side-by-side example of the conceptual strategy] just informed me of what it is and kind of showed me compared to. . . how I learned in school. [It] kind of gave me a comparison so I can see what. . . I was doing compared to what I needed to be doing to show her [Laura Jean].

Learning more about the conceptual strategies led to Linda developing more positive beliefs about the use of conceptual strategies. She initially expressed dissatisfaction with the adjustment strategy (Interactive Homework One; Researcher's Journal, pp. 1, 3; Session One), but by the third session, she conveyed that the addition of the number line on the fourth interactive homework assignment helped her learn the adjustment strategy, "Adjustment with the number line helped [me] to see how adjustment work[s]." Linda's understanding improved with the combination of two conceptual strategies she initially disliked and did not understand. Linda's greater familiarity with conceptual strategies and resulting improvements in her mathematics knowledge for teaching were illustrated when she relied on pictorial representations to help her daughter. For example, in the third session on Interactive Homework Four, Linda used the side-by-side examples of the base-10 model and traditional algorithm for subtraction to help Laura Jean understand regrouping:

> . . .when you look at just this number like that, then you don't have to borrow. Right? Right. So, if you did traditional, you'd have to like this, right? Cross that off, make that a seven, this would become 10. Right? Yes.

By the fourth session, Linda's improved mathematics knowledge for teaching was illustrated by her complete reliance on or full use of pictorial representations. She was then asking questions of Laura Jean such as," How does your picture represent three times? What part of your picture? Can you shade it?" These comments and resulting behaviors, reflective of Sessions Four and Five, directly contrast with her dependency on the traditional algorithm in Sessions One, Two, and Three.

Linda's understanding and mathematics knowledge for teaching were also demonstrated in her correct assessment of Laura Jean's abilities. By the end of the third session, which featured the fourth assignment, Linda desired a change in content because her daughter felt comfortable subtracting whole numbers in various ways, "Like she's comfortable with this stuff. Now she fully understands it. But. . . She's already [on] fractions [in class]." Her assessment matched that of the researcher, who had eight years of instructional experience. Moreover, although the intervention was initially created to develop a conceptual understanding of whole number operations, Linda requested practice computing with fractions once she observed that Laura Jean mastered subtracting whole numbers. The flexibility of the intervention was partially based on Linda's input, allowing the researcher to use the intervention to target additional areas of need. Thus, the fifth and sixth assignments were adjusted to include fractions (Researcher's Journal, pp. 5–6).

Linda's improved understanding of conceptual strategies was also indicated by how she recognized these strategies in other parts of her daughters' learning. She identified Interactive Homework Five's models in Laura Jean's DreamBox program, an online mathematics resource that teachers assigned as homework and classwork to reinforce mathematics learning. Linda also used this connection to help Laura Jean solve problems, ". . .these fractions are kind of like your square units that you were doing [on] that DreamBox. Remember how you're doing it yourself?" (Session Five). She also recognized the homework assignment's number line strategies in her younger daughter's class work (Session Three).

Initially, Linda's limited exposure to and understanding of conceptual strategies influenced her beliefs about them, as indicated by her concerns about their functionality, the second theme of Research Question Two. For example, when Laura Jean selected all the problems that corresponded to the adjustment strategy on the first assignment, Linda wrote, "Guess there are less steps in adjustment, but it can't be used in every problem and requires extra thinking." This observation demonstrated how she defined functionality by a strategy's versatility. Linda also defined a strategy's utility by the amount of effort required to use it, as indicated by her comments on the first assignment, which acknowledged advantages of the traditional algorithm for subtraction, "Traditional [algorithm] = straight forward and no extra thinking if it could work." The concept of the traditional algorithm's efficiency was reiterated in the third session, in which she discussed how the conceptual strategies "Force you to think about solving," thereby reducing their efficiency in the process.

The assignments were adjusted weekly to help Linda redefine functionality and change her mindset and develop confidence using conceptual mathematics strategies (Researcher's Journal, pp. 2–6). As a result, Linda began to understand that conceptual strategies promote understanding instead of achieving an immediate answer, the traditional algorithm's purpose (Interactive Homework Assignments Three through Six; Sessions Three through Five). For example, before the intervention, Laura Jean spent at least two months learning the traditional algorithm for subtracting whole numbers, and she developed the habit of reversing minuends and subtrahends. After completing Interactive Homework Assignments Three and Four, which featured the base-10 model for subtraction, Linda observed the visual representation's effectiveness when Laura Jean stopped reversing minuends and subtrahends for problems that required regrouping. This change in performance illustrates how visual representations remedied Laura Jean's misunderstandings, transitioning her to the correct use of the traditional algorithm (Interactive Homework Four; Session Three; Researcher's Journal, p. 4). Witnessing how learning conceptual strategies improved Laura Jean's understanding of the traditional algorithm was an eye-opener for Linda, "Yeah, I mean, I think the base-10 blocks actually make the most sense to me visually" (Session Three).

3.3. Research Question Three

Research Question Three is "What components of the interactive homework assignment program do parent participants identify as useful in helping them support their children with mathematics learning at home?" The themes of repetition, convenience, and side-by-side examples were identified as effective intervention components for developing mathematics knowledge. Repetition was defined as the participants' opportunities for repeated exposure to and practice with mathematics concepts. The side-by-side examples are about the diagrams of the pictorial strategies located in the interactive homework worksheets, and convenience is defined by the parent participant's noted ease of using the interactive homework assignments.

Linda discussed the importance of repetition, requiring additional practice on mathematics concepts. Laura Jean struggled to recall what she learned in school because she took notes on white boards and rarely had homework in previous grades (Session Two; Session Five). Linda appreciated the additional practice she received with the homework assignments, "I always wanted more stuff from the school. I'm just trying to help [with] the math because I didn't get it. And I took this opportunity as a chance to do that" (Session Four). Linda discussed how these additional resources and practice at home helped Laura Jean master concepts, "So it was helpful for her to have some extra math that wasn't too hard to kind of talk about and stuff to see" (Interview).

Repetition and Laura Jean's subsequent successes in mathematics led to her enjoyment of mathematics (Interview; Sessions Three through Five). This change in Laura Jean's mindset inspired Linda to provide more opportunities for mathematics engagement:

> But it was helpful for me to work with Laura Jean and for her to have somebody else kind of, I mean, we have the assignment to go along, but we had to do together and she actually enjoyed it. So, I think seeing her getting interested in studying math, I think it helped, and gave us a reason to do more math. And we just said, hey, let's just do some extra, like homework or sheets I found online. And so, it kind of gave us purpose to do more math and work on some of the stuff that she's been working on". (Interview)

Laura Jean's enjoyment of the interactive homework assignments and mathematics were also illustrated when she opted to do most of the assignments' problems instead of sharing the responsibility with Linda (Researcher's Journal, pp. 3–5; Sessions Two through Five). Linda appreciated how the side-by-side examples of the interactive homework assignments helped her understand the conceptual strategies. Linda was unfamiliar with these strategies and discussed the importance of using these examples to support Laura Jean:

> And so, I think having that sample like above and explaining the different forms to do it helps at least a parent that knows some bit about math and doesn't know the math isn't, you know, going to figure it out unless they're learning along with their kid. But yeah, I think the hardest thing was always that she was expected to do these different models. And I didn't know what they were. (Session Four)

Linda discussed how she did not receive sufficient resources from Laura Jean's teachers and how she appreciated how the side-by-side examples helped her compare conceptual strategies and the traditional algorithm to learn conceptual strategies:

> So, I didn't have any of the side-by-side stuff to show the different ways in what they're supposed to be doing and how they were learning and stuff because none of it ever came . . . home where she was explaining, I don't know what she was saying. So, so I guess [the examples] just informed me of what it is and kind of showed me compared to what how I learned in school kind of gave me a comparison so I can see what I was, what I was doing compared to what I needed to be doing to show her. (Interview)

The side-by-side strategies, serving as informative guides during remote instruction when access to teachers was more limited, and the resulting available problems informed Linda's understanding of efficiency. She emphasized the importance of these samples, "So I think having a sample of what the kids are supposed to be working on, I think, helps" (Session Five).

Linda praised the convenience of the interactive homework assignments, stating how prior inconvenient resources and tools detracted from Laura Jean's learning. She specifically cited the challenges of remote instruction that required students to complete assignments on the computer and Laura Jean's difficulties writing on the touch screen of her school-district-provided Chromebook. Her screen was too small to write on and a stylus had not been provided:

> It's like if the screen is bigger, it would be easy to work on but their computer screens are [too small and] it's like this big pain, and their little fingers are fat. And I guess if you had like little, like the pencils that write on screens. . .Yeah, that would make it easier. And I don't know if they even work with these things. But um, but yeah, it makes it really hard to write on there and then to erase and then you need to go back and take. Yeah, it's just a pain. (Session Three)

Linda believed that the typing feature of Pear Deck, an interactive app for student learning, posed additional challenges for setting up problems for solving:

> And if you just type and use the type part. . .trying to get it to like [to] type the problem and then line it up and hit their space and that ends up, it doesn't line up right. (Session Three)

She continued discussing challenges with Chromebooks in the fourth session, when she experienced difficulty locating Laura Jean's school assignments to print out (Researcher's Journal, p. 4). Given Linda's difficulties with technology, she was grateful for the easy access to the interactive homework assignments and the ability to write directly on them for problem solving (Session Three).

4. Conclusions

In alignment with existing literature, this study's findings revealed a need for providing parents with sufficient mathematics support to aid their child's learning [10–12]. Although Linda had an extensive mathematics content background, she did not have extensive pedagogical content knowledge. Therefore, she initially struggled to help her daughter learn mathematics concepts and strategies for which she had limited experience. Linda began to demonstrate improvements in mathematics self-efficacy in her ready use of conceptual strategies to support Laura Jean. Improvements in parent mathematics self-efficacy appear to stem from greater exposure to and practice with conceptual mathematics strategies and modifications to the intervention to suit participant needs [23,24,28,35].

Specifically, participants solved problems together using various strategies, and their understanding of these strategies improved through repeated practice and the use of question prompts and examples to evaluate and reflect on the strategies' effectiveness. As a result of practice and evaluation, their arsenal of strategies improved, leading to a deeper understanding of the nature of subtraction, addition, and the part-to-whole relationship of fractions. Moreover, participant learning and self-efficacy also progressed due to conditions that were controlled (e.g., the nature of parent–child discussion and opportunities for selecting different strategies) through weekly homework modifications based on participant feedback and need. Problem-solving tasks, collaboratively problem solving, direct instruction on instructional methods, self-guiding instruction, and emotional supports (e.g., a discussion question that asked participants to highlight each other's contributions and prompts that promoted positive interactions among participants) were contributing factors to improved parent mathematics self-efficacy and parent mathematics knowledge for teaching.

A bilateral relationship between the parent–child partnership in their understanding of conceptual strategies and confidence using the conceptual strategies was postulated to affect the intervention's outcomes. The improvement in Laura Jean's learning was remarkable. Although Linda was the participant of primary focus of this study, the interactions between Linda and Laura Jean propelled their progress during the intervention. Specifically, as Linda's mathematics self-efficacy and mathematics knowledge for teaching improved, Laura Jean's mathematics skills also improved, resulting in Linda's improved approach to her mathematics engagement with Laura Jean.

Additional themes that emerged from the intervention were initiative and ability. Initiative is defined as one's willingness to take charge of the learning process, and ability refers to levels of mathematics proficiency. Laura Jean's transformation was signaled by her willingness to take the initiative in solving problems in the homework. Session One was characterized by Laura Jean following Linda's directions on completing the problems (Researcher's Journal, p. 2). As Linda repeated the strategies' procedures, Laura Jean quietly obeyed. As Linda attempted to engage Laura Jean, she initially responded with one to three words, grunts, and shrugs. Moreover, Linda did not answer the second homework's discussion section question, "How did your partner help you learn about the strategies (e.g., encouragement, providing a great explanation)?", reinforcing the researcher's observations of how Laura Jean had not initially facilitated Linda's learning (Researcher's Journal, p. 2; Session One).

Their improved interactions mutually reinforced each other's growth, leading to Laura Jean's improved initiative to engage in mathematics tasks by the last session. At the beginning of the second session, Laura Jean maintained her passivity; however, after the base-10 strategy was explained, she volunteered to complete the rest of the interactive homework problems (Researcher's Journal, p. 3). By Session Three, as Linda began to follow the discussion prompts to examine the side-by-side examples more closely, facilitate productive dialogue, and encourage Laura Jean, Laura Jean volunteered to do most of the homework problems and complete problems outside of the intervention and schoolwork (Interview; Researcher's Journal, pp. 4–6; Sessions Three through Five). Linda's guidance and encouragement, facilitated by the assignment's prompts and questions, improved Laura Jean's understanding. As a result, she experienced success in mathematics, leading to more positive reinforcement and, thus, a greater desire to engage in mathematics. By the last session, Laura Jean created models for solving and articulated her problem-solving process using mathematical language without Linda's prompting. (Researcher's Journal, p. 6; Session Five).

Laura Jean's improved ability and rapid mathematics success appeared to be the linchpins that ignited improvements in Linda's mathematics knowledge for teaching and mathematics self-efficacy as she observed the effectiveness of conceptual strategies. Before the intervention, Laura Jean struggled for at least two months using the traditional standard algorithm for subtraction with regrouping. After two sessions using the base-10 block strategy led to her accurate use of the traditional algorithm, Linda requested additional practice with conceptual strategies to improve Laura Jean's understanding of fractions (Researcher's Journal, p. 4; Session Three). By Session Four, Linda guided Laura Jean's understanding of fractions through questioning and pictorial representations of the part-to-whole relationship. Linda had to correct Laura Jean's area models for subtracting fractions less than one; however, by Session Five, Laura Jean adroitly created number lines with equivalent fractions greater than one to regroup with subtraction while correcting her mistakes and explaining her process for solving. Laura Jean's self-corrections, articulation of her problem-solving process, and her adaptations of a conceptual strategy without prompting (e.g., placing equivalent fractions on the same number line to subtract) with more advanced concepts like fractions represent a marked departure from her mathematics engagement in the initial sessions. The parent participant's knowledge and confidence in using conceptual development strategies translated to the child participant's full embrace of mathematics as she began to experience success. As a result, the parent participant, initially resistant to using conceptual strategies, actively sought additional opportunities to work with her child to use conceptual strategies. Additional practice led to greater improvement in mathematics knowledge and confidence. Each participant fueled the other's progress, likely resulting in the relatively rapid growth in their overall conceptual knowledge and self-efficacy. The dynamic between parent and child collaborations cannot be underestimated as a factor in future mathematics interventions.

Implications for the Future

While the results of the intervention appear promising, there are many questions that future studies should examine. The parent participant had a strong mathematics background that she used to catapult her understanding of conceptual mathematics strategies. Furthermore, she was confident in her general mathematics abilities. How effective would the interactive homework assignments be for parent participants with less mathematics experience and confidence? These participants were also affluent, and what would the impact of this intervention be with participants from various SES or a much more diverse set of participants in general? This intervention involved two participants, but what could be its impact on a larger sample size of participants? Given the small sample size, the assignments could be tailored based on specific needs. How could this intervention be scaled up to meet the varying needs of several more participants?

As schools in the United States continue to grapple with the aftermath of the COVID pandemic's effects on student learning, new approaches must be taken to improve parent mathematics engagement. This study highlighted the impact of a parent participant, who, when appropriately supported, developed a new mindset and skills for helping her child overcome significant mathematics challenges and gaps in understanding. This study also revealed the importance of respecting parent perspectives; how valuing their needs, skills, and insight can work to accelerate their child's progress. Schools should provide resources to families that help them learn new mathematics concepts and approaches to teaching mathematics different from ways that they may have learned. Specifically, these resources should be embedded with question prompts that promote positive mathematics engagement between family member and child, as the study revealed that guided collaboration enhanced the participant understanding of the strategies and their motivation to continue learning. Moreover, schools should focus on gathering feedback from families about the resources they need to support their children. As the participants grew in their knowledge of pictorial mathematics strategies, the parent participant better advocated for additional materials to meet her daughter's needs. Improving the bridge between home and school through interactive tools that mirror classroom instruction and communication could be instrumental in improving this nation's mathematics trajectory for the future.

Author Contributions: L.M. designed the study, carried out the study with the participants, organized and analyzed the data, and prepared the initial manuscript. R.N.R. oversaw the development and

implementation of the study, assisted with the data analysis and the preparation of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by Homewood Institutional Review Board (protocol code HIRB00012604, date of approval: 20 October 2020).

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

Data Availability Statement: Data will be stored in a secure facility for seven years.

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

Appendix A

Interactive Homework Examples.

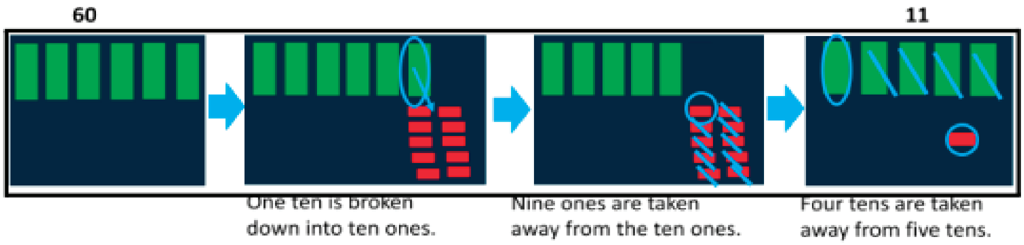
Interactive Homework Four

Understanding which strategies are the most efficient for solving is the focus of this assignment. Knowing **when** and **why** certain scenarios are more efficient is a crucial component of problem solving. Examples of the *removal, partial difference*, and *algorithm strategies* are below.

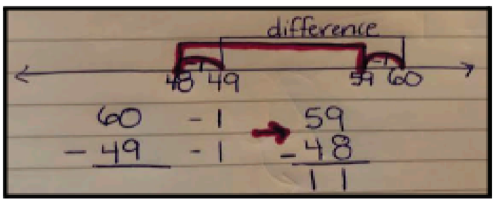
In this assignment, we explore different approaches to subtraction to strengthen your child's understanding of numbers and develop a deeper understanding of the operation of subtraction.

Katherine solved a word problem, which required her to solve 60-49. She used base-ten blocks, removal on a number line, and adjustment on a number line.

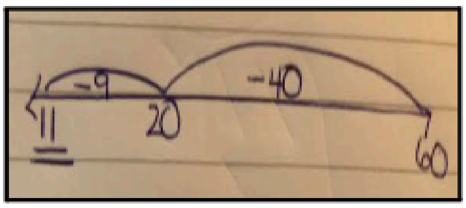
Base-Ten Blocks



Adjustment



Removal



Quick Takes

What do you notice?

What do you wonder?

How are these strategies related?

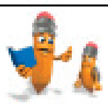
You have been presented with the following problems:

95-38, 47-39, 80-59, 72-25, 90-45

1. Let's talk about it! Which strategy would you use on each of these problems and why?
2. Next, student, select one problem from the list! Solve using a strategy from above.
3. Family Member, it is your turn! Select a different problem. Solve using a strategy from above.
4. Next, student, select a second problem from the list! Solve using a different strategy from above.
5. Family Member, it is your turn! Select a different problem. Solve using a different strategy that you have not used from above

Student Work	**Family Member Work**
Strategy 1 *Choose two different problems and solve using two different strategies from above.* *Strategy 2*	*Strategy 1* *Choose two different problems and solve using two different strategies from above.* - *Strategy 2*

Discussion Section



Let's talk about the strategies again! Refer to the different strategies you used to solve these problems.

1. Compare the strategies you used. Which strategy worked best for solving each problem? Please defend your response.
2. How did your partner help you learn about the strategies (e.g., encouragement, providing a great explanation)?

Family Member (Parent) Writing Section

Please respond in writing to the questions below.

1. What were the advantages of each strategy?
2. Can you predict which strategies would work best with these problems by inspection? Why or why not? Please explain.

Family Feedback Section

Dear Family Partner,

Please share your reactions to your student's work on this activity. Please check the statements that are true for you.

_____My student understood the homework.

_____My student was able to complete the homework.

_____I would like more information about how I can support learning out of school.

Interactive Homework Five

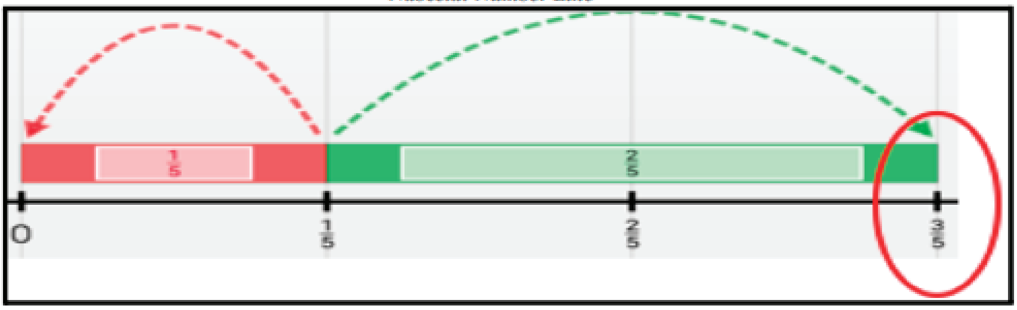
Understanding which strategies are the most efficient for solving is the focus of this assignment. Knowing **when** and **why** certain scenarios are more efficient is a crucial component of problem solving. Examples of the *removal, partial difference*, and *algorithm strategies* are below.

In this assignment, we explore different approaches to subtraction to strengthen your child's understanding of numbers and develop a deeper understanding of the operation of subtraction.

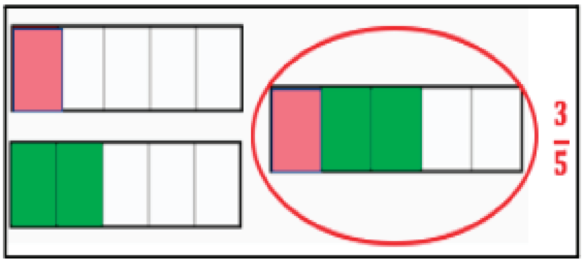
Nabeela, Shanyka, and Katherine reviewed and solved the word problem below using different strategies:

Raro ran 1/5 of a mile. Cassie ran 2/5 of a mile. How far did they run?

Nabeela: Number Line



Shanyka: Area Model



Katherine: Traditional Algorithm

$$\frac{1}{5} + \frac{2}{5} = \frac{3}{5}$$

Quick Takes

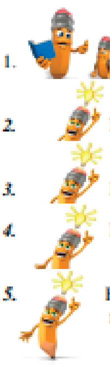
What do you notice?

What do you wonder?

How are these strategies related?

You have been presented with the following problems:

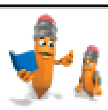
$\frac{3}{10} + \frac{5}{10} =$, $\frac{3}{7} + \frac{2}{7} =$, $\frac{1}{4} + \frac{2}{4} =$, $\frac{3}{12} + \frac{7}{12} =$, $\frac{5}{17} + \frac{9}{17}$



1. Let's talk about it! Which strategy would you use on each of these problems and why?
2. Next, student, select one problem from the list! Solve using a strategy from above.
3. Family Member, it is your turn! Select a different problem. Solve using a strategy from above.
4. Next, student, select a second problem from the list! Solve using a different strategy from above.
5. Family Member, it is your turn! Select a different problem. Solve using a different strategy that you have not used from above

Student Work	Family Member Work
Strategy 1 *Choose two different problems and solve using two different strategies from above.* *Strategy 2*	*Strategy 1* *Choose two different problems and solve using two different strategies from above.* *Strategy 2*

Discussion Section



Let's talk about the strategies again! Refer to the different strategies you used to solve these problems.

1. Compare the strategies you used. Which strategy worked best for solving each problem? Please defend your response.
2. How did your partner help you learn about the strategies (e.g., encouragement, providing a great explanation)?

Family Member (Parent) Writing Section

Please respond in writing to the questions below.

1. What were the advantages of each strategy?
2. Can you predict which strategies would work best with these problems by inspection? Why or why not? Please explain.
3. How did the strategies reflect the word problem?

Family Feedback Section

Dear Family Partner,

Please share your reactions to your student's work on this activity. Please check the statements that are true for you.

_____My student understood the homework.

_____My student was able to complete the homework.

_____I would like more information about how I can support learning out of school.

References

1. Gravemeijer, K.; Stephan, M.; Lin, F.; Julie, C.; Ohtani, M. What mathematics education may prepare students for the society of the future? *Int. J. Sci. Math. Educ.* **2017**, *15*, 105–123. [CrossRef]
2. Kang, H.; Cogan, L. The differential role of socioeconomic status in the relationship between curriculum-based mathematics and mathematics literacy: The link between TIMSS and PISA. *Int. J. Sci. Math. Educ.* **2022**, *20*, 133–148. [CrossRef]
3. Lawson, M.A. School-family relations in context: Parent and teacher perceptions of parent involvement. *Urban Educ.* **2003**, *38*, 77–133. [CrossRef]
4. National Council of Teachers of Mathematics. *Principles and Standards for School Mathematics*; National Council Teachers Mathematics: Reston, VA, USA, 2000.
5. Su, F. *Mathematics for Human Flourishing*; Yale University Press: New Haven, CT, USA, 2021.
6. United States Department of Education. Mathematics Equals Opportunity. 20 October 1997. Available online: https://files.eric.ed.gov/fulltext/ED415119.pdf (accessed on 20 June 2020).
7. Burrus, J.; Moore, R. The incremental validity of beliefs and attitudes for predicting mathematics achievement. *Learn. Individ. Differ.* **2016**, *50*, 246–251. [CrossRef]
8. Carnevale, A.P.; Desrochers, D.M. *The Democratization of Mathematics*; Semantic Scholar: Seattle, WA, USA, 2003.
9. National and Economic Importance of Improved Math-Science Education and H.R. 4272, the National Science Education Enhancement Act, 116th Cong. 2000. Available online: https://congressional-proquest-com.proxy1.library.jhu.edu/congressional/docview/t29.d30.hrg-2000-hew-0045?accountid=11752 (accessed on 14 November 2020).
10. Goldman, S.; Booker, A. Making math a definition of the situation: Families as sites for mathematical practices. *Anthropol. Educ. Q.* **2009**, *40*, 369–387. [CrossRef]
11. Jackson, K.; Remillard, J.T. Rethinking parent involvement: African American mothers construct their roles in the mathematics education of their children. *Sch. Community J.* **2005**, *15*, 51–73. Available online: http://www.adi.org/journalsearch/ (accessed on 5 December 2020).
12. Remillard, J.T.; Jackson, K. Old math, new math: Parents' experiences with standards-based reform. *Math. Think. Learn.* **2006**, *8*, 231–259. [CrossRef]
13. National Governors Association Center for Best Practices & Council of Chief State School Officers. *Common Core State Standards for Mathematics*; National Governors Association Center for Best Practices & Council of Chief State School Officers: Washington, DC, USA, 2010.
14. Silver, E.A.; Ghousseini, H.; Gosen, D.; Charalambous, C.; Strawhun, B. Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *J. Math. Behav.* **2005**, *24*, 287–301. [CrossRef]
15. Wei, R.C.; Darling-Hammond, L.; Andree, A.; Richardson, N.; Orphanos, S. *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the U.S. & Abroad (Technical Report)*; Learning Forward: Oxford, OH, USA; Stanford Center for Opportunity Policy in Education: Stanford, CA, USA, 2009.
16. Hatano, G.; Inagaki, K. Two courses of expertise. In *Child Development and Education in Japan*; Stevenson, H., Azuma, H., Hakuta, K., Eds.; Freeman: New York, NY, USA, 1986; pp. 262–272.
17. Givvin, K.B.; Stigler, J.W.; Thompson, B.J. What community college developmental mathematics students understand about mathematics, Part II: The interviews. *MathAMATYC Educ.* **2011**, *2*, 4–18.
18. Hackett, G. The role of mathematics self-efficacy in the choice of mathematics related majors of college women and men. A path analysis. *J. Couns. Psychol.* **1985**, *32*, 47–56. [CrossRef]
19. Jameson, M.M.; Fusco, B.R. Math anxiety, math self-concept, and math self-efficacy in adult learners compared to traditional undergraduate students. *Adult Educ. Q.* **2014**, *64*, 306–332. [CrossRef]
20. Berkowitz, T.; Schaeffer, M.W.; Maloney, E.A.; Peterson, L.; Gregor, C.; Levine, S.C.; Beilock, S.L. Math at home adds up to achievement in school. *Science* **2015**, *350*, 196–198. [CrossRef]
21. Berkowitz, T.; Schaeffer, M.W.; Rozek, C.S.; Maloney, E.A.; Levine, S.; Beilock, S.L. Response to comment on "Math at home adds up to achievement in school". *Science* **2016**, *351*, 1161. [CrossRef] [PubMed]
22. Schaeffer, M.W.; Rozek, C.S.; Berkowitz, T.; Levine, S.C.; Beilock, S.L. Disassociating the relation between parents with math anxiety and children's math achievement: Long-term effects of a math app intervention. *J. Exp. Psychol.* **2018**, *147*, 1782–1790. [CrossRef] [PubMed]
23. Jay, T.; Rose, J.; Simmons, B. Finding "mathematics": Parents questioning school-centered approaches to involvement in children's mathematics learning. *Sch. Community J.* **2017**, *27*, 201–230.
24. Tobon, G.; Hughes, T.M. Engaging Latino families as mathematicians. *Math. Teach.* **2020**, *113*, 201–208. Available online: https://www-jstor-org.proxy1.library.jhu.edu/stable/10.5951/mtlt.2019.0001?seq=1#metadata_info_tab_conconte (accessed on 4 December 2020). [CrossRef]
25. Mangram, C.; Metz, M.T.S. Partnering for improved parent mathematics engagement. *Sch. Community J.* **2018**, *28*, 273–294.
26. Papadopoulos, I. Opening inquiry mathematics to parents: Can they be engaged as teachers' partners in mathematical work? *J. Pedagog. Res.* **2017**, *1*, 1–20. Available online: https://search-proquest-com.proxy1.library.jhu.edu/docview/2265809190?accountid=11752 (accessed on 12 December 2020).

27. Mistretta, R.M. Conversations with family members about math. *Sch. Community J.* **2017**, *27*, 181–201. Available online: http://www.schoolcommunitynetwork.org/SCJ.aspx (accessed on 4 December 2020).
28. Williams, K.; Williams, H. Mathematics problem-solving homework as a conduit for parental involvement in learning. Evaluation of a pilot study. *Educ. Rev.* **2019**, *73*, 209–228. [CrossRef]
29. Panaoura, A. Parental involvement in developing students' perseverance in solving mathematical problems through the use of social media. *Int. J. Technol. Educ. Sci. (IJTES)* **2017**, *1*, 36–47. Available online: https://www.ijtes.net/index.php/ijtes/article/download/5/4 (accessed on 4 January 2020).
30. Westenskow, A.; Boyer-Thurgood, J.; Moyer-Packham, P.S. A window into mathematical support: How parents' perceptions change following observations of mathematics tutoring. *J. Res. Child. Educ.* **2015**, *29*, 258–475. [CrossRef]
31. Ball, D.L.; Thames, H.; Phelps, G. Content knowledge for teaching: What makes it special? *J. Teach. Educ.* **2008**, *59*, 389–407. [CrossRef]
32. De Corte, E.; Verschaffel, L.; Depaepe, F. Unraveling the relationship between students' mathematics-related beliefs and classroom culture. *Eur. Psychol.* **2008**, *13*, 24–36. [CrossRef]
33. Shulman, L.S. Those who understand: Knowledge growth in teaching. *Educ. Res.* **1986**, *15*, 4–14. [CrossRef]
34. Bandura, A. *Social Foundations of Thought and Action: A Social Cognitive Theory*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1986.
35. Van Voorhis, F.L. Costs and benefits of family involvement in homework. *J. Adv. Acad.* **2011**, *22*, 220–249. [CrossRef]

education sciences

MDPI

Article

The Impact of the COVID-19 Pandemic upon Mathematics Assessment in Higher Education

Eabhnat Ní Fhloinn [1,*] and Olivia Fitzmaurice [2]

[1] School of Mathematical Sciences, Dublin City University, D09 V209 Dublin, Ireland
[2] School of Education, University of Limerick, V94 T9PX Limerick, Ireland; olivia.fitzmaurice@ul.ie
* Correspondence: eabhnat.nifhloinn@dcu.ie

Abstract: Historically, the assessment of mathematics in higher education comprised closed-book, summative, proctored examinations. Related disciplines and subjects like statistics, mathematics education, and the history of mathematics lend themselves to a broader range of assessment techniques that have been reported to provide a more balanced picture of students' abilities. In 2020, an online environment for the teaching and learning of mathematics was imposed on the academic world globally as a result of the COVID-19 pandemic. In an effort to teach and assess remotely while maintaining institutional academic standards, the majority of lecturers were in a situation where closed-book, proctored assessments were not an option. As a result, other methods were adopted. This paper reports on an investigation into how mathematics lecturers worldwide assessed mathematics before the pandemic, during the initial lockdown restrictions, and in the immediate aftermath, while some restrictions were still in place, to see if any changes were sustained. There was a statistically significant difference in the proportion of respondents who used many of the assessment types investigated across the three time periods, including open-book timed, open-book untimed, closed-book, multiple-choice questions, online proctored, in-person proctored, presentations, projects, and assignments. The majority of those who favoured closed-book proctored examinations prior to the pandemic moved to timed open-book assessments. Differences between the weightings of final examinations versus continuous assessments were also statistically significant, with greater weight given to continuous assessment once the pandemic began. Respondents' satisfaction levels with their assessments were significantly different also, with the highest satisfaction levels prior to the pandemic and the lowest during the initial lockdown restrictions. Academic integrity was a key concern of the majority of respondents when assessing the learning outcomes of their modules and played a role in the vehicle of assessment they chose.

Keywords: COVID-19; mathematics assessment; emergency remote teaching; higher education mathematics; teaching mathematics online; technology in mathematics assessment

1. Introduction

The assessment of mathematics in higher education has long depended on closed-book, summative, proctored examinations (Iannone & Simpson, 2011; Davies et al., 2024). Arguments have been made to support the assertion that mathematics is in some way "different" as a discipline, in that it lends itself more to traditional forms of teaching and assessment, more so than might be the case in other subjects—but equally, more recently, such arguments have been refuted (Becher, 1994; Ní Fhloinn & Carr, 2017). Particularly since the early years of this millennium, research has been conducted into the possible modes of assessment into which mathematics could naturally expand, citing the fact that the traditional timed, closed-book assessment does not effectively assess skills such as problem-solving or the use of IT (Challis et al., 2003). It seems clear at least that mathematics assessment of this kind may fall short of the mark, as it may not give a fully comprehensive picture of a student's learning achievement (Burton & Haines, 1997). Although assessment of a broader range of skills can be achieved through approaches such as report writing, projects, and oral examinations (Niss, 1998), it appears that alternative modes of assessment are only more prevalent in modules of statistics, history of mathematics, mathematics education, and final-year projects (Iannone & Simpson, 2011).

During the initial university closures due to COVID-19 in 2020, mainly out of necessity, a broader range of formative and summative assessment methods were embraced by the mathematics teaching community (Fitzmaurice & Ní Fhloinn, 2021). This was a positive outcome of the move to remote teaching, as alternative assessment strategies assess a broader range of learning outcomes (Pegg, 2003). However, while this change happened during the first months of the pandemic, when lecturers had to pivot to online assessment with little or no time to plan, it is of interest to determine what happened the following year, when lecturers still often had to assess mathematics online but had significantly more notice of the fact.

Two decades ago, online assessment in mathematics started becoming more prevalent as the internet became more conducive to mathematics (Engelbrecht & Harding, 2004). The use of technology in education is seen as an inherent component of a teaching and learning environment that seeks to fulfil the diverse needs of students in the 21st century (Valdez & Maderal, 2021). Digital technologies offer compelling tools to conduct formative assessments effectively in mathematics (Barana et al., 2021). Online assessment, or E-assessment, comprises an extensive range of assessment types, including but not limited to online essays and computer-marked online examinations (James, 2016). Online examinations are an efficient means of conducting diagnostic, formative, and summative assessments and providing students with the opportunity to perform to the best of their ability (Valdez & Maderal, 2021).

Recent years have seen exponential growth in the different modes of online assessment that are available. Research by Davies et al. (2024), conducted in 2024, however, indicates that Computer-Aided Assessment (CAA) in tertiary mathematics remains underutilised despite these reported advancements in assessment methods. The continued over-reliance by university-level mathematics lecturers on closed-book written exams referred to above prompts questions about whether CAA can provide a more effective alternative, particularly in formative assessment. Some studies have documented that CAA has been shown to improve examination performance (Greenhow, 2015), although Greenhow (Greenhow, 2015, 2019) recommends that it should complement rather than replace traditional mathematics assessments. When used effectively, online formative assessment has the potential to nurture a learner- and assessment-centred focus using formative feedback and enhanced learner engagement with worthwhile learning experiences (Gikandi et al., 2011). However, it should be noted that distance assessment is not always experienced as a positive for students, as they can struggle with access to technology and resources or simply with feelings of isolation (Kerka & Wonacott, 2000). Multiple-choice questions have also been shown to have the potential for bias in relation to students with varying learning styles or in relation to confidence levels (Sangwin, 2013). Specific training or knowledge of question creation is also needed, as questions assessed via computer-aided assessment are different from those graded by hand (Greenhow, 2015).

Iannone and Simpson's more recent investigation into summative assessments in the UK (Iannone & Simpson, 2022) found that closed-book examinations are still tremendously popular; however, E-assessment has increased significantly in the decade since their last study on mathematics assessment. Systems like STACK and NUMBAS are widespread in many universities, mainly as part of some coursework components of first-year modules. These systems are attractive because of the time-saving aspects of electronic marking and the provision of rapid feedback (Iannone & Simpson, 2022). STACK is an open-source tool that can seamlessly integrate with learning management systems like Moodle, enabling the generation of randomised, automatically graded mathematics questions. Davies et al.'s (2024) work ratifies STACK as an assessment vehicle that delivers instant feedback, with adaptability in assessing complex mathematical concepts. NUMBAS is another open-source, web-based assessment tool designed specifically for mathematics. Similar to STACK, it provides interactive, automatically graded assignments and instantaneous feedback to participating students, which they tend to put great value on (Lishchynska et al., 2021). While the questions are customised and randomised, this can be quite a time-consuming process for lecturers who choose it (Lishchynska et al., 2021).

Valdez and Maderal (2021) state that online assessments in mathematics are increasing in use and popularity over the traditional paper-and-pen type as they evaluate student learning without the need for everyone to be physically present in the same room. The decrease in cost and increase in the availability of powerful technology have altered how many mathematics lecturers assess their modules (Stacey & Wiliam, 2012). Yet Iannone and Simpson (2022) found that there remains a relatively low level of variety in what they call 'the assessment diet' in mathematics in HE in the UK. There is a question over whether reasoning can be assessed online as efficiently as in a traditional examination setting; however, Sangwin (2019) demonstrated that typical closed-book exam questions in linear algebra could be replicated in an e-assessment system.

Academic integrity is a primary concern when selecting modes of assessment. Universities must uphold the academic veracity and exit standards of their degrees to preserve their reputation, and a move to online assessment can coincide with grade inflation if students are given increased time to complete assessments (Henley et al., 2022). Instances of academic misconduct have been shown to increase when assessment is fully in online format. Contract cheating refers to instances where students hire someone else to complete their work or provide answers on their behalf (Liyanagamage et al., 2025). Lancaster and Cotarlan (2021) found a 196% increase in contract cheating requests across five STEM subjects when comparing the periods from April to August 2019 and April to August 2020. This rise coincided with the shift to online assessments due to the COVID-19 pandemic. Trenholm (2007) contends that proctoring truly is the only method to eliminate cheating in online exams. However, Eaton and Turner's's (2020) research on E-proctoring, the systematic remote visual monitoring of students as they complete assessments, may have a detrimental impact on students' mental health and well-being. Sarmiento and Prudente (2019) demonstrated that ways around this are achievable. Their work illustrates that it is possible to assess online and limit opportunities for copying. They used MyOpenMath to generate individualised homework assignments for students. They found it not only limited copying but also had a significant positive impact on students' homework that was submitted and on their summative assessment performance (Sarmiento et al., 2018).

While there is an abundance of studies that examine student perceptions about online learning, there is a dearth in the literature on online assessment of mathematics (Valdez & Maderal, 2021), specifically which areas of mathematics lend themselves more to being assessed online. The university closures due to the COVID-19 pandemic forced lecturers worldwide out of their comfort zones and normal practices when it came to assessing their students. This research investigates the extent to which lecturers migrated from their conventional assessments when pressurised to do so, the lessons learnt during this time, the assessment changes that were preserved when they had a little more time to think and plan, and those that were discarded on their second attempt at online assessment. The research questions we explore in this paper are as follows:

1. What assessment types were used by mathematics lecturers before the pandemic, during the initial university closures, and during the academic year 2020/2021?
2. Were there changes to the weightings given to final examinations versus continuous assessment during these time periods?
3. Did mathematics lecturers observe any changes in grade distribution within their modules during these time periods?
4. Were mathematics lecturers satisfied with their assessment approaches during these time periods?
5. What do mathematics lecturers believe are the easiest and most difficult aspects of mathematics to assess online?

2. Materials and Methods

2.1. Sample

The profile of the respondents in the survey can be seen in detail in Table 1, with gender, age, years of experience in teaching mathematics in higher education, and employment status given. There was a total of 190 respondents to the survey. The gender breakdown of the respondents was 52% female, which does not reflect the population of mathematics lecturers in higher education, as this is predominantly male. The survey was sent to a mailing list of female mathematicians in Europe, which likely accounts for the high response rate from female mathematicians. The age profile was fairly evenly spread, with 85% of respondents between 30 and 59 years of age and a similar percentage in permanent employment. Their teaching experience in higher education reflected the age profile, with 40% of respondents having more than 20 years of experience.

Table 1. Profile statistics of survey respondents (*n*= 190), showing their gender, age, years of experience teaching mathematics in higher education, and current employment status.

	Number	%
Gender		
Male	87	46%
Female	99	52%
(Blank)	4	2%
Age		
20–29 years	8	4%
30–39 years	52	27%
40–49 years	60	32%
50–59 years	50	26%
60+ years	19	10%
(Blank)	1	1%
Experience teaching maths in higher education		
0–1 year	2	1%
2–3 years	11	6%
3–5 years	18	10%
5–10 years	28	15%
10–15 years	39	21%
15–20 years	17	9%
20+ years	75	40%
Employment status		
Ph.D./Postdoc	2	1%
Short-term contract (≤1 yr)	4	2%
Long-term contract (>1 yr)	22	12%
Permanent	162	85%

The respondents were based primarily in Europe, with only 12% based outside of the continent, most of these in the United States. The highest proportion by far was based in Ireland, as are the two researchers in this study. In total, respondents from 27 different countries answered the survey. Further details can be found in (Ní Fhloinn & Fitzmaurice, 2022).

Respondents were also asked about their mathematics teaching in the academic year 2020/2021 to provide context for their responses. Half of the respondents lectured students taking non-specialist (service) mathematics, while 60% taught students undertaking a mathematics major. We defined "small" classes as those having less than 30 students, "medium" as being those between 30 and 100 students, and "large" as more than 100 students. From our sample, 58% had small classes, 57% had medium classes, and 37% had large classes. Almost three-quarters of respondents did all their teaching online that year, with a further 17% doing it almost all online. Prior to the pandemic, 75% of respondents had done no online teaching of any kind, with a further 13% having done only a little.

2.2. Survey Instrument

This study falls under the umbrella of a larger investigation into the remote lecturing of tertiary-level students of mathematics during the COVID-19 pandemic. Our research design comprised the creation and dissemination of a purposely designed survey to investigate how lecturers and students were coping with remote mathematics education during the initial stages of the COVID-19 pandemic around the world. In 2020, we created an initial survey that would shed light on how mathematics lecturers were responding to the challenge of a spontaneous and forced move to remote teaching. The survey questions were original questions devised by the authors. The outcome of this research is reported in (Ní Fhloinn & Fitzmaurice, 2022).

To get a comprehensive insight into this experience and period of time where there were varying levels of restrictions enforced in 2021, we disseminated a follow-up survey in order to make comparisons between remote teaching in 2020 with that 12 months later. It is this follow-up survey on which this study focuses. It was largely based on the previous survey, with questions adapted to account for the fact that lecturers now had a year of remote teaching and assessment experience. It was an anonymous survey created in Google Forms, with a consent to participate checkbox on the landing page. We piloted the survey with a panel of experts within two university mathematics departments to increase the reliability and validity of the instrument. The lecturers were asked to review the questions to check for clarity of phrasing and if the questions were unbiased and aligned with the intended constructs. Suggestions for items we had not included were also welcomed. We asked the panel of experts to identify ambiguities, inconsistencies, or potential misinterpretations in a bid to improve content validity. Their feedback helped us refine question wording and format, thus increasing reliability. This process was conducted so that the survey would accurately address our research questions and produce consistent results across different respondents. The relevant survey questions are shown in Appendix A.

The survey began with a number of profiling questions, the results of which are shown in Table 1. There were then sections on their mathematics teaching allocation, the types of technology they used and the purpose of this, student experience, remote teaching experience, and personal circumstances, as well as the section which is the focus of this paper—online assessment in mathematics. Within this section, there were nine questions, four of which were open-ended.

2.3. Data Collection

The survey was made available exclusively online, using Google Forms, and was advertised via mailing lists for mathematics lecturers, as well as at mathematics education conferences relevant to those working in higher education.

2.4. Data Analyses

Respondents were asked in many places to give their approach at three specific time points: prior to the pandemic, during the initial university closures from March–June 2020, and during the academic year 2020/2021. Throughout the paper, we will refer to these as Period 1, Period 2, and Period 3 for simplicity.

Quantitative analysis was done using SPSS (version 29). For investigating binary changes over the three time periods, we used Cochran's Q test, as this is a non-parametric test for comparing binary outcomes across three or more time periods for the same group of subjects. In our results reported below, a statistically significant result is one for which the p-value is less than 0.05. Where the results of Cochran's Q tests were significant, we conducted post-hoc McNemar tests to ascertain where the significant differences lay. In this case, we needed to utilise a Bonferroni adjustment on the results because of making multiple comparisons, and so a statistically significant result is one for which the p-value is less than 0.017. To investigate changes in Likert-scale data over the three time periods, a Friedman test was used, as this is a non-parametric test for comparing ordinal data across three or more time periods for the same group of subjects. Where the results of the Friedman test were significant, we conducted post-hoc Wilcoxon signed-rank tests as above for the McNemar tests, with a similar Bonferroni adjustment.

For the qualitative responses, grounded theory was utilised, specifically general inductive analysis. Both researchers independently coded the responses initially to provide greater reliability in the results. Inter-coder reliability was 91%. We used the open discussion to categorise the disputed data points. This revealed minor differences in the interpretation of some data that, when clarified, led to almost perfect agreement.

3. Results

3.1. Assessment Types

Respondents were asked about the types of assessments that they conducted at three specific time points: Period 1, Period 2, and Period 3. The results are shown in Figure 1.

A Cochran's Q test was conducted on each assessment type, and the outcomes are shown in Table 2 below. From this, we found that there was a statistically significant difference in the proportion of respondents using most of these assessment types over time: namely, open-book timed, open-book untimed, closed-book, MCQ, online proctored, in-person proctored, presentations, projects and assignments.

Further post-hoc McNemar tests were conducted for the assessment types that were statistically significant, and the results are shown in Table 3 below. From this, it can be seen that for open-book timed and online proctored assessments, there was a statistically significant difference between each of the three time periods when considered in pairs—so there was a significant difference between Periods 1 and 2, Periods 1 and 3, and Periods 2 and 3. For open-book untimed, closed-book, MCQ and in-person proctored assessments, there was a statistically significant difference between Periods 1 and 2 and Periods 1 and 3, but not between Periods 2 and 3. For presentations, there was no significant difference between Periods 1 and 3, although the other two pairings were significant. For projects and assignments, the only significant difference was between Periods 2 and 3.

Table 2. Results of Cochran's Q tests conducted on each assessment type, where a *p*-value of $p < 0.05$ represents a statistically significant difference in the proportion of respondents using each assessment type.

Assessment Type	Cochran's Q Value	*p*-Value	Outcome
Open-book timed	138.078	<0.001	Significant
Open-book untimed	19.818	<0.001	Significant
Closed-book	163.538	<0.001	Significant
MCQ	27.167	<0.001	Significant
Online proctored	44.4	<0.001	Significant
In-person proctored	107.079	<0.001	Significant
Presentations	15.826	<0.001	Significant
Oral Assessments	2.583	0.275	Not significant
Essays	4.545	0.103	Not significant
Screencasts	4.750	0.093	Not significant
Projects	22.067	<0.001	Significant
Assignments	8.615	0.013	Significant

Table 3. Results of McNemar tests conducted on the assessment types that showed as statistically significant in Table Y. A Bonferroni adjustment was applied so that *p*-values below 0.017 were considered significant. P1 = Period 1, P2 = Period 2 and P3 = Period 3.

Assessment Type	Periods Compared	*p*-Value	Outcome
Open-book timed	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	<0.001	Significant
Open-book untimed	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	0.832	Not significant
Closed-book	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	0.064	Not significant
MCQ	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	0.307	Not significant
Online proctored	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	<0.001	Significant
In-person proctored	P1vP2	<0.001	Significant
	P1vP3	<0.001	Significant
	P2vP3	0.092	Not significant
Presentations	P1vP2	0.002	Significant
	P1vP3	0.864	Not significant
	P2vP3	<0.001	Significant
Projects	P1vP2	0.027	Not significant
	P1vP3	0.021	Not significant
	P2vP3	<0.001	Significant
Assignments	P1vP2	0.481	Not significant
	P1vP3	0.096	Not significant
	P2vP3	0.004	Significant

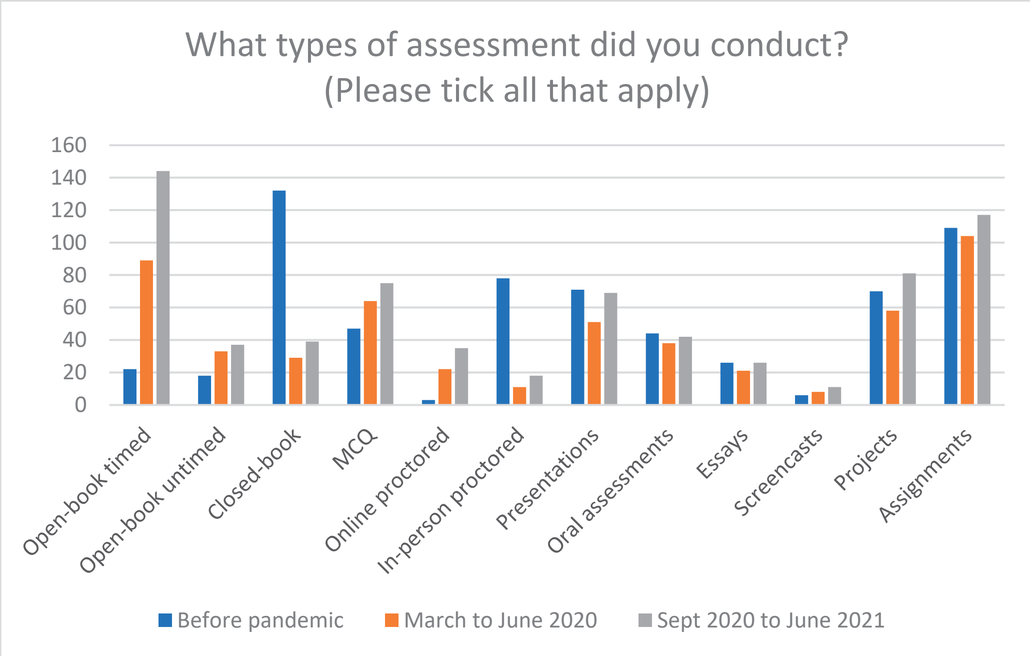


Figure 1. Responses to the question "What types of assessment did you conduct?" (n = 190). Respondents were permitted to tick more than one response, as appropriate.

3.2. Assessment Weightings

To measure the difference in weightings between different assessment types, respondents were asked how much their final examination was worth in comparison with continuous assessment in their modules. They had to select a percentage band, which was 20% wide (e.g., 100% final exam, 80–99% final exam, etc.). The results are shown in Figure 2. As can be seen in the chart, for Period 1, the most common approach was to have between 60–79% of the marks for the module assigned to the final examination (n = 68), which dropped during Periods 2 and 3 (n = 46) but remained the most common weighting throughout, although 40–59% became a very close second (n = 42) in Period 3. When a Friedman test was conducted on this data, there was a statistically significant difference between the weightings of final examinations and continuous assessments, measured across the three time periods ($\chi^2(2) = 27.752$, $p < 0.001$). Post-hoc analysis was conducted using Wilcoxon signed-rank tests with a Bonferroni correction applied, and this found statistically significant results between Period 1 and Period 2 ($Z = -4.112$, $p < 0.001$) and Period 1 and Period 3 ($Z = -3.879$, $p < 0.001$), but not between Period 2 and Period 3 ($Z = -0.866$, $p = 0.387$).

In fact, if we consider the changes for any individual respondent, shown in Figure 3, we find that 112 respondents did not alter how much their final exam was worth in relation to the continuous assessment at any point during the three time periods under investigation. Very few put more weight on final examinations between Periods 1 and 2, with most who made changes moving towards a greater weighting on continuous assessment (either a 1–19% increase or 60–79% increase, suggesting they either made minor weighting adjustments or substantial ones). Between Periods 2 and 3, the biggest adjustments were plus or minus 1–19% change in weightings, with respondents almost equally adjusting towards more final examination weighting or more continuous assessment weighting. Overall, looking at the change between Periods 1 and 3, there was a noticeable shift back towards increased weighting for continuous assessment, although again the most common response was a 1–19% increase in weighting.

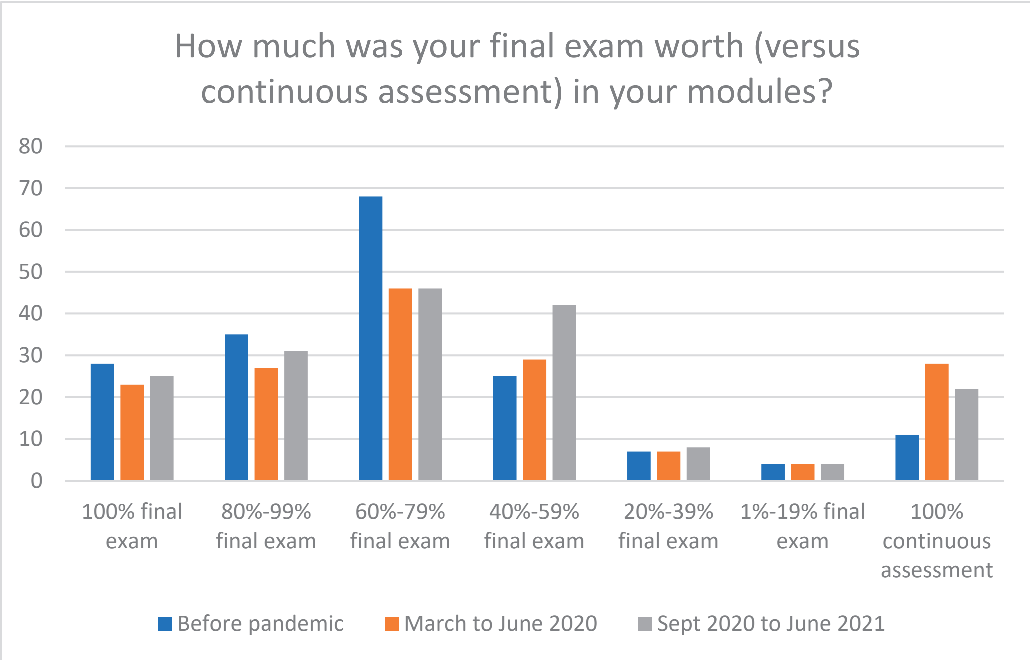


Figure 2. Responses to the question "How much was your final exam worth (versus continuous assessment) in your modules?" (n = 178).

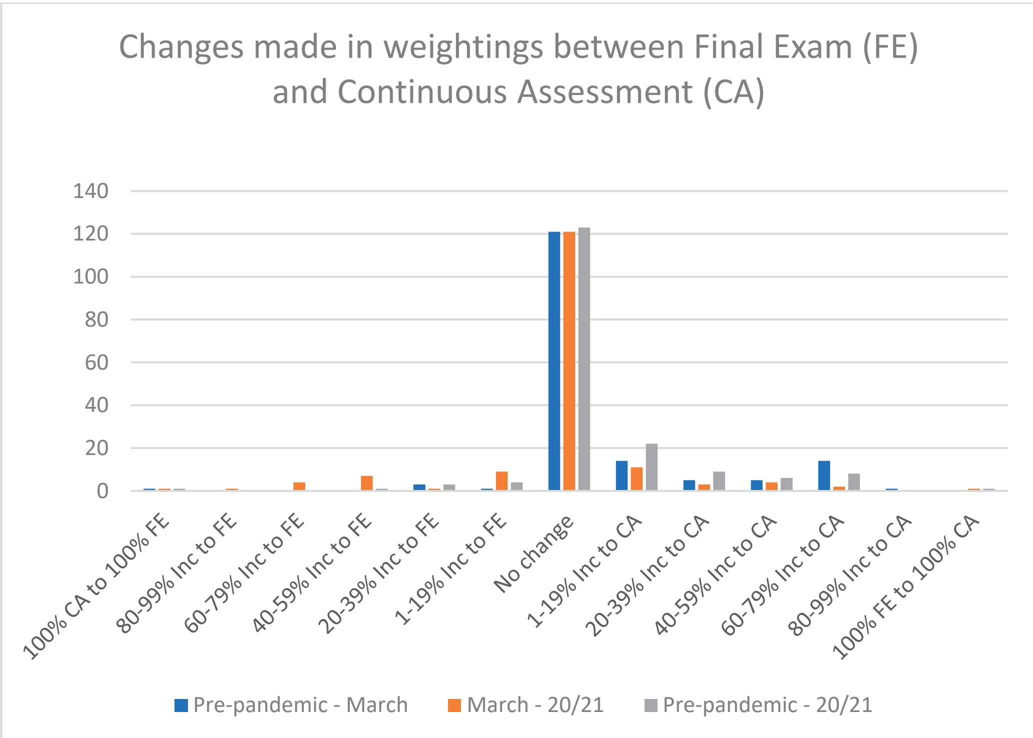


Figure 3. Changes in individuals' module weightings between the Final Exam (FE) and Continuous Assessment (CA) between (i) pre-pandemic and March–June 2020 (ii) March–June 2020 and the academic year 2020–2021 and (iii) pre-pandemic and the academic year 2020–2021.

There were 73 additional comments on this question, many describing specifics about their approach, but some elaborating on reasons for the choices made in terms of assessment at the various stages. The most common themes are highlighted in Table 4 below, along with a sample comment made under each theme.

Table 4. Responses given to the question "Any comments on changes to your approach to assessment this academic year?" (*n* = 73).

Themes	Number of Responses	Sample Comment
Cheating	10	Open book online assessments led to lots of cheating (e.g., Chegg)
No choice	5	The university rules did not allow changing the balance of exam vs. continuous assessment
Previous experience	4	Multiple choice exam last June in Linear Algebra was awful to create and very bad at assessing students—average mark was 90%. This year's Stack based questions for quizzes and exams on Moodle was much more discriminating
Time-consuming	4	Assessment became much more difficult. In particular, the process of creating an exam and finding appropriate problems (which could not be solved trivially using online software or Math Stack Exchange) was extremely time-consuming

3.3. Grade Distribution

Respondents were then asked if they observed any differences in the distribution of grades within their modules during Period 3. Only just over 3% felt that there was a big difference in grade distribution that year, with 22% stating that there was no difference at all and the remaining responses falling somewhere between these poles, as shown in Table 5.

Table 5. Responses given to the question "Was there a difference in the distribution of grades within your modules this academic year?", where 1 = no difference and 5 = a big difference (*n* = 181).

Ranking	Number of Responses
1	40
2	52
3	54
4	29
5	6

When asked for their interpretation of any differences they observed, there were 89 further comments made. The most common themes to emerge are shown in Table 6.

Table 6. The five most common themes mentioned by respondents when asked, "If you saw a difference (in grades), why do you think this was?".

Theme	Number of Occurrences	Sample Comments
Exam type	40	Grades are possibly higher due to the large amount of coursework/continuous assessment. I don't think I would have the same number of A's if it was the usual proctored exam.
Copying	19	Students adapt to online exams and are more likely to cheat. It doesn't appear that there was more cheating than there usually is.
Student divide	19	More bimodal—perhaps reflecting the difficulties a subgroup of students had with online learning. Struggling students withdrew but focused students had more time to study.
Engagement	9	Engagement in sessions was not as high during online learning as students were not as accountable as they would be in-person. Some students didn't/couldn't engage with the online environment.
Lecturer experience	9	It was hard to pitch the level of difficulty on Webwork. . .I think that was too generous and the resultant marks were slightly higher. Students scored higher in some classes because my questions were a bit too 'traditional'.

3.4. Satisfaction with Assessments

Respondents then rated their satisfaction levels with their assessments during the three time periods in question. The results are shown in Figure 4. It can be seen that three-quarters of respondents to this question were either satisfied or very satisfied with their assessments during Period 1. This dropped as low as two-fifths of respondents for Period 2 and recovered to 55% for Period 3. When a Friedman test was conducted, there was a statistically significant difference between satisfaction levels with assessments over time ($\chi^2(2) = 84.874$, $p < 0.001$). Post-hoc analysis was conducted using Wilcoxon signed-rank tests with a Bonferroni correction applied, and this found statistically significant results between each of the three time periods when compared with each other in terms of respondent satisfaction with their assessments (Period 1 vs. Period 2: $Z = -7.333$, $p < 0.001$; Period 2 vs. Period 3: $Z = -4.564$, $p < 0.001$; Period 1 vs. Period 3: $Z = -5.286$, $p < 0.001$).

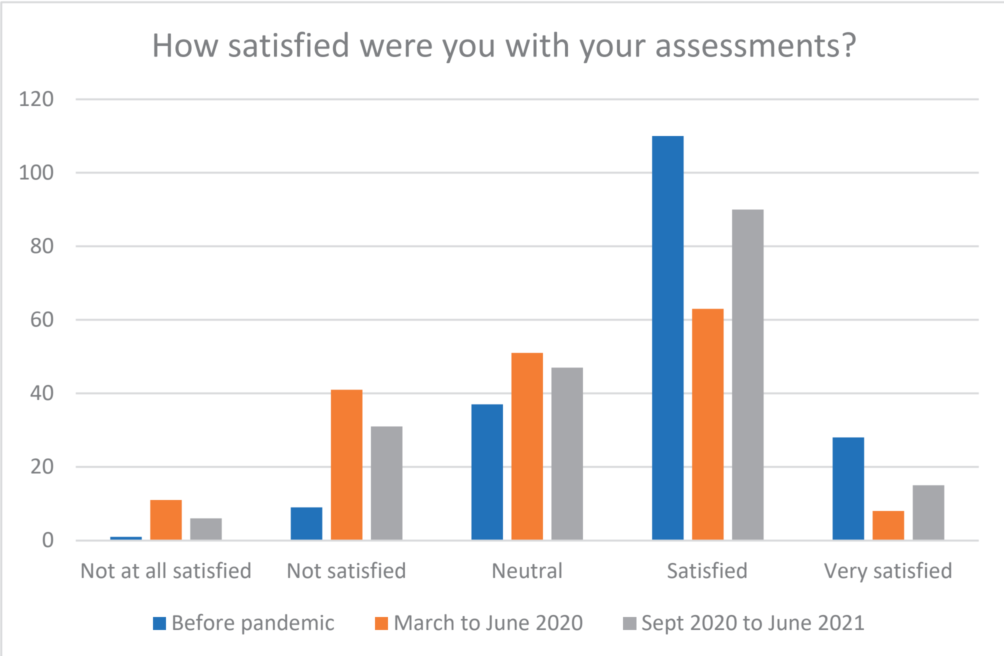


Figure 4. Responses to the question: "How satisfied were you with your assessments?" where respondents could select one response from a five-point Likert scale as shown in the chart. Respondents were asked the question in relation to the three time periods: before the pandemic; from March to June 2020; and during the academic year 2020–2021.

3.5. Easiest/Hardest Aspects of Mathematics to Assess Online

Finally, respondents were asked two more general questions about assessment in mathematics: namely, what aspects of mathematics they considered to be most difficult and easiest to assess online. The most common themes to emerge from the 123 responses to each question are shown in Table 7.

Table 7. The most common themes that emerged when respondents were asked "What aspects of maths do you think are most difficult to assess online?" and "What aspects of maths do you think are easiest to assess online?" ($n = 123$).

Themes: Hardest to Assess	Frequency	Themes: Easiest to Assess	Frequency
Proof	29	Computation	38
Computation	26	None	13
Cheating	24	Understanding	11
Understanding	16	MCQs	8
Online solving	13	Cheating	5
Theory	12	Proof	5
Skills	9	Programming	4
Reasoning	7	Statistics	4

Tables 8 and 9 show the most common themes under "hardest to assess" and "easiest to assess", respectively, along with the frequency of the theme and some sample comments made by respondents.

Table 8. The most common themes emerged when respondents were asked, "What aspects of maths do you think are most difficult to assess online?" along with sample comments under each theme (n = 123).

Theme	Number of Occurrences	Sample Comments
Proof	29	I believe there is much merit in asking students to learn to state theorems and to prove them. But this is tricky to assess online.
Computation	26	I think any procedural questions for service maths modules. . .are difficult to assess online—the fact that there are so many websites where fully worked solutions to questions can be easily generated makes it very difficult to stand over the integrity of online assessments.
Cheating	24	The fact that there are so many websites where fully worked solutions to questions can be easily generated makes it very difficult to stand over the integrity of online assessments.
Understanding	16	As 'bookwork' questions (e.g., proofs of theorems) are not possible. . .it can be difficult to assess a student's understanding of the theory of any particular topic.
Online solving	13	The availability of online computational software and problem-solving sites makes it very difficult to assess students online. I don't see any way around this without radically altering what we assess for. To some extent, this might be worth discussing even in the absence of online education.
Theory	12	Theoretical content
Skills	9	Basic skills
Reasoning	7	The problem I had was assessing their ability to think with insight into unfamiliar problems.

Table 9. Most common themes that emerged when respondents were asked "What aspects of maths do you think are easiest to assess online?" along with sample comments under each theme (n = 123).

Theme	Number of Occurrences	Sample Comments
Computation	38	Questions that have parameters are relatively easy to vary without changing the difficulty of the question.
None	13	There is really no aspect of Maths that is easy to assess online.
Understanding	11	More abstract concepts are possibly a little easier to assess online as the understanding of the students can be tested in different ways.
MCQs	8	Multiple choice questions and quizzes can be easily developed and completed which does help with assessment between sessions.
Cheating	5	However, there is no way to control or prevent students to contact each other during a timed exam they sit at home. Therefore, it is not clear if the exams assess if a particular student has understood the material.
Proof	5	Standard questions that are not (easily) solved on webpages or computer software like Mathematica. Such as 'show this function is continuous', 'show this subgroup is normal'.
Programming	4	Programming. . ., by design, has been assessed online for a very long time—anything else wouldn't really make sense.
Statistics	4	Giving students real data to work with and setting them online questions based on their individual data and/or project work. Students are very engaged too as they feel it is useful.

4. Discussion

In this study, we explored the assessment approaches taken by mathematics lecturers at three distinct periods to determine whether the forced inclusion of online assessment as a result of the COVID-19 campus closures might alter their opinions towards diversifying their assessment longer-term. Assessment in mathematics typically over-emphasises replication of content (facts) and skills (techniques) (Pegg, 2003). Assessment design should be heavily influenced by the mathematics that is considered most important for students to learn (Stacey & Wiliam, 2012). The respondents in this sample conducted almost no online teaching in Period 1, exclusively online teaching in Period 2, and almost all online teaching in Period 3. As such, the latter two periods did not reflect the "normal" situation for many of them but instead acted as a unique snapshot of time whereby they had to use online assessment.

4.1. What Assessment Types Were Used by Mathematics Lecturers Before the Pandemic, During the Initial University Closures, and During the Academic Year 2020/2021?

Our first research question explored the assessment types used by mathematics lecturers during the three periods under investigation. We found statistically significant changes in assessment types used by our sample in relation to 9 of the 12 assessment types investigated. We established that our sample was typical of the mathematics lecturer population during Period 1, in that 70% of respondents used closed-book examinations and 57% used assignments, both of which would be commonly used mathematics assessment approaches (Iannone & Simpson, 2022). By Period 3, open-book timed examinations had replaced the more popular closed-book examinations, even more so than in Period 2, when many respondents did not have sufficient time to implement such changes in their assessment. The findings clearly show a broader range of assessment modes in use in Period 3. After open-book timed assessments, projects, assignments and MCQs are the most popular modes of assessment. This is of interest as, according to Niss (1998), relaxing timing restrictions for students allows one to assess a broader range of concepts and skills. Suurtamm et al. (2016) say that implementing a range of assessment approaches allows students multiple opportunities to utilise feedback and demonstrate their learning. When a lecturer relies on one type of assessment, students often become experts at predicting likely assessment areas and choose their areas of revision accordingly (Burton & Haines, 1997). Seeley (2005) recommends designing assessments by "incorporating problem-solving, open-ended items, and problems that assess understanding as well as skills", making assessment an integral part of teaching and learning. Of the three assessment types where no significant difference was observed, oral assessments were the most commonly used (by just over 20% of respondents). Although this was an under-utilised assessment type, it does appear to have been one that was able to be used across all three time periods, regardless of whether the assessment was in-person or online. This makes it an assessment type worthy of further consideration in relation to mathematics (Iannone et al., 2020).

4.2. Were There Changes to the Weightings Given to Final Examinations Versus Continuous Assessment During These Time Periods?

The proportion of lecturers who gave 80–100% final exams remained steady between Periods 1 and 3. There was a shift from the proportion who offered a 60–79% final exam towards a 40–59% exam. We do observe a statistically significant increase in lecturers favouring 100% continuous assessment in Period 3 when compared with Period 1. The number of those in favour of a 100% continuous assessment is in line with those in favour of a 100% final exam.

4.3. Did Mathematics Lecturers Observe Any Changes in Grade Distribution Within Their Modules During These Time Periods?

Overall, lecturers did not observe a vast difference in grade distribution between the three periods in question, although they gave varying reasons for this. The most prevalent theme was that of the type of examination that the students undertook. Some comments related this to a higher weighting on continuous assessment ("*Grades are possibly higher due to the large amount of coursework/continuous assessment*"), while others referenced the "*binary right/wrong marking*" of some online assessment tools, particularly in relation to multiple-choice questions. A number of respondents spoke of the impact of using open-book examinations where previously they would have had closed-book ("*I don't think I would have the same number of A's if it was the usual proctored exam*"). However, opinions were mixed on whether this was more or less difficult for the students, with one respondent stating, "*open book exams plus timed online format skewed harder than in previous years as no book answers*" while another felt that "*open book made it easy for students to look at worked examples*". The next most common theme was that of copying—either another student's work or using an online tool to cheat. Respondents observed that "*cheating online is very easy*" and that "*students adapt to online exams and are more likely to cheat*". Conversely, one respondent felt that "*it doesn't appear that there was more cheating than there usually is*"; however, this respondent stated that their module was fully assessed by continuous assessment both before the pandemic and during the academic year 2020/2021. Additionally, several respondents remarked on how, although they suspected there were students cheating, there was not as big a difference in the grade distribution as might have been expected because "*there was probably cheating, but at the same time some students were less motivated*", and others echoed approaches similar to that in (Sarmiento & Prudente, 2019). Another strong theme to emerge was that of an increased divide in student performance, with many respondents pointing out how difficult it was for some of their students ("*More bimodal—perhaps reflecting the difficulties a subgroup of students had with online learning*"). Several respondents felt that, although it was more difficult for some students, others coped well ("*I feel that some students were not able to concentrate online. The high achievers were able to perform well in both situations*"), or in fact that some students excelled in this situation ("*Struggling students withdrew but focused students had more time to study*"). Several respondents also mentioned student engagement as a reason for greater disparity in grades, stating that "*some students didn't/couldn't engage with the online environment*" and "*engagement in sessions was not as high during online learning as students were not as accountable as they would be in-person*". This could partly be attributed to the pandemic conditions, in which students from lower socio-economic backgrounds, or those with caring responsibilities, were impacted more in terms of engaging with online learning (Ní Fhloinn & Fitzmaurice, 2021). Finally, a number of respondents also alluded to their own lack of experience in setting different types of assessments as a reason for a differing grade profile, with some stating that they felt they made the assessment too easy as a result ("*it was hard to pitch the level of difficulty on Webwork. . .I think that was too generous and the resultant marks were slightly higher*"), while others felt the opposite had occurred ("*Students scored higher in some classes because my questions were a bit too 'traditional'*"). Others simply observed that "*there is a learning curve in terms of making a good online multiple-choice exam when all study aids are allowed. It is possible but requires some clever thinking*". This echoes the warning of Greenhow (2015) in relation to the creation of online assessments.

4.4. Were Mathematics Lecturers Satisfied with Their Assessment Approaches During These Time Periods?

Although the academics had increased the number of assessment methods in use, fewer respondents were satisfied with their assessments in Period 3 than before the pandemic in Period 1. The differences in satisfaction levels between all three time periods were significant, with the highest satisfaction ratings given pre-pandemic, and the next highest in Period 3. Overall satisfaction was significantly lower in Period 2, which was to be expected, as this was when lecturers had to pivot to new assessment types with very little time or resources to implement or plan these new assessments.

4.5. What Do Mathematics Lecturers Believe Are the Easiest and Most Difficult Aspects of Mathematics to Assess Online?

Our final research question was in relation to what aspects of mathematics lecturers believed were easiest or most difficult to assess online. The end of Period 3 represented a unique snapshot in time to obtain this information, whereby the majority of lecturing staff had been obliged to conduct online assessments in mathematics, regardless of their previous experience with online teaching (Ní Fhloinn & Fitzmaurice, 2021). We particularly wanted to investigate if mathematics lecturers perceive an online environment to be more conducive to the assessment of some mathematical areas than others. Respondents mostly spoke about the topic being assessed. The most common theme among "hardest to assess" was that of "proof". Many respondents pointed out the difficulty of assessing theorems online in the same manner as they would have in a closed, proctored examination ("*I believe there is much merit in asking students to learn to state theorems and to prove them. But this is tricky to assess online*"). Other respondents noted that it was difficult to assess students' ability to "*reproduce a long technical proof*", although some respondents did question whether there was evidence of understanding previously in this approach ("*Understanding (or, perhaps, if we are very honest, memory of) proofs.*"), and observed that "*it does make you think about key aspects of a proof not just memory and regurgitation*". The small number of instances of "proof" being mentioned in terms of being easiest to assess online leaned towards "*standard questions that are not (easily) solved on webpages or computer software like Mathematica. Such as 'show this function is continuous', 'show this subgroup is normal'*" or "*applications of theorems*".

Notably, "computation" appeared as the second-most common theme in terms of "most difficult to assess" and as the most common theme in terms of "easiest to assess". Those who found it easiest to assess tended to make comments around the ease of using software to develop "*questions that have parameters are relatively easy to vary without changing the difficulty of the question*". In contrast, many of those who found it hardest to assess linked this difficulty to the issue of students cheating, either by getting the solution from their classmates or online software ("*I think any procedural questions for service maths modules. . .are difficult to assess online—the fact that there are so many websites where fully worked solutions to questions can be easily generated makes it very difficult to stand over the integrity of online assessments*"). Similarly, some mentioned the difficulty of assessing basic computational techniques in open-book examinations ("*As we only used open-book methods when assessments were online, it is hard to gauge if students have gained mathematical skills or if they are very dependent on copying procedures from worked examples*").

Yet more mentioned the drawbacks of "*MCQs [multiple-choice questions] that involve lots of arithmetic/algebraic manipulation—it is difficult to give half marks/quarter marks to students who have done it partly correct but reached an incorrect result*".

Plagiarism, copying or "cheating" was the third most common theme in the aspects that are hardest to assess. This was frequently linked to either the "proof" or "computation" themes, but also often cited on its own, with one respondent stating that the hardest thing to assess was "*anything where you want to ensure the work is the student's own*" and another that it was "*anything you can find the answer to online*". The difficulty of detecting such activity was summed up by one respondent who said that "*catching collusion can be tricky, as sometimes there is basically one obvious way to get to the right answer*". Another observed that "*the fact that there are so many websites where fully worked solutions to questions can be easily* *generated makes it very difficult to stand over the integrity of online assessments*". This theme emerged a few times also among the comments in relation to the easiest aspects to assess; usually as an addendum to other comments, such as "*However, there is no way to control or prevent students to contact each other during a timed exam they sit at home. Therefore, it is not clear if the exams assess if a particular student has understood the material*".

The theme of "understanding" occurred almost equally frequently in both the hardest and easiest aspects to assess online. Among those who deemed it hardest to assess online, many simply mentioned "*conceptual understanding*" without elaboration; however, a few mentioned that it was difficult to know "*whether students understand definitions (because open book)*" and "*as 'bookwork' questions (e.g., proofs of theorems) are not possible. . .it can be difficult to assess a student's understanding of the theory of any particular topic*". For the respondents who deemed "understanding" easiest to assess, again several simply mentioned "*conceptual understanding*". However, a couple explained that "*more abstract concepts are possibly a little easier to assess online as the understanding of the students can be tested in different ways*" and that misconceptions could be exposed *"using basic mcqs or true false. It doesn't require any written mathematics and can expose key misunderstanding. Although why those exist might require alternative approaches*".

The theme "*online solving*" made specific reference to students using apps, websites or software to solve the questions they were asked in their assessments and only appeared in relation to the hardest aspects to assess. One respondent linked this to a need to consider what it is that is assessed and the purpose of such assessment, even outside the need to assess online ("*The availability of online computational software and problem-solving sites makes it very difficult to assess students online. I don't see any way around this without radically altering what we assess for. To some extent, this might be worth discussing even in the absence of online education*").

The three other most common themes of "*theory*", "*skills*" and "*reasoning*" again appeared only in the hardest aspects to assess online. Respondents did not elaborate much on any of these themes, mostly just mentioning "*theoretical content*", "*basic skills*" or "*mathematical reasoning*". However, one respondent did elaborate in terms of reasoning to explain that "*the problem I had was assessing their ability to think with insight into unfamiliar problems*".

In terms of the easiest aspects to assess online, the second-most common was "*none*" with no further comments made on this. A number of respondents mentioned "*multiple-choice questions (MCQs)*", stating "*multiple choice questions and quizzes can be easily developed and completed which does help with assessment between sessions*". However, another respondent cautioned that "*aspects that can be validly assessed by multichoice questions lend themselves to being tested in hands-on mode . . . but the design of good MCQs is very challenging, and not all knowledge and skills can be covered*".

Finally, both "*programming*" and "*statistics*" were mentioned a number of times as being the easiest to assess online. In relation to the former, one respondent observed that "*programming. . ., by design, has been assessed online for a very long time—anything else wouldn't really make sense*". For statistics, another respondent suggested "*giving students real data to work with and setting them online questions based on their individual data and/or project work. Students are very engaged too as they feel it is useful*".

5. Conclusions

To conclude, only one significant change in assessment approach was observed across the three periods of study; a move from closed- to open-book timed examinations. Academic integrity was clearly a strong consideration of the lecturers who participated in this study. For them, this refers to cheating and significantly influences the assessment strategy they adopt. This echoes the findings of Henley et al. (2022), whose large-scale survey of mathematics departments across the UK and Ireland found academic misconduct on the part of students to be of significant concern over the same time period as this study. The authors state that a community-wide approach will be necessary moving forward if open-book online assessments are to be continued. They found that increasing the time available for students to complete the assessments remotely led to increases in instances of plagiarism and cheating (Henley et al., 2022). Assigning randomised and personalised assessments is one method suggested in recent studies for reducing this misconduct (Eaton & Turner's, 2020; Sarmiento & Prudente, 2019). While this may sound very laborious and time-consuming, Henley et al. (2022) state that this is not necessarily true. Respondents to their survey listed several methods of achieving this, including, for example, LaTeX Macros, which can be used to create personalised mathematical problems and generate random values for assessments. Online proctoring may reduce incidents of cheating (Lancaster & Cotarlan, 2021) but then result in negative mental health outcomes for others (Eaton & Turner's, 2020).

Moving away from closed-book, proctored assessments in any capacity is likely to lead to grade inflation for a variety of reasons. Our findings echo those of Iannone and Simpson (2022), who in 2021 found that there is little variety in the assessment strategies of mathematics departments in Higher Education institutions in the UK. They too found that the weighting of final assessments was shifting towards the inclusion of other methods, but progress is slow. Despite the unprecedented visit of a pandemic and a drastic sudden move to remote teaching and assessment, assessment in mathematics remained largely unchanged. Research conducted since the pandemic provides evidence that remote assessment of mathematics posed a momentous challenge for many lecturers during this unprecedented period (Cusi et al., 2023). More worryingly, a study conducted with almost 470 university mathematics lecturers in Kuwait and the UK during the same time period as reported in this study found that the lecturers were very much unconvinced of the merits of online assessment in mathematics and would likely revert to their tried and trusted methods when restrictions were lifted (Hammad et al., 2025). Continued professional development programmes that support the effective implementation of practical and fair assessment practices and provide practical strategies and hands-on training are necessary if progress forward instead of backwards is to be sustained (Cusi et al., 2023).

The future of mathematics education holds great promise for technological innovations and the integration of Generative Artificial Intelligence (GenAI). Personalised learning is expected to become more widespread, with adaptive learning platforms and intelligent tutoring systems analysing students' needs and providing tailored instruction and assessment. This study was undertaken when the capacities of GenAI were considerably weaker than they are at the time of writing, and given the concerns expressed by lecturing staff in relation to academic integrity at this earlier time, we feel that this is likely to be at the forefront of any decisions taken in relation to new assessment approaches to mathematics in the future. A follow-up mixed-methods study is planned to investigate the current status of assessment in mathematics globally, with the emergence of GenAI and its availability to both lecturers and students.

There are a number of limitations which should be taken into account in relation to this study. Firstly, 36% of respondents were based in Ireland, with a total of 88% of respondents based in Europe, meaning that the results can only be considered reflective of practice in European countries and may not be generalisable further afield. In addition, the survey was only available in English; it was distributed and conducted online via mailing lists and advertisements at relevant mathematics education conferences. As a result, it is unknown how representative it is of the general population of mathematics lecturers in higher education.

Author Contributions: Conceptualization, E.N.F. and O.F.; methodology, E.N.F. and O.F.; validation, E.N.F. and O.F.; formal analysis, E.N.F. and O.F.; investigation, E.N.F. and O.F.; writing—original draft preparation, O.F.; writing—review and editing, E.N.F.; visualization, E.N.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Dublin City University (protocol code: DCUREC/2020/079; date of approval: 1 May 2020).

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

Data Availability Statement: Data are contained within the article.

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

Appendix A. Survey Questions on Assessment

How much of your assessment was conducted online versus in-person this academic year (2020–2021)? (5-point Likert scale)
What types of assessment did you conduct? (Please tick all that apply)

Assessment Type	Before the Pandemic	During Initial University Closures (March–June 2020)	Academic Year 2020–2021
Open-book timed			
Open-book untimed			
Closed book			
Multiple-choice questions (MCQ)			
Online proctored			
In-person proctored			
Presentations			
Oral assessments			
Essays			
Screencasts			
Projects			
Assignments			

How much was your final exam worth (versus continuous assessment) in your modules?

% Weighting	Before the Pandemic	During Initial University Closures (March–June 2020)	Academic Year 2020-2021
100% Final Exam			
80–99% Final Exam			
60–79% Final Exam			
40–59% Final Exam			
20–39% Final Exam			
1–19% Final Exam			
100% Continuous Assessment			

Any comments on changes to your approach to assessment this academic year?
Was there a difference in the distribution of grades within your modules this academic year? (5-point Likert scale)
If you saw a difference, why do you think this was?
How satisfied were you with your assessments? (5-point Likert scale)

What aspects of maths do you think are most difficult to assess online?
What aspects of maths do you think are easiest to assess online?

References

Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. *Education Sciences*, *11*(6), 279.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, *19*(2), 151–161. [CrossRef]

Burton, L., & Haines, C. (1997). Innovation in teaching and assessing mathematics at university level. *Teaching in Higher Education*, *2*(3), 273–293.

Challis, N., Houston, K., & Stirling, D. (2003). *Supporting good practice in assessment in mathematics, statistics and operational research*. Available online: https://www.advance-he.ac.uk/knowledge-hub/supporting-good-practice-assessment-mathematics-statistics-and-operational-research (accessed on 17 December 2024).

Cusi, A., Schacht, F., Aldon, G., & Swidan, O. (2023). Assessment in mathematics: A study on teachers' practices in times of pandemic. *ZDM–Mathematics Education*, *55*(1), 221–233. [CrossRef]

Davies, B., Crisan, C., Geraniou, E., & Smart, T. (2024). A department-wide transition to a new mode of computer-aided assessment using STACK. *International Journal of Research in Undergraduate Mathematics Education*, *10*, 850–870. [CrossRef]

Eaton, S. E., & Turner, K. L. (2020). Exploring academic integrity and mental health during COVID-19: Rapid review. *Journal of Contemporary Education Theory & Research (JCETR)*, *4*(2), 35–41. [CrossRef]

Engelbrecht, J., & Harding, A. (2004). Combing online and paper assessment in a web-based course in undergraduate mathematics. *Journal of Computers in Mathematics and Science Teaching*, *23*(3), 217–231.

Fitzmaurice, O., & Ní Fhloinn, E. (2021). Alternative mathematics assessment during university closures due to COVID-19. *Irish Educational Studies*, *40*(2), 187–195. [CrossRef]

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*(4), 2333–2351.

Greenhow, M. (2015). Effective computer-aided assessment of mathematics; Principles, practice and results. *Teaching Mathematics and Its Applications*, *34*(3), 117–137. [CrossRef]

Greenhow, M. (2019). Embedding e-assessment effectively. *MSOR Connections*, *17*(3), 63.

Hammad, S., Wood, R., Hunter, G., & Dimitriadis, C. (2025). University mathematics lecturers' experiences of teaching and learning during COVID-19 Pandemic: A comparative study between Kuwait and United Kingdom. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(3), em2603. [CrossRef]

Henley, M., Grove, M., & Hilliam, R. (2022). University mathematics assessment practices during the COVID-19 pandemic. *MSOR Connections*, *20*(3), 5–17. [CrossRef]

Iannone, P., Czichowsky, C., & Ruf, J. (2020). The impact of high-stakes oral performance assessment on students' approaches to learning: A case study. *Educational Studies in Mathematics*, *103*, 313–337. [CrossRef]

Iannone, P., & Simpson, A. (2011). The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and Its Applications*, *30*(4), 186–196. [CrossRef]

Iannone, P., & Simpson, A. (2022). How we assess mathematics degrees: The summative assessment diet a decade on. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, *41*(1), 22–31. [CrossRef]

James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. *International Journal of Educational Technology in Higher Education*, *13*, 1–13.

Kerka, S., & Wonacott, M. E. (2000). *ERIC clearinghouse on adult, career and vocational education: Practitioner file*. ERIC.

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: A COVID-19 pandemic perspective. *International Journal for Educational Integrity*, *17*(1), 3. [CrossRef]

Lishchynska, M., Palmer, C., & Cregan, V. (2021). Online mathematics supports in lieu of university tutorials during COVID-19: Student and lecturer perspectives. *Teaching Mathematics and Its Applications*, *40*(4), 417–435. [CrossRef]

Liyanagamage, N., Asylguzhina, E., Vel, P., Khan, Z. R., Fernando, M., & Lancaster, T. (2025). Methodological and ethical approaches to studying contract cheating: A systematic literature review. *Journal of Academic Ethics*, 1–27. [CrossRef]

Niss, M. A. (1998). Dimensions of geometry and assessment. In *Perspective on the teaching of geometry for the 21st century: An ICMI study* (pp. 263–274). Kluwer Academic Publishers.

Ní Fhloinn, E., & Carr, M. (2017). Formative assessment in mathematics for engineering students. *European Journal of Engineering Education*, *42*(4), 458–470. [CrossRef]

Ní Fhloinn, E., & Fitzmaurice, O. (2021). Challenges and opportunities: Experience of mathematics lecturers engaged in emergency remote teaching during the COVID-19 pandemic. *Mathematics*, *9*(18), 2303. [CrossRef]

Ní Fhloinn, E., & Fitzmaurice, O. (2022, June 24–25). *Getting to grips with online mathematics education during the COVID-19 pandemic* [Conference session]. Science and Mathematics Education Conference (SMEC) 2022, Dublin, Ireland. [CrossRef]

Pegg, J. (2003). Assessment in mathematics. In J. M. Royer (Ed.), *Mathematical Cognition* (pp. 227–259). Information Age Publishing Inc.

Sangwin, C. (2013). *Computer aided assessment of mathematics*. OUP Oxford.

Sangwin, C. (2019, February 6–10). *Developing and evaluating an online linear algebra examination for university mathematics* [Conference session]. Eleventh Congress of the European Society for Research in Mathematics Education, Utrecht, The Netherlands.

Sarmiento, C. P., Lapinid, M. R. C., & Prudente, M. S. (2018). The effect of online homework in the performance of mathematics of accounting and finance students. *Advanced Science Letters*, *24*(11), 7956–7960.

Sarmiento, C. P., & Prudente, M. S. (2019). Preventing Homework copying through online homework in a math class. *The Normal Lights*, *13*(1), 27–55. [CrossRef]

Seeley, C. L. (2005, September). Assessing to learn and learning to assess. *NCTM News*.

Stacey, K., & Wiliam, D. (2012). Technology and assessment in mathematics. In *Third international handbook of mathematics education* (pp. 721–751). Springer.

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. Springer Nature.

Trenholm, S. (2007). A review of cheating in fully asynchronous online courses: A math or fact-based course perspective. *Journal of Educational Technology Systems*, *35*(3), 281–300. [CrossRef]

Valdez, M. T. C. C., & Maderal, L. D. (2021). An analysis of students' perception of online assessments and its relation to motivation towards mathematics learning. *Electronic Journal of E-Learning*, *19*(5), 416–431.

education sciences

MDPI

Article

(Up)Grading: A (Re)Humanizing Assessment Process with a Focus on Feedback

Stefanie D. Livers [1,*], Kristin E. Harbour [2] and Patrick L. Sullivan [3]

[1] School of Inclusive Teacher Education, Bowling Green State University, Bowling Green, OH 43403, USA
[2] Department of Teacher Education, University of South Carolina, Columbia, SC 29208, USA; kharbour@mailbox.sc.edu
[3] Department of Mathematics, Missouri State University, Springfield, MO 65897, USA; patricksullivan@missouristate.edu
* Correspondence: slivers@bgsu.edu

Abstract: Researchers across two universities and three different mathematics education courses implemented their vision of a novel grading approach called (Up)grading. (Up)grading shifts the focus of assessment from grades to growth. Key features of implementing (Up)grading included (a) providing students with opportunities to reflect upon and grow from their learning experiences, and (b) giving them a voice in determining their course grade. The findings suggest that most students perceived (Up)grading as a positive experience in their learning and as an assessment approach. The features of (Up)grading students believed contributed to the positive experience included giving them opportunities to reflect on their work and learn from their mistakes, as well as targeted feedback, enabling them to independently move their thinking forward. Tensions in the process did arise, including students' initial anxiety with the norm shift from grades to growth and instructors' management of the flow of assignments.

Keywords: assessment; upgrading; rehumanizing practices; self-assessment

1. Introduction

In the quest for teaching mathematics for understanding, the connection between learning and assessment often becomes disjointed if the assessment practices are based on compliance and completion. The call for assessment practices of teacher candidates that mirror K-12 began in the 1990s [1] as standards-based instruction became the focus. The standards-based movement was grounded in learning theory and environments that support constructivist practices [2]. Although teacher education programs have incorporated performance assessments and have been revised to align more with those in their P-12 counterparts, the requirement of grading and particular grading scales hinder a focus on assessment for growth, reflection, and feedback. The grading system never was meant to increase engagement but fostered a climate where "mistakes are unwanted, unhelpful, and punished" [3] (p. 30). It is noted that mathematics teacher educators (MTEs) wishing to "rehumanize" [4] (p. 1) elements of their instructional practice to better align with perspectives on teaching and learning are often faced with challenges from their departments and institution [5]. Despite the move to be more standards-based and student focused, grading remains something we *do* to students rather than *with* students, meaning there is not a shared authority in the process. Problematically, "grades [. . .] have contributed to the ever-widening divide of learners based on race, socioeconomic status, sex, gender identity, ability, and more, granting even more access and opportunities to those who already had access and opportunity" [6] (p. 163). Because the act of grading causes a detachment from the purpose of the learning experience, students tend to focus on checking off requirements versus internalizing the content and practice [6]. The focus on earning a grade rather than engaging in the learning process can be troublesome for teachers and teacher candidates as they navigate the complexities of the teaching and learning of mathematics. With the need for teachers to focus on student understanding in mathematics classrooms, the preparation and development of teachers should support and provide exposure to assessment practices aligned with fostering understanding.

2. Literature Review

In the following section, we outline research and practices related to humanizing assessment to create an environment of growth and development rather than compliance and hierarchy.

2.1. Standards-Based Assessment

Despite assessment practices being promoted to determine what students know and can do, grading practices have encompassed non-cognitive factors like behavior, dispositions, and compliance. As early as the late 1950s, it was noted that grading practices in elementary schools were based on a mix of academics and non-cognitive factors [7], even with calls from measurement research to be based on achievement. Likewise, and decades later, Brookhart [8] noticed similar assessment practices and believed they were deeply rooted in societal expectations and norms and included factors that were not about learning, growth, and achievement. Because of the comingling of academics and non-cognitive factors, traditional grading practices are unsound in capturing student achievement [9].

Standards-based assessment practices are primed to keep the focus on what students know and can do in alignment with the standards. Students are assessed on specific content standards in terms of mastery or proficiency [10,11]. Standards-based assessment practices communicate precise areas where students need support, allowing students to address the gaps in learning, and provide targeted feedback to students and families [11,12]. Although there is no consensus on the approach to standards-based assessments [13,14], research does indicate three specific criteria that are essential in standards-based assessment practices: (1) grades are based on standards and often include multiple grades in place of a single content grade; (2) performance categories are used to help communicate proficiency; and (3) academic grades are presented separately from non-cognitive grades [14,15].

2.2. (Re)Humanizing Assessment

The humanity of students is at the forefront of rehumanizing mathematics education and mathematics identity development [16]. Addressing power, status, and agency structures and practices within the teaching and learning of mathematics works to humanize students [16]. Ensuring the assessment practices are aligned with a humanizing approach is vital to supporting positive identities and success. "Grades are not good incentive, not good feedback, not good markers of learning, encourage competitiveness over collaboration, don't reflect the idiosyncratic, subjective, often emotional character of learning, and they are not fair" [17] (p. 28). Variations in grading practices that veer away from traditional practice can provide "mirrors" through which teachers view their growth and "windows" [4] (p. 1) through which teachers can gain new views of practice. These mirrors and windows put the focus of assessment on feedback, reflection, and growth and do not lower standards or rigor [18].

Within a more humanized approach to assessment, grading policies should be "responsive to [learners], sustain, and revitalize" [19] (p. 86). By including students in the grading process, the authority shifts away from the instructor and provides a space for feedback that helps teachers examine their teaching practice [19]. Optimal feedback must be valued by the student, provide open dialogue between the student and the teacher, and help to establish trust [20]. Additionally, the purpose of intentional feedback is for students to "develop their capacity to calibrate their own judgements and appreciate the qualities of their work and how it might otherwise be improved" [21] (p. 4). The shared authority increases communication and a shared relationship between students and their instructors and thus cultivates personal responsibility [22].

A rehumanizing approach to grading may require an increased involvement from the instructor and does require more work and flexibility [18]. Teaching philosophies can also be a factor in the implementation of rehumanizing grading approaches [23]. For example, by allowing for flexible deadlines, there are issues since assignments often build upon each other causing difficulty for instructors and students if the students are on varying paths and levels of understanding [18]. "Teachers become designers and sustainers of the learning milieu; establishing conditions in which students can operate with agency" [21] (p. 710). Moreover, student-focused assessment has proven to support student learning in higher education [24].

Blum [6] emerged with her conceptualization of rebuilding and redesigning assessment practices in what she coined "ungrading". She notes that others have called it de-grading or going gradeless. The "ungrading" process changes the cornerstone from grades to learning [6]. As we have worked within this philosophy, we put an emphasis on feedback as the cornerstone of our assessment practices. As such, we refer to our approach as (Up)grading to denote the shared partnership around growth, feedback, and reflection. We believe that it is an elevated assessment approach that (up)lifts students' voices and agency in the process.

3. Theoretical Framework

We entered this work with a lens for equitable and just instructional practices. We centered ourselves in TODOS's essential actions that include eliminating deficit views, eradicating mathematics used as a gatekeeper, engaging in the sociopolitical turn of mathematics education, and elevating the professional learning of mathematics teachers and leaders with a dual focus on mathematics and social justice [25]. Additionally, we draw on the National Council of Teachers of Mathematics' *Catalyzing Change* series key recommendation to "[implement] equitable instructional practices to cultivate students' positive mathematical identities and a strong sense of agency" [26] (p. 45). More specifically to assessment practices, frameworks that have informed our different approaches include labor-based grading [27], self-evaluation using progress processing reflective writing/metacognition (e.g., [17]), practices that promote social justice and equity [4], and diminishing hierarchies and promoting student contributions [28–30].

We draw specifically on the last two assessment frameworks for the purpose of this study and our approach to assessment: practices that promote social justice and equity and seek to increase student involvement within the assessment process. The essential aspect of incorporating (Up)grading practices is a shift in the process and the importance of feedback. Instead of traditional feedback, the instructor is challenged to "feedforward". The practice of feedforward involves providing the learner with statements and questions that provide the potential to advance the nature of their thinking and/or improve the quality of their responses via resubmissions. Hirsch [31] argued that feedforward feedback is about assisting students with repair. Instead of providing ratings and judgment based on past performance, feedforward focuses on development, growth, and additional learning through repairing misconceptions and mislearning [31]. Specifically, there are six attributes to the approach of feedforward: (1) regenerates talent to increase engagement, (2) expands possibilities and supports creativity and opportunities, (3) particular and laser-focused on essentials, (4) authentic and based on practice, (5) provides impactful feedback that is practical and useful, and (6) refines group dynamics [31] (p. 7). We argue that the use of feedforward feedback provides a more equitable and inclusive learning environment, centering student contributions and providing them with more agency and authority in their learning [4,28–30].

4. Purpose of the Study

The purpose of this study was to explore students' lived experience, via reflection, as they participated in a grading experience that was a significant disruption from their traditional experiences with grading and grades. The two research questions that we sought to address were: (1) What perceptions did students have about their (Up)grading experience?, and (2) What tensions did students and MTEs experience as they navigated the (Up)grading experience?

5. Methodology

5.1. Participants and Setting

(Up)grading was implemented in three different courses, across two universities, with a total of 65 teacher candidates and practicing teachers. At one public, doctoral/professional institution located in the Midwest region of the United States, the (Up)grading practice was implemented in two different undergraduate semester-long courses for teacher candidates with differing foci, elementary mathematics methods ($n = 18$) and a statistics and probability course for future teachers ($n = 15$). Additionally, (Up)grading was implemented in a graduate course ($n = 32$) at a large, research-intensive university in the southeast region of the United States. The graduate course consisted of practicing K-12 mathematics teachers and focused on mathematical discourse and high-quality tasks. Each author received approval from their university Intuitional Review Board (IRB-FY2020-695, Pro00079356, and IRB-FY2024-414). The IRB approval for one MTE required student consent forms, which were presented to students during the first week of class. These consent forms were kept secure until final grades were posted; then, the MTE reviewed consent and pulled survey data and reflections for the consenting participants. The IRB for the third author at the same institution required consent after the close of the class and the MTE contacted students for consent. The IRB at the other institution (second author) did not require consent forms. This MTE used anonymous surveys that were sent after final grades of the course were posted.

5.2. (Up)Grading Approaches

Within each of the three courses, an overarching goal was a more humanistic approach that valued student voice in assessment through what we call (Up)grading. Features that were consistent across the three courses were: (a) students were given opportunities to revise their thinking to show an advancement in understanding after receiving targeted feedback from the MTE, and (b) students had a substantial voice in determining their final grade for the course. Despite student work not being "graded" in a traditional sense, students were made aware that there would be a grade attached to their efforts at the end of the course and that they would be the one responsible for providing a grade with supporting evidence to justify that grade. It was communicated early in the semester that the assessment process would be based on labor (i.e., completing assigned work) and evidence of growth in understanding of the key concepts of the course. We sought to focus students' attention on the journey of continuous reflection and revision with the goal of attaining higher levels of academic rigor and greater retention of conceptual ideas. We are learning that context matters in how we approach (Up)grading; as such, each MTE approached (Up)grading in a slightly different way as described below.

Elementary Mathematics Methods. For the elementary mathematics methods course, students set goals at the beginning of the semester, and the (Up)grading process was explained. For smaller assignments, the codes "Got it" or "Missed something" were used. For larger assignments attached to their work in schools, a weighted scale was used to help students see the impact of their work on elementary students and the placement teacher. Students were provided detailed descriptions outlining the expectations of the assignment. Target due dates were used coupled with an expanded window for submission. If the assignment was turned in within this window, assignments could be resubmitted after adhering to the feedback. Weekly, students communicated their participation level, preparation level, and key understandings, and they could ask additional questions within a folder system for communication and documentation. The folder system included a physical folder with reflective prompts each week that supported two-way communication between the MTE and students. Students had weekly entries to document attendance, participation, reflections, and questions. The focus of the feedback was to help students to reflect, rework, and resubmit their work to increase understanding. At the midpoint of the semester, there was a check-in for both the process and student understanding. To culminate the semester, students provided a critical reflection and justification for the overall grade complying with university expectations. Students could refer to their folder for the weekly documentation.

Statistics and Probability Course for Future Teachers. Within the statistics and probability course for future teachers, students were told at the beginning of the semester that they would not be receiving traditional numerical or letter grades on assignments, mini-projects, or assessments. In place of these markings, the instructor would provide targeted feedback to support their efforts to improve their understanding and, when appropriate, a phrasing that would suggest the level of work needed to be done to demonstrate an advancement of understanding. One example of these ratings was "Nailed It", "Almost There", "Not Yet", and "Help". Throughout the semester, the instructor would assign reflective written assignments, asking students to compare their thinking to responses that the instructor believed had "Nailed It". These assignments typically asked students to reflect on what was missing from their response and specifically what they now understand that they did not before. Throughout the semester, it was stressed that there was always an opportunity for growth on any work that was submitted. The challenge for this course is the unfamiliar content and the modeling of high-quality teaching of statistics and probability with the hope that the students will be comfortable implementing these same activities and practices in their own classroom someday.

Mathematical Tasks and Discourse Graduate Course. For the graduate course on discourse and tasks, the focus was allowing teachers the space to apply their learning in their classrooms without fear of "grades" limiting their risk-taking, as they were being asked to push themselves out of their comfort zones with their mathematics teaching. The goal was to focus on growth, not on grades. To do this, students were asked to write critical reflections and rationales for their grades on all assignments. For instance, reading responses were completed weekly; however, "scores" were obtained through three retrospective reflections by the students, while the instructor provided feedback each week. For the major assignment in the course, the students were asked to work through the 5 Practices [32]. The 5 Practices (Smith and Stein, 2018 [32]) provides a model for facilitating productive and high-quality discourse in mathematics classrooms through: (a) setting goals and selecting tasks—Practice 0, (b) anticipating student thinking—Practice 1, (c) monitoring students' work—Practice 2, (d) selecting student strategies to share—Practice 3, (e) sequencing the strategies to be shared—Practice 4, and (f) connecting shared strategies—Practice 5. Assignments were designed to allow teachers time to dig into each practice and implement it within their various classroom settings. For example, for Practice 0, setting goals and selecting tasks [32], the teachers worked to identify the learning goal and develop the task they would ultimately implement with their own K-12 students. The MTE's role was to provide feedback, wherein this was an iterative process where the practicing teacher could revisit and refine the task. The teachers would then "grade" themselves through a critical reflection on their process, focusing on their learning and growth through this phrase of the 5 Practices [32], and ultimately scoring themselves out of 10 points possible with a detailed justification for their score. The same process occurred across each of the major course assignments. While no rubrics were provided, detailed descriptions for assignments were used, and revisions and refinements were encouraged.

5.3. Data Collection

Data focused on the MTEs' initial implementation of (Up)grading and consisted of an open-ended survey given to students seeking perceptions of their (Up)grading experience. Sample open-ended survey questions included: (a) How has the (Up)grading experience shaped your perception of "growth" versus "grades?" (b) Did the (Up)grading process provide you with a voice and ownership in your learning and assessment? If so, how? (c) What aspects of the (Up)grading process were most helpful in advancing your learning? The survey also asked students about how to improve the assessment process. Survey data were collected for all the enrolled students in the elementary methods course and for statistics content course; however, the task and discourse graduate course had a 53% return rate (n = 50 total surveys, n = 18 for elementary methods course, n = 15 for statistics content course, n = 17 for task and discourse graduate course). Additionally, each MTE kept anecdotal and semester-long reflections, including a mid-point check-in. The reflections included open-ended questions to monitor how students were doing with the process and helped monitor student needs along the way. These were often part of exit slips in class. Each MTE did these a little differently, but the intent was to help inform supports and student growth. Each MTE also kept notes of their thinking and changes throughout the semester.

5.4. Data Analysis

Each MTE conducted a case study analysis of their students' responses to the survey. The goal of this analysis was to identify themes in students' lived experiences with (Up)grading, aspects of (Up)grading that student found most helpful, and ways the (Up)grading experience could be improved. After this analysis was completed, a comparative analysis [33] of emerging themes across the three courses for each question was conducted to identify consistent themes across the data. The first step was to code for perceptions and tensions within the survey data, reflections, and the MTE anecdotal notes. To ensure reliability of the study, data obtained were analyzed first by the MTE of record, then checked by another member. Once we analyzed the student perceptions, we organized the perceptions according to the six feedforward attributes as defined by Hirsch (2017) [31] as depicted in Figure 1. Additionally, each MTE coded these perceptions; then, to ensure reliability, another MTE checked the initial coding. Disagreements were decided by the third MTE.

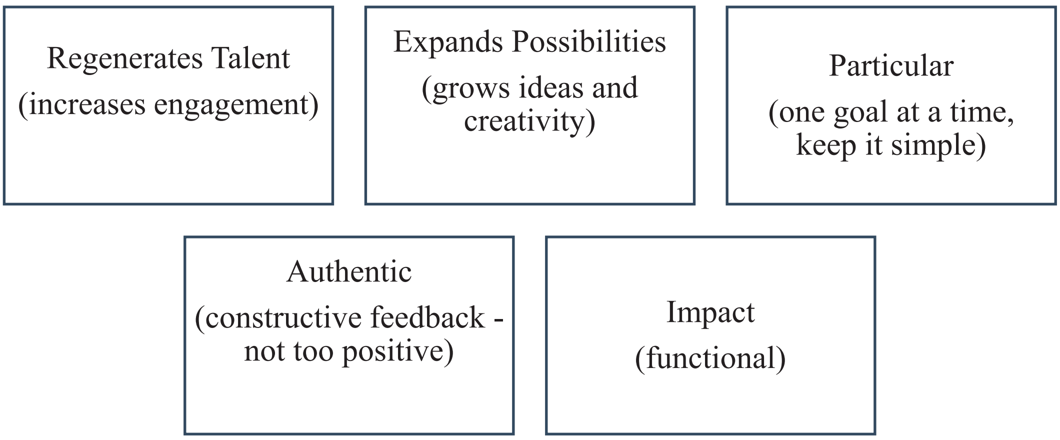


Figure 1. Hirsh's [31] feedforward attributes.

6. Findings

6.1. Student Perceptions of (Up)Grading

Collectively the students' perceptions revealed students had a favorable response to (Up)grading. Their perceptions were rooted in the shift from a focus on grades to a focus on growth in understanding. Three main themes emerged as students reflected on aspects of (Up)grading that were most helpful to them. These included: (a) depressurization of grades and stress, (b) opportunities to self-reflect, and (c) opportunities to learn from mistakes and explore ideas. We expand on each theme in the following paragraphs.

As indicated, one significant theme that emerged across all courses was the depressurization of grades, as evidenced by this student who noted, "Having the pressure of grades released allowed me to take time with the projects and fully learn the material rather than just getting it done for points". Another student noted, "I loved the ungrading experience because I felt like I could focus more on my actual learning and what I wanted to do in the classroom, other than turning in an assignment and worrying about a grade". Other words and phrases students used to describe the depressurization of grades were "liberated", "freedom", "ownership", "reducing fear", "take away stress", and "motivated".

(Up)grading allowed students to take a more holistic approach to their learning, allowing them a multiple of opportunities to self-reflect and have ownership in their growth and subsequent grades. For instance, one student indicated, "This was a foreign experience for me. As a student I am not used to being in such a large degree of ownership for grading. I have owned my learning before, but this layer allowed me to own my learning, growth and results". Additionally, a student expressed, "The upgrading experience definitely helped me to reflect on what I did, what I learned, how I interacted with the course and my peers. I found it to be much more effective because of the reflection piece. In my own teaching, I reflect constantly so I found it to be much more easier and my work was better". These student responses highlight how the (Up)grading assessment approach allowed them to be involved in their learning and grading process, rather than having it done to them or working to simply check something off a list of tasks. Students also noticed that the purpose of the courses and assignments was about them and their growth and not a particular class ranking. A student commented, "It has showed me how it matters more about the content of your work and the understanding behind it than the letter grade result".

Another advantage of (Up)grading for students was the opportunities to learn from mistakes and explore ideas. For instance, a student realized that the work and resubmission process was focused on the opportunity to learn from mistakes or misunderstanding, noting "It has made me see that I am able to keep trying till I master the material, rather than what I get is what I get. I was able to fix my learning and get feedback from the teacher". The idea of expanded opportunities to learn was highlighted by one student who indicated, "The upgrading experience has been enlightening. I focused on my growth and my meeting of expectations much more than the number grade. I felt like I was critical but in doing so I found room for improvement and was able to go back through and improve!". The process allowed for students to make changes without the feeling of stress caused in traditional grading approaches as noted by one student's comment: "I find that the upgrading process really allowed me to look at my work and make changes. I never really think much about grades. There was definitely more reflection on my part". Additionally, students appreciated the time and having more flexibility in their working pace as essential to relieving the stress of worrying about grades.

Students also provided insight into ways to improve the (Up)grading process. For instance, students indicated that additional guidance and time to become comfortable with (Up)grading would be beneficial. One student expressed the idea of a more gradual integration of (Up)grading, saying it may be helpful to "Ease students into the notion of upgrading. Rather than completely removing all grading aspects, leave some structure, maybe some rubrics, to allow students to learn to reflect and become more comfortable". This sentiment was echoed by another student who noted, "The upgrading process requires a shift in thinking. As you introduce it to students for the first time, consider using a combination of traditional teacher-graded activities and ungraded activities. It took me some time to become at ease with the technique". Additionally, students noted that providing more opportunities to reflect on learning, think about the nature and structure of assignments could also help to improve the (Up)grading process. One student provided a suggestion of providing examples to "show what that looks like (an example of a written response)" to support the development of self-reflection.

Connection to Framework. As we worked to offer feedforward feedback, we relied on Hirsh's [31] recommendation with repair as the key purpose of feedback and growth. In analyzing students' perceptions, we organized the perceptions as they related to the six attributes of feedforward feedback. We sought to see if the influence of the attributes were noted in the student reflections (see Table 1 for examples).

Table 1. Examples of feedforward attributes.

Feedforward Attribute	Influence on Students' Examples
Regenerates Talent	I felt like I could focus on my actual learning and what I wanted to do in the classroom, instead of just turning in an assignment and worrying about a grade. I got to focus on what I was doing in the classroom both in class and at my placement school rather than working on assignments.
Expands Possibilities	Growth allows students to make mistakes, explore ideas, get feedback. The lack of a grade provided better motivation.
Particular	I found it very helpful that you were very clear on exactly what my understanding was. You told me what I needed to elaborate on, what I was off, and what I was almost there. In addition, you let me know when I nailed something. The explicit feedback.
Authentic	Being able to redo assignments after getting useful feedback from teachers. To reflect on how I implemented a given task in my classroom. This provided me with the energy to give myself grace when it did not go well but to continue implementing it in order to improve.
Impact	That it was not about a grade, it was about actually learning the information and how we put that information into effect in the classroom. Upgrading made me do more than I probably would have done in the first place. It's easy to phone things in when someone else is grading your work. When you are critiquing yourself, it's hard to lie to yourself.
Refines Group Dynamics	The most helpful aspect of upgrading experience is the collaborative element in it. We can all grow quicker and stronger if we're all trying together. One big aspect that helped me was hearing other students share. I learned just as much from other students' understanding of the concept as I did from you giving us the information.

As we analyzed the data, we realized that these reflective assignments with the support of the feedforward feedback provide a "window" into the students' lived experience that extends far beyond the grade. Often, the instructor has no idea as to each student's journey and learning processes within a given assignment or even course. As we read over the student reflections, we were taken with the impact these experiences have had on not only students' intellectual development but also their social-emotional development. The following reflection from a student encompassed the benefit of the overall process and feedforward feedback:

> "This was an interesting and thought-provoking assignment. It had me answer questions and think in a manner that is the exact opposite of everything I have experienced regarding grades in school. Not only does it make me think critically about what I know, I find it difficult to advocate for myself in any assignment, so this assignment is challenging. In the beginning and still a bit now I feel like I'm giving myself a pat on the back for what I've accomplished, which once again is something I'm not prone to doing. Had I not done this assignment I would have never recognized those things about myself. For that reason, I am already grateful for this assignment, in the sense that it showed me things about myself that clearly need some growth because I should be proud of what I learned and the effort I have put in".

6.2. Student and MTE Tensions around (Up)Grading

Several themes emerged from the analysis of tensions around (Up)grading, including clarity of assignment expectations and criteria, redundancy of reflection, and navigation of a new assessment process. As each MTE navigated the (Up)grading process, the student tensions were similar in that students still wanted the comfort of step-by-step guides to how to assess rather than realizing that, for teachers, a cookie-cutter approach does not work as each grows in their knowledge and instructional practice. For instance, a student suggested, "...maybe show what that looks like as an example of a written response". Additionally, students still wanted to quantify their work. For instance, one student said, "Maybe a rubric that explains...give yourself 10 points if you did "xyz" 9 points if you did "abc". As the focus shifted to the students with an increase in reflection, students found some of the reflection activities to be redundant. An example from a student included this, "there were times where I felt like I was being redundant in my self-reflection".

(Up)grading represented a disruption to the norm of grading that students had experienced throughout their schooling. This disruption brought on a range of emotions from students that required the MTE to empathetically navigate: "my initial thoughts and feelings towards upgrading was that it was unfamiliar. This unfamiliarity brought on a sense of fear and anxiousness as it was an experience I have never encountered. It has been strange and difficult because it is so different because it is so different...however, it is a fabulous concept".

The MTEs also experienced tensions within implementing the (Up)grading within their courses. A significant tension the MTEs faced in implementing (Up)grading was managing the flow of assignments. In traditional grading, there is a linear flow of assignments: students turn in the work, the teacher provides a grade, and the work is returned to the student. However, in (Up)grading, the process is more cyclical in nature, with iterations of MTE feedback and revisions. The cyclical nature of grading can cause the process of coordinating assignments to become more demanding on the MTE because they are providing feedback on the same assignment multiple times. Additionally, balancing hard and soft deadlines to assist students with structure added to the pressures of a constant influx of assignments to provide feedback and guidance. Likewise, figuring out how to use (Up)grading on smaller assignments that support the authentic application assignments (e.g., reading/article reflections) was a difficult process to navigate for the MTEs. It was easier to focus and foster a feedforward focus on the authentic application assignments verses the supporting assignments, perhaps indicating that (Up)grading is not a one-size-fits=all approach. Another tension for the MTEs was shaking themselves from the traditional thinking of point values on assignments. MTEs still wanted to weigh or handle assignments in a way that felt more traditional and comfortable.

7. Discussion

The goal of (Up)grading is to focus students' attention on the journey of learning through continuous reflection and revision of assessment responses with the goal of attaining higher levels of academic rigor and greater retention of conceptual ideas. Previous research has indicated that with traditional grading, students focus on the number or grade and do not attend to the feedback (e.g., [34]). To rehumanize the assessment process, we have found a way to provide feedforward feedback within a climate of (Up)grading, putting students' growth and reflection at the forefront of the learning experience. Specifically, this exploratory study provides insights into students' perceptions of (Up)grading as an assessment approach, students' perceptions of the influence of feedforward feedback as part of the (Up)grading approach, and the tensions felt by both students and MTEs as they experienced and implemented (Up)grading.

Each MTE approached (Up)grading differently; like Brown and Robbins [35] noted, a "one size fits all approach" is not possible when seeking to focus on rehumanizing assessment practices (p. 64). However, looking across the three different cases, insight was gained around specific strategies that students believed best supported their learning, as well as ways to improve the (Up)grading experience. Like Gorichanaz [22], we found that students were able to focus more on the feedback received by the MTEs rather than grades assigned to their work. The assessment approach of (Up)grading provided new ways of learning. Students across all courses were able to see the bigger picture of the assignments in that the courses were benefiting their development as learners and teachers. It is important to note that using a feedforward approach within (Up)grading yielded strong positive perceptions and growth. Each MTE worked to facilitate a positive learning experience with the purpose of growing reflective practitioners [21]. As each MTE worked to provide feedback, monitor the student growth, and optimize the experience, trust was established between the MTEs and their students [20]. We believe that once students experience the benefits of (Up)grading and there is better initial communication by the instructor of features of (Up)grading, the tensions students face as they experience (Up)grading for the first time can be lessened.

Limitations

We recognize that there are limitations in the current study since the main source of data is a single end-of-course survey and the MTE reflections and notes from the student reflections. It is also an important limitation that most of the studies like this one have been with smaller class sizes and not large lecture courses (e.g., [35]). Additional research is needed to understand how to incorporate (Up)grading within those larger lecture courses to make the assessment practices more humanizing.

8. Closing

The detractors of (Up)grading would argue that there is a "fuzziness" around the idea of upgrading. This is not necessarily the case. What it does is position the student as an advocate for themselves and an active participant in the process [21]. Often, grades are based on numerical data that do not consider the students' journey. The often-quoted phrase "It is not the destination but the journey that matters" is appropriate here. More traditional grading practices centered on the destination with students are typically focused on a numerical value (e.g., weighted average) without regard for what has been done to reach that destination. This destination is often a product of grades that are typically assigned by the instructor.

Moreover, we would argue that a feedforward approach could be used in both an (Up)grading and a more traditional grading approach that allows resubmissions. Let us be clear that all grades, even those within our (Up)grading approach, are subjective. It is the instructor who more than likely chooses what is to be assessed, how it is assessed, and how student responses are evaluated. Along the same vein, the student sees the grade that is assigned to those evaluation instruments as a final destination, fixed in time with no opportunity to benefit from reflection. We learn more from our mistakes than we will from what we did correctly. Giving students opportunities to reflect on those mistakes and advance understanding is motivating while, at the same time, it shifts from a fixed destination mindset to a continued growth, journey-driven mindset. In these situations, benefits arise for both the student and the teacher; the student has an opportunity to be rewarded for their intellectual advancement, while the instructor has visible evidence of these advancements in learning. As we work to prepare and support teachers, it is essential that the focus be on helping them grow within their practice and knowledge and giving up the constraints and hierarchy of traditional grading practices. We hope the exposure to student-focused assessment practices will yield more humanizing assessment practices in P-12 schools.

Author Contributions: Conceptualization and methodology, S.D.L., K.E.H. and P.L.S.; project administration, S.D.L., K.E.H. and P.L.S.; writing—original draft preparation, S.D.L. and P.L.S., writing—review and editing, S.D.L., K.E.H. and P.L.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: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Missouri State University (protocol code IRB-FY2020-695, approved on 6 June 2020 and IRB-FY2024-414, approved on 30 May 2024) and the Institutional Review Board of the University of South Carolina (protocol code Pro00079356 approved on 29 March 2019).

Informed Consent Statement: Informed consent was obtained from subjects involved in the study when required by the IRB process. The IRB at one institution did not require consent forms.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

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

References

1. Shapson, S.M.; Smith, N. Transformative Teaching: An Agenda for Faculties of Education. *Educ. Can.* **1999**, *39*, 4–6.
2. National Research Council. *Knowing What Students Know*; National Academies Press: Washington, DC, USA, 2001.
3. Feldman, J. *Grading for Equity: What it is, Why it Matters, and How it Can Transform Schools and Classrooms*; Corwin Press: Thousand Oaks, CA, USA, 2018.
4. Gutiérrez, R.; Goffney, I.; Gutiérrez, R. (Eds.) The Need to Rehumanize Mathematics. In *Annual Perspectives in Mathematics Education: Rehumanizing Mathematics for Black, Indigenous, and Latinx Students*; National Council of Teachers of Mathematics: Reston, VA, USA, 2018.
5. Newton, J.R.; Williams, M.C.; Feeney, D.M. Implementing Non-traditional Assessment Strategies in Teacher Preparation: Opportunities and Challenges. *J. Cult. Values Educ.* **2020**, *3*, 39–51. [CrossRef]
6. Blum, S.D. *Ungrading: Why Rating Students Undermines Learning (and What We Do)*; West Virginia University Press: Morgantown, VA, USA, 2020.
7. Parsons, T. The School Class as a Social System: Some of its Functions in American Society. *Harv. Educ. Rev.* **1959**, *29*, 297–318.
8. Brookhart, S.M. Teachers' Grading: Practice and Theory. *Appl. Meas. Educ.* **1994**, *7*, 279–301. [CrossRef]
9. Brookhart, S.M.; Guskey, T.R.; Bowers, A.J.; McMillan, J.H.; Smith, J.K.; Smith, L.F.; Welsh, M.E. A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Rev. Educ. Res.* **2016**, *86*, 803–848. [CrossRef]
10. Iamarino, D.L. The Benefits of Standards-based Grading: A Critical Evaluation of Modern Grading Practices. *Curr. Issues Educ.* **2014**, *17*, 1–10.
11. Link, L. Teachers' Perceptions of Grading Practices: How Pre-service Training Makes a Difference. *J. Res. Educ.* **2018**, *28*, 62–91.
12. Link, L.J.; Guskey, T.R. Is Standards-based Grading Effective? *Theory Into Pract.* **2022**, *61*, 406–417. [CrossRef]
13. Knight, M.; Cooper, R. Taking on a New Grading System: The Interconnected Effects of Standards-based Grading on Teaching, Learning, Assessment, and Student Behavior. *NASSP Bull.* **2019**, *103*, 65–92. [CrossRef]
14. Welsh, M.E.; Guskey, T.R.; Brookhart, S.M. (Eds.) Standards-based Grading. In *What We Know About Grading*; Association for Supervision and Curriculum Development: Alexandria, VA, USA, 2019; pp. 113–144.
15. Tierney, R.D.; Simon, M.; Charland, J. Being Fair: Teachers' Interpretations of Principles for Standards-based Grading. *Educ. Forum* **2011**, *75*, 210–227. [CrossRef]
16. Morales, H.; DiNapoli, J. Latinx Bilingual Students' Perseverance on a Mathematical Task: A Rehumanizing Perspective. *J. Res. Math. Educ.* **2018**, *7*, 226–250. [CrossRef]
17. Stommel, J.; Blum, S.D. (Eds.) How to Ungrade. In *Ungrading: Why Rating Students Undermines Learning (and What We Do)*; West Virginia University Press: Morgantown, VA, USA, 2020; pp. 25–41.
18. Alex, P. Time to Pull the Plug on Traditional Grading. *Educ. Next* **2022**, *22*, 38–43.
19. Richmond, G.; del Carmen Salazar, M.; Jones, N. Assessment and the Future of Teacher Education. *J. Teach. Educ.* **2019**, *70*, 86–89. [CrossRef]
20. Boud, D. Feedback: Ensuring that it Leads to Enhanced Learning. *Clin. Teach.* **2015**, *12*, 3–7. [CrossRef] [PubMed]
21. Boud, D.; Molloy, E. Rethinking Models of Feedback for Learning: The Challenge of Design. *Assess. Eval. High. Educ.* **2013**, *38*, 698–712. [CrossRef]
22. Gorichanaz, T. "It Made Me Feel Like It Was Okay to Be Wrong": Student Experiences with Ungrading. *Act. Learn. High. Educ.* **2024**, *25*, 67–80. [CrossRef]
23. Campbell, R.; Clark, D.; OShaughnessy, J. Introduction to the Special Issue on Implementing Mastery Grading in the Undergraduate Mathematics Classroom. *Primus* **2020**, *30*, 837–848. [CrossRef]
24. Nieminen, J.H.; Lahdenperä, J. Assessment and Epistemic (In) Justice: How Assessment Produces Knowledge and Knowers. *Teach. High. Educ.* **2024**, *29*, 300–317. [CrossRef]
25. del Rosario Zavala, M.; Andres-Salgarino, M.B.; de Araujo, Z.; Candela, A.G.; Krause, G.; Sylves, E. The Mo(ve)ment to Prioritize Antiracist Mathematics: Planning for This and Every School Year; TODOS: Mathematics for All. 2020. Available online: https://scholarworks.wm.edu/educationpubs/179 (accessed on 15 January 2023).
26. National Council of Teachers of Mathematics (NCTM). *Catalyzing Change in Early Childhood and Elementary Mathematics: Initiating Critical Conversations*; The National Council of Teachers of Mathematics: Reston, VA, USA, 2020.
27. Inoue, A. *Labor-based Grading Contracts: Building Equity and Inclusion in the Compassionate Writing Classroom*; WAC Clearinghouse: Savannah, GA, USA, 2019.

28. Featherstone, H.; Crespo, S.; Jilk, L.; Oslund, J.; Parks, A.; Woods, M. *Smarter Together! Collaboration and Equity in the Elementary Math Classroom*; National Council of Teachers of Mathematics: Reston, VA, USA, 2011.
29. Horn, I. *Strength in Numbers: Collaborative in Secondary Mathematics*; National Council of Teachers of Mathematics: Reston, VA, USA, 2012.
30. Turner, E.; Celedón-Pattichis, S. Problem Solving and Mathematical Discourse Among Latino/a kindergarten Students: An Analysis of Opportunities to Learn. *J. Lat. Educ.* **2011**, *10*, 1–24.
31. Hirsch, J. *The Feedback Fix: Dump the Past, Embrace the Future, and Lead the Way to Change*; Rowman & Littlefield: Lanham, MD, USA, 2017.
32. Smith, M.S.; Stein, M.K. *5 Practices for Orchestrating Productive Mathematics Discussions*, 2nd ed.; The National Council of Teachers of Mathematics: Reston, VA, USA, 2018.
33. Miles, M.B.; Huberman, A.M.; Saldana, J. *Qualitative Data Analysis: A Methods Sourcebook*, 4th ed.; Sage Publications: Thousand Oaks, CA, USA, 2022.
34. Butler, R.; Nisan, M. Effects of No Feedback, Task-related Comments, and Grades on Intrinsic Motivation and Performance. *J. Educ. Psychol.* **1986**, *78*, 210–216. [CrossRef]
35. Brown, R.D.; Robbins, K.R. Developing and Reconceptualizing an Equitable Grading System in Undergraduate Education. *Teach. Educ. J.* **2023**, *16*, 50–70.

education sciences

MDPI

Article

The Use of Guided Reflections in Learning Proof Writing

Kathleen Hoffman [1,*], Tory H. Williams [2] and Kerrie Kephart [3]

[1] Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
[2] Office of Research Protections and Compliance, University of Maryland, Baltimore County, Baltimore, MD 21250, USA; twilliams@umbc.edu
[3] Faculty Development Center, University of Maryland, Baltimore County, Baltimore, MD 21250, USA; kkephart@umbc.edu
* Correspondence: khoffman@umbc.edu

Abstract: We investigated written self-reflections in an undergraduate proof-writing course designed to mitigate the difficulty of a subsequent introductory analysis course. Students wrote weekly self-reflections guided by mechanical, structural, creative, and critical thinking modalities. Our research was guided by three research questions focused on the impact of student self-reflections on student metacognition and performance in the interventional and follow-up class. To address these questions, we categorized the quality of the students' reflections and calculated their average course grades within each category in the proof-writing, the prerequisite, and the introductory analysis courses. The results demonstrated that writing high-quality self-reflections was a statistically significant predictor of earning higher average course grades in the proof-writing course and the analysis course, but not in the prerequisite course. Convergence over the semester of the students' self-evaluations toward an experts' scorings on a modality rubric indicates that students improve in their understanding of the modalities. The repeated writing of guided self-reflections using the framework of the modalities seems to support growth in the students' awareness of their proof-writing abilities.

Keywords: self-reflection; proof writing; real analysis

1. Introduction

In the US, the transition from early university mathematics courses that are calculational in nature to proof-based courses—where students are expected to read, write, and assess proofs—has been a significant modern pedagogical challenge [1]. Even students who were successful in prior calculational courses, such as calculus and differential equations, may struggle when confronted with a proof-based course [2]. Historically and inter-institutionally, an alarmingly high number of students earn non-passing grades in their first proof-based course. At many institutions of higher education, a first course in proof writing aligns with one of two categories: (1) a course with a primary focus and learning objectives that address specific mathematical content, which we describe as a Content-Based Introduction to Proof (CBIP); or (2) a course for which the primary focus is on proof structure and techniques, which we call a Fundamental Introduction to Proof (FIP) [3]. In CBIPs, proofs (and the associated linguistic and logical content) are taught through the lens of other mathematical content, for example, linear algebra, abstract algebra, or real analysis. Alternatively, FIPs teach the fundamentals of proof writing through symbolic logic, sets, relations, and elementary number theory.

In the early 1980s, there was a concerted national movement to address students' difficulty in transitioning to proof-based courses by creating FIP courses [4]. These courses aimed to teach students how to effectively communicate in the language of mathematics and, in particular, how to write formal proofs such as those required in upper-level courses. Today, most collegiate mathematics departments in the U.S. have incorporated some form of an FIP course as a requirement for baccalaureate math programs, although the syllabus and learning objectives of FIPs vary widely [3–5]. Marty [4] addressed the effectiveness of CBIP and FIP courses through a 10-year study at their institution. Longitudinal data, based on grade outcomes in future courses, were compared between student populations who took CBIP-type courses and student populations who took FIP-type courses. The findings, based on a population of approximately 300 students, led the author to suggest that it is more effective to focus on developing students' approaches to mathematical content (as emphasized in FIPs) than to focus on the mathematical content itself. The study concluded that FIPs also increase students' confidence and ability to take ownership in their mathematical maturation.

Developing best practices for teaching proof writing in the mathematical sciences has been the focus of much scrutiny and has proven to be a formidable task [6–14]. Alcock's study [15], based on interviews with mathematicians teaching introductory proof material, addressed the complexity of the thought processes involved in proof writing. The article concludes with the suggestion that the transition to proof courses should address four independent thinking modes: instantiation, structural, creative, and critical modes. The study of Moore [2] looked specifically at the cognitive difficulties for students transitioning to proofs at a large university, concluding that the three main difficulties are (a) foundational concept understanding, (b) mathematical language and notation, and (c) starting a proof. The FIP in our study focuses on skills in (b) and (c) through novel self-reflective writing. Consequently, we identified a previously unaddressed opportunity to directly analyze the effectiveness of transition-to-proof reflective writing strategies.

The present study focused on an FIP course we recently developed to bolster the success rate of students in the subsequent Introduction to Real Analysis class. The latter course is considered difficult by many students and instructors. The modified course sequence begins with (1) a prerequisite introductory linear algebra course, followed by (2) the optional FIP course, predicted to prime student success in (3) the subsequent introductory analysis course. This sequencing allows for a direct comparison of students who choose the path of linear algebra to real analysis and those who choose to take the FIP between these core courses. The newly developed FIP course includes active learning techniques that have been shown to increase the efficacy of teaching mathematical proofs [4]. While building from the thinking modality work of Alcock [15], our approach of using reflective writing as a framework for students to develop self-awareness of their understanding of their proof-writing skills represents a novel interventionist [13] pedagogical approach compared to the traditional pedagogical approaches to transition-to-proof courses.

The goal of the present work was to assess the innovative use of student self-reflective writing-to-learn (WTL) exercises, scaffolded by a modality-based prompt and assessment rubric. Specifically, we developed a modality rubric that consists of a Likert-scale evaluation and field(s) for open-ended response self-evaluations for each of four modalities: mechanical (the mechanical modality corresponds to remembering, which is different from [15] the instantiative modality, which corresponds to a deeper understanding of the definitions), structural, creative, and critical thinking modalities of proof writing. Students were also asked to evaluate their use of the type-setting program LaTeX. However, here, we focus on the conceptual elements of proof writing, that is, the mechanical, structural, creative, and critical modes of thinking (see Appendix B for example modality rubrics and Section 2 for detailed descriptions of the modalities). The framework of these modalities and the modality-based rubric are inspired by prior, related work [8,16,17]. In weekly homework assignments, students were asked to reflect on their performance specifically in the context of their comfort with the modalities. With this structure, we posed three research questions:

- Does having students write reflections in the interventional (FIP) course support their ability to be metacognitive about their own proof-writing processes?
- Does having students write reflections in the interventional course impact their performance and success in that course?

- Does having students write reflections in the interventional course impact their performance in the subsequent Introduction to Real Analysis course?

We addressed these questions over three semesters of the FIP course through analyzing students' weekly self-reflections (both with Likert ratings and open-ended responses) and correlating with students' grades in the prerequisite, interventional, and subsequent introductory analysis courses. As such, we addressed a gap in the research in mathematics education regarding the cognitive processes involved in proof writing [1]. Specifically, we investigated the effectiveness of a novel approach that has students engage in reflective writing in a transition-to-proof course. While there is extensive research on reflective writing and student success, our novel contribution is combining the modality rubric, one of the best practices for teaching proof writing, with reflective writing.

While the focus of this work is on the impact of reflective writing on student achievement in an introductory proof-writing course, there is a related question of the impact of the new course as an intervention relative to a student's overall success. We analyze the impact of the new course in a subsequent paper [18] using data from a pre–post-assessment and analyzing students' grades in the prerequisite course, the interventional course, and the subsequent course for students who took and did not take the interventional course. Preliminary results show a positive effect of the interventional course on student learning and success [18].

The next section lays out the theoretical framework for our study design and methodologies. The Materials and Methods section describes the interventional course, the learning modalities, the instruments used, and the qualitative and quantitative methods used to analyze the student reflections. The Results section includes a discussion of student performance in the progression of courses as a function of the quality of their reflections, an analysis of their growth in metacognition, and an analysis of student reflections relative to the grader's evaluation of student performance. Finally, the Discussion section contains a summary of our results, conclusions, and other considerations.

2. Theoretical Framework and Related Literature

In this section, we synthesize theory and the relevant literature to provide framing for the pedagogical practice of having students engage in reflective writing in mathematics and how and why to provide students with instructional scaffolding to support their reflection on their proof-writing processes. We also trace the origins of our framing of the four thinking modalities involved in proof writing.

2.1. Reflective Writing-to-Learn in Mathematics

Writing stimulates thinking and promotes learning [19–23]. In the process of composing, writers put into words their perceptions of reality. As Fulwiler and Young describe it, "... language provides us with a unique way of knowing and becomes a tool for discovering, for shaping meaning, and for reaching understanding" [22] (p. x). Writing enables us to construct new knowledge by symbolically transforming experience [24] because it involves organizing ideas to formulate a verbal representation of the writer's understanding.

For more than 50 years, research in the movements known as writing-to-learn (WTL) and writing in the disciplines (WID) has shown the connections between writing, thinking, and learning and has demonstrated that writing may contribute to gains in subject area knowledge [25], and specifically in mathematics [26–30]. One strand of theorization around the mechanisms by which writing may lead to gains or changes in understanding in a particular subject area holds that the act of writing spontaneously generates knowledge, without attention to or decisions about any particular thinking tools or operations [31,32]. The writing process commonly referred to as free writing is associated with this strand of theory. Another conceptualization of the role of writing in thinking and learning is the idea that when thinking is explicated in the written word, the writer (and readers) can then examine and evaluate those thoughts, which may allow for the development of deeper understanding [33], as cited in [32]. Pedagogical approaches that follow from this view include reflective and/or guided writing processes. Other theories of the relationship between writing and learning focus on the role of attention to genre conventions [34] and rhetorical goals [35] and how attending to these concerns allows writers to transform their understanding. Together, these theories of the connection between writing and learning, along with empirical studies of the effects of writing-to-learn, argue for the value of having students articulate their thinking about a subject matter through the written word.

Indeed, a recent meta-analysis of more than 50 studies of WTL activities in math, science, and social studies at the K-12 level found that writing in these subject areas "reliably enhanced learning (effect size = 0.30)" [36] (p. 179). Furthermore, Bangert-Drowns et al.'s earlier (2004) [25] meta-analysis of writing-to-learn studies had found that scaffolding students' writing through prompts that guided them to reflect on their "current level of comprehension" (p. 38) of the topic was significantly more effective than other kinds of prompts or than unguided writing.

Such metacognitively focused reflective writing is a mode of WTL that supports learning by engaging the writer in the intentional exploration and reconstruction of knowledge and personal experience in a way that adds meaning [37,38]. In the process of reflection, the learner's own experience and understanding become the focus of their attention. Done well, reflection in writing enables the writer to abstract from, generalize about, and synthesize across experiences [39]. When reflective WTL is supported pedagogically through well-constructed writing prompts and instructor and peer feedback, the writing process may enable student writers to become aware and in control of their own thinking and learning processes, in other words, to become metacognitive and self-regulating [38]. Such an ability to self-regulate their learning processes also contributes to learners' perceptions of greater self-efficacy [40,41]. Thus, guided, reflective WTL activities hold particular promise for helping learners to organize their knowledge, thereby deepening their understanding, gaining more awareness and control over their learning processes, and experiencing a sense of greater self-efficacy.

In mathematics, guided, reflective WTL aims to support learners' making connections to prior knowledge, developing awareness of and improving problem-solving processes (i.e., metacognition), bringing to consciousness areas of confusion or doubt as well as the development of understanding, and/or expressing the learner's feelings and attitudes toward math [42]. A small body of the literature on reflective WTL in college-level mathematics exists, but even fewer studies have aimed to connect reflective WTL with improved student outcomes in mathematics courses. Much of the existing literature on reflective WTL in mathematics focuses on students' feelings and attitudes, specifically helping students to express and alleviate math anxiety [42,43]. In a notable departure, Thropp's controlled study aimed to show benefits from reflective writing for student learning outcomes in a graduate statistics course. Students who participated in reflective journal writing in the interventional statistics course performed significantly better on an assignment and a test than students in the control section who did not engage in reflective journaling [44].

2.2. Scaffolding Reflection on Proof Writing

For reflection in writing to be productive, many students need some support and guidance [25,30,38,45–49]. Guiding questions and other forms of "scaffolding" [50] can help students focus on the elements of their thinking and problem solving that are important for their success in the course. Bangert-Drowns et al.'s meta-analysis [25] found that the metacognitive scaffolding of writing-to-learn activities, such as interventions that had students "reflect on their current understandings, confusions, and learning processes", were associated with gains in student learning. Applying it to proof writing, we posit, therefore, that focusing students' attention on modes of thought associated with constructing and justifying proofs might help them to understand proofs better and become more adept at writing them [6,51]. Another more streamlined approach toward the thinking modes involved in proof writing is Alcock's four modes [15]: instantiation, structural thinking, creative thinking, and critical thinking. Engaging students on these fronts likely requires classroom approaches and techniques beyond standard lecture. Therefore, our study adapted the above approaches to guide students' reflection on their proof writing using four thinking modalities, as outlined below.

2.3. The Modalities for Thinking about Proof Writing

The mechanical modality refers to the precise use of definitions and formal manipulation of symbols. Accurate and precise use of language in definitions is a new skill for students as they begin to write proofs [52,53], and they often find it difficult to understand and apply many of the definitions of advanced mathematics [2,54]. Instantiation, as described by Alcock [15], goes beyond the simple memorization of a definition and includes understanding the definitions to the extent of successfully applying them to an example. Instead of Alcock's instantiation [15], we chose to focus the mechanical modality on the more basic skill of memorization, since this was an introduction to proof writing class. Alcock's instantiation [15] might be better for the real analysis course, so students can build on the mechanical modality as their proof-writing acumen increases. The mechanical modality as used here does not require the level of understanding of Alcock's instantiation [15], and instead refers to memorizing, a very basic learning skill.

The structural modality focuses on viewing the whole proof as a sum of its constituent parts. For instance, standard proof writing begins with stating the hypothesis and ends with stating the conclusion. To prove an if-then statement $(A \rightarrow B)$, one must justify how the hypothesis leads to the correct conclusion. The proof of an if-and-only-if statement $(A \leftrightarrow B)$, however, must include proving both the forward $(A \rightarrow B)$ implication and backward implication $(B \rightarrow A)$. Thus, the proof comprises two pieces. This is similar to "zooming out" as described by Weber and Mejia-Ramos [55] and reminiscent of the idea of the "structured proof" studied by Fuller et al. [56]. The structural modality corresponds to more complex cognitive skills than the mechanical modality since it requires students to see both the whole proof and the parts that contribute to the whole.

The creative modality involves making appropriate connections between concepts to correctly ascertain the crux of the statement/proof. For a simple if-then style proof, the creative modality is the crucial idea that connects the hypothesis to the conclusion. Still, there is no algorithmic method to teaching students how to connect the hypothesis to the conclusion because it will be different for each proof. The creative modality is similar to the "zooming in" strategy proposed by Weber and Mejia-Ramos [55] as "filling in the gaps" or a line-by-line strategy or "key idea" or "technical handle" as suggested by Raman et al. [57]. This modality is more challenging than either the structural or mechanical modalities because it requires creating a logical sequence of statements.

The critical modality involves ascertaining the overall truth or falsity of a statement/proof, thus verifying a sequence of logical steps or producing a viable counterexample [58]. Critically assessing whether a proof is correct, well written, valid, and has the right amount of detail for readers to follow is an advanced skill that results in differences in interpretation even among mathematical experts [7–9,12,57,59–62]. Students find this modality challenging for several reasons. For one thing, they typically read less skeptically than experts. They also tend to trust every line of proof by default if it is written by a mathematician [63]. To verify the validity of a proof, students must think abstractly to decide if the hypotheses and line-by-line details warrant the conclusion [64]. Students struggle with the critical modality because it involves both "zooming in" and "zooming out", as referenced by Weber and Mejia-Ramos [55] (p. 340).

3. Materials and Methods

3.1. Context of Research

3.1.1. FIP Interventional Course Description

We piloted a 4-credit class, Introduction to Mathematical Reasoning, with a maximum capacity of 30 students. The course was offered to students who had not yet taken, or attempted, a CBIP and had previously completed a second-year introductory course in linear algebra with a grade of "C" or better. The textbook we used was Edward R. Scheinerman's (2013) *Mathematics: A Discrete Introduction, 3rd Edition*, Brooks/Cole [65]. Twice a week, students met with the faculty instructor for 75 min for a content-focused lecture and guided examples. A subsequent weekly, 50 min discussion section, led by a teaching assistant (TA), engaged students actively in small groups working on low-risk assessments. Students were expected to complete homework assignments independently each week. Throughout the semester, students took three unit exams and a cumulative final exam with questions aligned to student learning objectives (see Appendix A). Overall, the FIP course format facilitated deliberate practice of proof-writing skills through both traditional (i.e., lecture) and active learning (i.e., group work) methods.

The focal interventional course was developed by one of the authors of this manuscript, some colleagues in the Department of Mathematics and Statistics, as well as pedagogical experts in the Faculty Development Center, with the aim of decoupling the structural foundational difficulties of proof writing from specific mathematical concepts, such as mathematical analysis and abstract algebra. To this end, the students developed proof-writing skills based on topics including basic number theory ideas (e.g., odd, even, prime, and divisibility), sets, logic and truth tables, and relations. The course focused on basic proof techniques such as if-then and if-and-only-if but also introduced more advanced proof techniques, including induction, contradiction, contraposition, and smallest counterexample. The final weeks of the course focused on applying these techniques to set-based function theory using ideas such as injective, surjective, and bijective. We addressed these competencies among others through the eight specific learning objectives listed in Appendix A. The scope of the interventional course was consistent with 81% of FIPs offered at R1 and R2 universities [3]. During this study, this newly designed FIP course was taught by the same instructor (who is an author and the course developer) and two different TAs for three semesters from Fall 2019 through Fall 2020. Data collection for this study was impacted by the COVID-19 pandemic. This course abruptly switched to an online course during the second semester and was taught completely online for the third semester of data collection. We further comment on the impact of the pandemic in the Discussion.

3.1.2. The Role of the Teaching Assistant

Each semester the course was offered, there was one teaching assistant (TA) assigned to the course. Over the three semesters that data were collected for this study, two different individuals served as the teaching assistant for this course: one individual for Fall 2019 and Spring 2020, and a different individual for Fall 2020. The role of the teaching assistant for this course is multifaceted. The TA, as mentioned previously, led a 50 min discussion section each week. In addition, the TA graded the weekly homework assignment, which consisted of the worksheets given during the discussion, questions from the book, and the modality rubric. For the purposes of this study and after grading the discussion worksheet and questions from the book, the TA evaluated each student on a Likert scale for their proficiency in each modality each week. As a TA for the course, the individual had extensive experience in writing proofs and was fluent in the meaning of each of the modalities. The TA was not required to write self-reflections such as the ones that the students wrote, nor did they participate in the part of the research that involved reading and rating the students' reflections.

3.1.3. Participants

Prospective participants for this study included students who completed the interventional FIP course, the subsequent introductory analysis course, Introduction to Real Analysis, and the prerequisite core course, Introduction to Linear Algebra, with a grade of C or better. Prospective participants reviewed an informed consent letter, approved by the Institutional Review Board, before actively joining or declining to join the study. Only students consenting to the study ($N = 36$) had their course data included in research analyses. All participants were enrolled in the interventional FIP course during one of the first three semesters that it was offered at the focal institution, a mid-sized, public, minority-serving research university in the mid-Atlantic region of the United States. According to institutional data, 39% of student participants in this study identified as female and 61% as male. In total, 38% of the participants identified their race/ethnicity as White, while 62% identified as people of color.

3.1.4. Reflection Guided by Thinking Modalities

The development of a new course allows opportunities for innovations in pedagogy to better achieve course goals. The innovation described here involved students in writing self-reflections on their proof-writing processes and abilities, scaffolded by a rubric that defined four thinking modalities of proof writing: mechanical, structural, creative, and critical. The goals of having students write self-reflections included: (1) supporting students' metacognitive awareness of their proof-writing processes and abilities by giving them a framework for identifying specific areas of struggle; (2) assessing students' development in these skills over the semester; and (3) allowing the instructor to evaluate the role of interventional self-reflection in students' learning of proof writing. The four thinking modalities were described to students routinely throughout the entire interventional course, both in lectures and in the discussion sections. Each weekly homework prompted students to evaluate their proof-writing performance through the lens of each modality. Students began this reflective process by evaluating their performance on a Likert scale, followed by writing justifications of their Likert ratings in an open-ended response field (see Appendix B). The Likert ratings served a multifaceted purpose. They (1) corresponded to a modality rubric containing descriptors developed by the instructor to approximate novice-to-mastery level achievement with each modality; (2) encouraged students to carefully consider their achievement-level(s) using each modality, thereby preparing them to write a more meaningful self-evaluation in the open-ended response portion of the reflection; and (3) provided a means for the instructor to quickly compare students' perceptions of their modality skills to that of the grader, a TA considered to have expert-level fluency with the modalities and proof writing.

The cognitive difficulty associated with each of the modalities increased over the semester. Early in the course, the mechanical modality consisted of learning basic definitions such as "even", "odd", "prime", and "composite" (among others). As the semester progressed, students learned new definitions, such as "relation" and "equivalence relation", which were more abstract and challenging for the students to remember and apply. Similarly, the structural modality was concrete for the proof of an "if-then" statement, introduced near the beginning of the semester. However, toward the end of the course, students were asked to construct more complex proofs, such as an "if-and-only-if" statement in conjunction with a hypothesis and conclusion that involved "set equality". The structure of such a proof is more advanced because it comprises four separate parts. Given this increase in complexity, we would not expect linear progress in students' development of these modalities over the course of the semester. The open-ended response justifications and explanations in the written reflections provide insight into the students' thinking and learning processes throughout the course.

3.2. Research Methodology

3.2.1. Instruments

While students worked toward mastery of the learning objectives (Appendix A) through a variety of formative and summative assessments, the focal instrument for this study was a novel reflection tool assigned to students along with each weekly homework assignment. Students' evaluations were guided by a modality rubric with an accompanying Likert scale and field(s) for open-ended response self-evaluations for each modality (see Appendix B). Students were assigned to complete one reflection for each of 12 weekly homework assignments. The TA graded students' weekly homework and evaluated students' performance of each modality using the same Likert scale that students used to self-evaluate. While the TA provided written feedback on weekly homework problems, the TA did not comment on the students' written reflections, and the students were not aware of the TA's Likert scale ratings on the modalities. Students were given credit for completing the weekly modality rubric and reflection.

Over the three-semester study, the self-reflection prompts changed in response to direct feedback from the students and the recognition of a misalignment between the instructor's expectations for open-ended response content versus the observed content submitted by students. The rubric for each semester can be found in Appendix B. In the first semester of the study, some students provided vague or off-point comments in relation to modalities, which provoked re-examination of the instrument's phrasing. For example, the original mechanical modality required "insightful" use of definitions. Recognizing that this was vague, we changed the requirement to a more basic skill of memorizing definitions "precisely". Likewise, student feedback drove a change to the Likert scale range used to self-rate performance. Below, we describe these changes in detail.

During the first semester, the students self-rated their mastery of each thinking modality using a four-point Likert scale, attached to the descriptors superior, proficient, acceptable, and poor (see Appendix B). In addition to these four categories, students could choose "not applicable for this homework". Subsequently, students were prompted to write a short self-reflection on "substantial gaps or significant improvement" with respect to their mastery of the modalities. Students typically focused on only one of the modalities, and their reflective responses lacked specificity. During the first semester of the study, students used a Learning Management System tool to enter both the Likert scale responses and open-ended response reflections/justifications. For the subsequent two semesters of the project, these responses were entered into a Google form.

The student feedback and instructor's evaluation of student responses drove a change in the Likert scale students used to rate their performance during the second and third semesters that the course was offered (see Appendix B). The Likert scale for each modality was expanded to include seven increments, where a rating of "7" indicated confident mastery, and a "1" indicated a significant need for more instructional support. Consistent with the first semester, students were also given the choice of "not applicable for this homework". The explanation for each increment on the Likert scale changed as well. For example, the highest self-rating or TA rating of superior on the mechanical modality was described during the first semester as "Student shows flawless and insightful use of definitions and logical structures, and formal manipulation of symbols". During the second semester of the study, the revised rating of "7" on the Likert scale read "I consistently and correctly use definitions, logical structures, and formal manipulation of symbols". While the first description tacitly asked the students to rate their performance and the language implied an almost unattainable perfection, the revised prompt is more student-centric (such as the use of "I") and explicitly asks students to evaluate their own skills. In an effort to elicit more details from the students, the prompt for the open-ended self-reflection was also revised in the second semester to read "Please comment on why you chose each of the ratings above" in an effort to guide students to comment on each of the four modalities, every week. During the second offering of the course, the instructor explicitly discussed the reflection prompt with students and reinforced expectations with an example of an ideal reflection.

During the third semester of the interventional FIP course, the seven-point Likert scale on which students evaluated their performance with respect to each modality was retained. However, in a further effort to elicit high-quality self-reflections for each of the modalities, the prompt for the written self-reflection instructed students to respond to each modality separately by writing four different self-reflections. In addition, more explicit instructions were given to the students (see Appendix B for details). The prompt for the mechanical modality, for example, became "Using the format claim $\rightarrow$ evidence $\rightarrow$ reasoning, justify your chosen ranking for the Mechanical Modality. Be sure to include specific examples for evidence to support your claim, and carefully describe your reasoning for how your evidence supports your claim".

3.2.2. Qualitative Methods

Three raters (the authors) independently read and rated all the reflections provided by $N = 36$ participants, for a total of $M = 310$ responses, out of a possible 432 responses. (The difference between the total and possible numbers of responses reflects the fact that some participants did not write all 12 reflections during the semester.) The overall quality of each participant's set of reflections was initially rated as one of four levels: exceptional, acceptable, developing, or incomplete. The definitions of each quality level were refined ad hoc as the raters made iterative passes through the data and patterns of completion and specificity emerged from the data, a method adapted from analytic coding techniques as described by Coffey and Atkinson [66]. While the prompts were refined in the second and third offerings of the course, the criteria by which we rated students' responses were consistent over all three semesters of data. Table 1 provides descriptions with which each participant's open-ended response reflections were rated for quality. In order for a student's response profile to be categorized as exceptional, acceptable, developing, or incomplete, criteria had to be satisfied for both the aligned "Completion of Assigned Reflections" and "Adherence to the Prompt" columns of Table 1. For example, if a student's reflections were 75% complete over the semester, yet only 50% of the reflections adhered to the prompt, then the student's response profile would be categorized as Acceptable, not Exceptional. Assessing reflection quality is crucial to addressing our research questions.

Table 1. Parameters for rating the quality level of students' written reflections

Rating	Completion of Assigned Reflections (% over the Semester)	Adherence to the Prompt (% of Reflections Possible)
Exceptional	75–100%	In total, 66–100% specifically or fully addressed the prompt, including explication of the thinking modalities, justifying the self-rating, and exemplifying it using specific elements of the proof in question
Acceptable	≥66%	In total, ≥50% specifically addressed the prompt
Developing	≥50%	In total, ≥25% either vaguely or specifically addressed the prompt; vague reflections are representative
Incomplete	0–49%	In total, 0–24% contained reflections that specifically addressed the prompt; many responses omitted

In developing the rating categories for students' reflections, we considered the percent of reflection assignments completed over the semester. However, the percent of reflections completed could not be the sole definition of quality since several students completed most assignments with little effort or adherence to the prompt. In order to specifically address the prompt, reflections went beyond stating, for example, that the homework was "easy" or "hard"; ideal reflections provided explanations and reasoning for any of the four modalities the student struggled with or felt confident about. Rephrasing the prompt without additional context did not constitute a highly specific reflection, nor did commenting on one's rating of the respective homework assignment. Likewise, several students provided thoughtful, targeted reflections in some instances but neglected to write them or wrote vague ones in other instances. It is important to note that the rating process did not address the correctness of any mathematical claims students made since the prompt did not require them to make mathematical claims. Rather, the focus of the reflection was on their own thinking, with reference to the thinking modalities. Therefore, our ratings of the quality of student reflections is based on the extent to which they showed such metacognition, not on mathematical accuracy.

After independently coding $M = 310$ written reflections, the three raters deliberated and reached consensus, following [67], on appropriate quality ratings for all 36 participants' reflections. The methodological flow diagram in Figure 1 shows how these 310 individual written reflections were categorized throughout the research process.

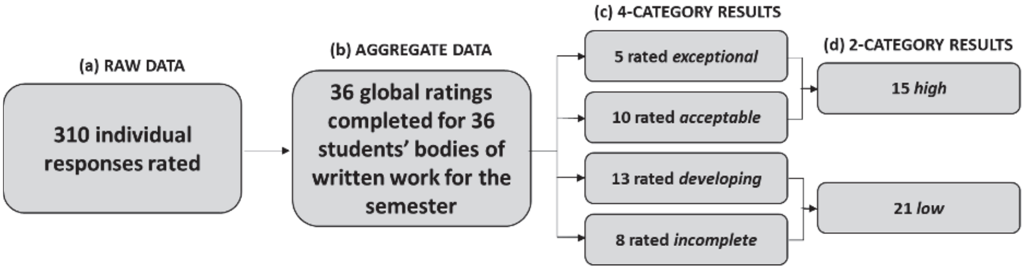


Figure 1. Data analysis flow diagram: This figure shows the data analysis process we conducted and the relationships among (**a**) the raw data, or individual student responses; (**b**) the aggregate data, or the global rating assigned to each student's collection of responses; and the analytical categories these collections of responses first assigned to (**c**) four categories, which were later collapsed into (**d**) two categories.

3.2.3. Quantitative Methods

Quantitative analyses treated written reflection data on a broader level of two groups: low or high quality. Developing and incomplete reflections were grouped together in a "low quality" category. Likewise, acceptable and exceptional data were treated as a group of "high-quality" reflections. We decided to explore the broader categories for two reasons. First, while deliberating and reaching consensus for rating the reflections qualitatively, we often struggled with distinguishing incomplete reflections from developing reflections, and acceptable reflections from exceptional ones. Second, given the small sample size and variance, we expected that creating two larger categories would allow relationships to emerge.

To determine whether a relationship existed between content-specific reflection quality and students' performance in relevant mathematics coursework, quality identifiers for each participant's written reflections were aligned with their respective final grades in the prerequisite course, the interventional course, and the subsequent required core course (on a five-point scale: $A = 4, B = 3, C = 2, D = 1, F$ and $W = 0$). The real analysis course was chosen because it is a required course for all math majors at the focal institution, and students from the FIP course would then enroll in that course. The majority of the grade in each of these courses was determined by in-class exams and a cumulative final exam. Quantitative analyses accompanied qualitative analyses to provide support for emerging patterns. The weekly modality rubric prompted students to rate their use or understanding of the four modalities on a Likert scale (see Appendix B). We calculated the difference between grader and student modality rubric ratings for each modality, per student, per week of the semester. These difference data were used in a Pearson's Correlation test to determine if they shared a significant linear relationship with time in the interventional course (assignment week over the duration of a semester). With this test, we asked, as the weeks passed, do grader–student differences in rubric ratings decrease? A smaller difference in grader–student ratings was intended as a proxy for students' modality use reaching expert level.

3.2.4. Statistical Analyses

Descriptive quantitative statistics were applied to determine whether course grade data sets followed a normal distribution, for which statistical tests were warranted. Both Shapiro–Wilk and Kolmogorov–Smirnov tests for normality revealed that none of these data sets of interest met the criteria for parametric tests, except for tthe modality rubric data. Therefore, non-parametric tests were applied to compare multiple groups of unpaired, non-normal data, while parametric tests were used to analyze the modality rubric data, as described below.

After categorizing students' reflections as one of two quality groups, the corresponding grade data were analyzed with independent samples Mann–Whitney U tests. This test was selected because we wished to compare the same variable (grade earned) across two completely different groups of subjects (those producing "low-" vs. "high-quality" reflections). Similarly, we applied the independent samples Kruskal–Wallis test to compare the course grades of four completely different groups of subjects, those whose set of reflections were rated as (1) exceptional, (2) acceptable, (3) developing, or (4) incomplete quality.

Pearson's Correlation test was applied to the normative modality rubric numeric data to elucidate possible direct or inverse relationships between subjects' course grades and the quality of their written reflections. Likewise, parametric beta linear regression analysis with independent variables of linear algebra grades $(4, 3, 2, 1, 0)$ and "high-quality" written reflections and the dependent variable of the Real Analysis couce grade $(4, 3, 2, 1, 0)$ allowed us to investigate the significance on "beta" for the binary variable; the binary variable was whether students earned a passing grade in the interventional course. In other words, the selected statistical test considered whether some combination of prerequisite course performance, interventional course performance, and/or written reflection quality could reliably predict or explain course performance in the next, rigorous course in sequence.

Data were organized in Microsoft Office Excel, and all statistical tests were performed in IBM SPSS Statistics 26. The study was classified by UMBC's Office of Research Protections and Compliance as exempt (IRB protocol Y20KH13003), 25 July 2019.

4. Results

4.1. Analyses of Reflections for Quality, Metacognition, and Variability Across Cohorts

The three raters independently rated the body of each student's written reflections over the 12 homework assignments to provide a single, holistic rating for each student's work. In determining these ratings, we agreed that typical "exceptional" reflections explicitly addressed the prompt by explicating multiple modalities and consistently provided examples and/or reasoning. Reflections rated "acceptable" often explicitly responded to the prompt, sometimes provided specific examples of areas of struggle or success from the student's proofs, and/or provided some reasoning for the self-rating. A holistic rating of "developing" was assigned if the student neglected to submit some responses and/or the reflections responded to the prompt in nonspecific ways. A rating of "incomplete" was assigned where a student did not complete the reflections or the writing was extremely vague with respect to the prompt. See Table 1. Example responses characteristic of each quality level are shown in Table 2. One participant completed over 50% of reflection opportunities, yet their reflections were rated as "incomplete" (instead of "developing", as suggested by the completion parameter) because the majority of their reflections were a single word, which was insufficient for addressing the prompt. Nearly all participants whose reflections were rated as exceptional completed at least 11 of 12 assignments. However, the reflections of one participant who completed only nine assignments were rated as exceptional due to their overall quality.

As shown in Table 2, students whose writing was rated as "developing" or "incomplete" skipped writing the reflections or wrote them using very vague language. They generally did not describe their struggles or successes and did not refer to the modalities or any specific elements of the proofs they wrote that week. Thus, their reflections did not provide good evidence that they understood the four modalities and could appreciate how to operationalize them in order to write their proofs.

Table 2. Example participant reflections characteristic of four quality levels (one example provided per semester per level).

Reflection Quality	Semester, Participant, Homework Assignment Number (HW) , Excerpt of Written Reflection	Raters' Comments
Exceptional	Semester 2, Student 22, HW 7: "Mechanical: I chose superior because I feel like I was able to know which definitions to use and where to put each definition. [Structural]: I chose proficient because I struggled with knowing how many parts are supposed to go in a proof, especially with the smallest counter example proofs. Creative: I chose superior because I feel like [I] was able to adequately fill in the middle of the proofs, knowing what the beginning and the end should be. Critical: I chose superior because although we didn't have true or false, I was able to know the converse or contrapositive form of the given statements and then prove those statements to see a contradiction properly".	• Accurately explicates each of the four modalities • Provides reasoning for self-ratings • Refers to specific elements of the proofs (e.g., "converse or contrapositive forms of the given statements")
Acceptable	Semester 1, Student 10, HW 11: "I felt substantial gaps in the [structural] and critical modality, which then prevented me from getting to the creative modality portion of the homework"	• Demonstrates understanding of the inter-relationships among modalities • Does not explicate or exemplify any modalities
Developing	Semester 3, Student 23, HW 4: "I gave myself a 7 because I used a lot of logic and had strong arguments"	• Provides vague reasoning • Does not explicate any of the modalities or provide any examples
Incomplete	Semester 3, Student 24, HW 7: "I had issues with the proofs, especially 21"	• Does not refer to the modalities • Refers to but does not reflect on a specific problem • No reflection on what the writer struggled with

4.2. Analysis of Student Growth in Metacognition

In this section, we analyze student growth in metacognition based on a longitudinal analysis of student reflective writing over the semester, as well as a longitudinal comparison of student ratings with the ratings of the grader, whom we consider an expert in this context.

4.2.1. Longitudinal Analysis of Student Reflective Writing

To frame our analysis of metacognition in the students' reflective writing over the semester, we looked for evidence of their awareness of and ability to control their own thinking and learning processes, specifically their use of the four thinking modalities to self-regulate their learning of the proof-writing process, as well as evidence of metacognitive growth in their reflective writing over the semester. Students who were able to articulate their thinking processes through accurate use of the modalities to explain their approaches to the homework in their reflections and/or were able to clearly express the extent of their understanding of how the modalities supported their ability to write the homework proofs were judged as showing evidence of metacognition. All 15 students whose sets of reflections were rated exceptional or acceptable showed evidence of metacognition in one or more of the reflections they wrote.

As an example of metacognition in reflection, we highlight a few of Student 13's reflections, which were selected for their typicality as well as their richness in detail. This student's set of reflections was rated exceptional overall and showed early and consistent evidence of metacognition. In Homework 2, Student 13 wrote:

> Mechanical: I think I did a better job at using symbols than homework 1, but I'm still not sure whether it's greater than, as good, and hopefully not worse than homework 1. . .
>
> (Student 13, Homework 2)

In the first clause of this statement ("I think I did a better job at using symbols than homework 1"), Student 13 shows awareness of improvement in their use of symbols compared to in the previous week's homework. In the second clause ("but I'm still not sure whether it's greater than, as good, and hopefully not worse than homework 1"), they also demonstrate awareness of the limits of their current understanding, mentioning what they struggled with or were still unsure about. Like all students whose reflections were rated exceptional ($N = 5$), Student 13 cited specific elements of that week's proof exercise that showed they understood how the modalities relate to the thinking involved in writing those proofs. See, for example, the rest of Student 13's Homework 2 reflection:

> . . . [Structural]: I think I was able to show off parts of the statement in the truth table in order to fully determine if two statements were equivalent, but there could be something that I could be missing from the tables. Creative: I was able to show some connection between statements in order to determine if statements were logically equivalent, but I feel like I wasn't able to fully explain some of the statements as to why they were either true or false. Critical: I think I was able [to] show how statements were either true or false, and was able to show my thought process with the truth tables.
>
> (Student 13, Homework 2)

In response to the creative modality, Student 13 was able to articulate an awareness of strengths ("I was able to show some connection between statements") but also identify areas of improvement ("I feel like I wasn't able to fully explain some of the statements"). Thus, Student 13 is demonstrating signs of metacognition in their reflections by being able to identify their own strengths and weaknesses in writing the proofs in Homework 2.

While Student 13's set of reflections are representative of writing that was rated exceptional overall, it is noteworthy that only five students out of 36 across the entire data pool wrote reflections that were rated this highly. In total, 9 of the 10 students whose sets of reflections we rated as acceptable wrote some entries that were highly metacognitive and self-critical (39 out of 108 instances of reflection across 12 homeworks for 9 students), but they did not consistently, throughout the entire semester, ground their references to the modalities in actual performance or the particulars of that week's proofs. In total, 8 of 10 students whose set of reflections were rated acceptable showed inconsistent evidence of progress in metacognition over the semester. However, two students whose reflections were rated acceptable showed consistent development in their ability to reflect metacognitively about proof writing over the semester. Student 9's reflections, for example, are representative of such growth, and were selected to highlight here because the contrast between their early and later writings is vivid. Compare Student 9's reflection on Homework 3 with what they wrote for their final reflection on Homework 12:

> I believe that the proofs I provided used adequate detail and included logical connections between statements in this homework.
>
> (Student 9, Homework 3)

> I believe that I have made a lot of improvement on how to approach a proof problem. I think I really built the mechanical and [structural] modality being that I can see a proof and immediately be able to break it down into the multiple subsections in order to solve it in [its] entirety. However, going from one point to the next is where I think I struggle. I have a tendency to set up each part of the proof or create a form of a shell of what needs to be filled in and then I have a hard time filling in some of the gaps. I think the repetition of proofs in class and on the homework have allowed me to improve on doing so, however I think I need a bit more practice on my own over the winter break. . .
>
> (Student 9, Homework 12)

In the earlier reflection on Homework 3, Student 9 wrote in generalities with no explicit reference to the modalities or to their process, whereas by Homework 12, Student 9 articulated strengths and weaknesses in their proof-writing process with explicit reference to the modalities. Thus, while Student 13, classified as writing exceptional reflections, showed signs of metacognition in their responses as early as Homework 2, Student 9, classified as writing acceptable reflections, showed development of metacognition over the semester.

4.2.2. Longitudinal Analysis of Student–Grader Differences

We examined the question of whether grader–student differences in rating decrease over time using Pearson's Correlation test and found that as the semester progressed, student and grader ratings became more similar in the structural and creative modalities ($p < 0.001$, Table 3). A similar trend is evident for both the mechanical and critical modalities, but not significantly so. The negative Pearson Correlation values show that the linear relationship for all four modalities is negative: As time passed, the difference in ratings between graders and students decreased.

Table 3. Correlations between time and the difference between grader and student reflective rubric ratings.

	Mechanical	Structural	Creative	Critical
Pearson Correlation	−0.065	−0.298 **	−0.226 **	−0.043
Sig. (2-tailed)	0.220	0.001	0.001	0.478
N	359	345	344	276

The symbol (∗∗) indicates that the p-value of significance is less than 0.01. Through analyzing student reflections as well as comparing student–grader differences over time, there is evidence of student growth in metacognition over the semester. Below, we will show that students who wrote high-quality reflections demonstrated strong performances in both the interventional course and the advanced course.

4.3. Comparison across Semesters

The proportion of the students classified as writing either high- or low-quality reflections was fairly consistent over the three semesters (see Table 4). Notably, the relative proportions of reflection quality observed in the second iteration of the course, which coincided with an abrupt transition to emergency remote instruction in spring 2020 due to the COVID-19 pandemic, stand out in comparison to the semesters prior and following.

Table 4. Descriptive statistics of participant reflections by quality.

Reflection Quality	n	Number of Completed Reflections (X ± SD)	Proportion of Semester 1 Reflections	Proportion of Semester 2 Reflections	Proportion of Semester 3 Reflections
High	15	11.3 ± 1.3	0.36	0.44	0.41
Low	21	6.7 ± 3.1	0.63	0.55	0.6

In summary, students were grouped by the quality of their reflection responses over the course of the semester. Student reflections showed both evidence of metacognition as well as growth in reflective writing and development of metacognition. Despite changes to the prompt in an effort to elicit better self-reflections and an abrupt transition to online instruction, the proportions of reflection quality remained fairly consistent over the three semesters.

4.4. Analyses of Quantitative Trends

We compared the achievement of participants in the prerequisite course (linear algebra), the interventional course (FIP), and the subsequent course (Introduction to Real Analysis) based on participants' quality of self-reflections in the interventional course. Specifically, we analyzed math course grade data on the basis of reflection quality. We proceeded with two statistical perspectives: (1) comparisons of students' course grades between multiple groups of written reflection quality and (2) beta linear regression analyses to evaluate the possible role(s) of linear algebra grades, FIP grades, and written reflection quality as predictors of course grades in Introduction to Real Analysis.

Course performance diverged between those students who had high-quality (exceptional or acceptable quality) reflections in the interventional course compared to those who had low-quality (developing or incomplete quality) reflections as shown in Figure 2. This difference in performance between the reflection quality groups is more prominent in both the interventional and introductory analysis course grades, as compared to prerequisite core course grades. Figure 2 illustrates that the performance differential between the students who wrote high-quality reflections and the students who wrote low-quality reflections increases with course progression. A non-parametric, individual samples test (Mann–Whitney U) was applied to the data, with reflection quality (high vs. low) as the group qualifier. Statistical results suggest that differences between course grades among those students ($N = 36$) who wrote high- vs. low-quality reflections were significant in the interventional ($p = 0.003$) and introductory analysis courses ($p = 0.007$), and not significant in the prerequisite core course ($p = 0.072$).

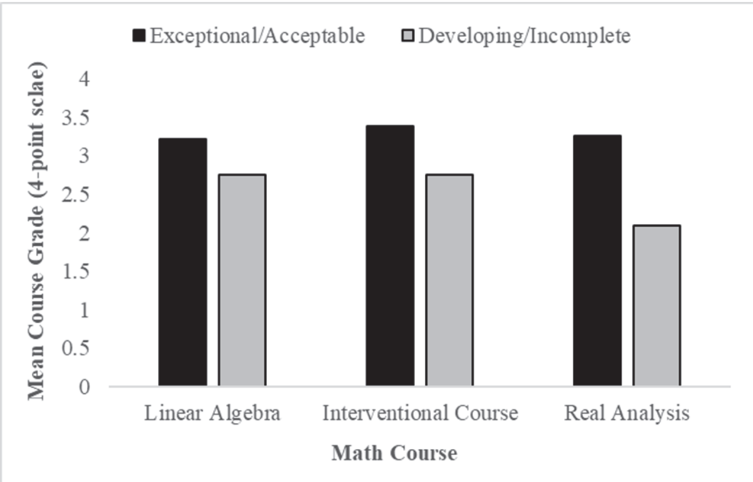


Figure 2. Relationship between average course grades and two categories of reflection quality: We classified each participant's responses as either high quality ("exceptional/acceptable", black bars) or low quality ("developing/incomplete", gray bars); the participants' grades in the prerequisite core course (linear algebra), the interventional (FIP) course, and the introductory analysis course (real analysis) were determined. The course grade results are shown in the bar graph, where a grade of 'A' corresponds to 4, 'B' corresponds to 3, 'C' corresponds to 2, 'D' corresponds to '1', and 0 to 'F' or 'W' (withdraw). While each course shows a higher average grade for the "exceptional/acceptable" participants (black bars), only the grades in the interventional FIP course ($p = 0.003$) and introductory analysis course ($p = 0.007$) are statistically significant.

4.5. Reflection Quality and Prior Student Course Achievement as Predictors of Future Success

A beta linear regression test shows that reflection quality, treated as two category levels, is a significant predictor of the Real Analysis grade ($beta = 1.171$, $p = 0.004$). A separate beta linear regression test shows that the Linear Algebra grade (scale of 0–4) is not a significant predictor of the Real Analysis grade ($beta = 0.290$, $p = 0.312$). A third beta linear regression test shows that the FIP grade is a significant predictor of the Real Analysis grade ($beta = 0.628$, $p = 0.022$), where the beta variable represents whether students passed the interventional course with a C grade or higher. In summary, whether we apply regression analysis or a group comparison approach to the quantitative data, we see similar results. Specifically, reflection quality during the FIP is a significant predictor of future success in the Real Analysis course, while grades in the prerequisite Linear Algebra course are not. Likewise, group comparison results suggest that differences between course grades among those students who wrote high- vs. low-quality reflections were significant in the interventional (FIP) and introductory analysis courses, and not significant in Linear Algebra. Together, these aligned findings reinforce our confidence that having students write reflections throughout the interventional FIP had a positive impact on students' success in both that course and the subsequent introductory analysis course.

5. Discussion

5.1. Summary

Based on best practices in the literature [3–5], we created an interventional FIP course, Introduction to Mathematical Reasoning, to mitigate the difficulty of the CBIP course, Introduction to Real Analysis. The interventional FIP course utilized active learning and the novel approach of self-reflections focused on four learning modalities within mathematical reasoning—mechanical, structural, creative, and critical—inspired by [15]. In this study, we analyzed the quality of self-reflections pertaining to students' proof-writing self-efficacy for evidence of meaningful learning. We first rated each student's reflections in the categories of exceptional, acceptable, developing, or incomplete. We then combined these four categories of student reflections into two groups: high quality, i.e., those meeting exceptional/acceptable criteria, and low quality, i.e., those aligned with developing/incomplete criteria. Both classification systems demonstrated similar quantitative results. Figure 2 illustrates the average letter grade for each of the three sequenced courses based on the two-category classifications, respectively. Specifically, the increased performance gap was statistically significant for the interventional and introductory analysis courses, but not for the prerequisite course. In other words, success in the prerequisite course alone does not predict success in the upper-level CBIP course.

5.2. Addressing the Research Aims

5.2.1. Does Having Students Write Reflections in the Interventional (FIP) Course Support Their Ability to Be Metacognitive about Their Own Proof-Writing Processes?

In answering our first research question, we see evidence that some students are marshaling the concepts behind the four thinking modalities post hoc to write their reflections. Student 13, whose reflections were rated exceptional, showed consistent evidence of metacognition throughout the semester. Student 9's reflections, on the other hand, were rated acceptable, but they also showed growth in their metacognitive abilities over the semester. While we are able to draw conclusions about the metacognitive abilities of students who write high-quality reflections, we are not able to draw any conclusions about students who either did not write reflections every week or whose reflections were vague or nonspecific. However, our study design did not allow us to determine whether or not any students are consciously applying the modalities while they are in the process of writing proofs. A future study that has students think aloud as they engage in proof writing might shed more light on their use of various thinking strategies, including the four modalities, during the process.

5.2.2. Does Having Students Write Reflections in the Interventional Course Impact Their Performance and Success in That Course?

Figure 2 shows that there is a statistically significant relationship between the high-quality (exceptional/acceptable) reflections and students earning higher average grades in the interventional course. Thus, the intervention may have impacted the performance of the students who thoughtfully completed the self-reflections. Specifically, based solely on the average course grade, students who wrote better self-reflections performed better in the interventional course than students who wrote low-quality (developing/incomplete) reflections. Thus, corroborating [25], it appears that merely assigning students to write self-reflections does not necessarily affect their performance and success in the interventional course. However, on average, writing high-quality reflections correlated with higher grades in both the interventional course and subsequent introductory analysis course, thus suggesting a positive impact of the intervention. Students whose reflections qualified as high quality did not necessarily have greater success in the prerequisite course, as evidenced by the lack of significant correlation between writing a high-quality reflection and respective performance in the prerequisite course. However, on average, these students did perform better in both the interventional and subsequent introductory analysis courses.

5.2.3. Does Having Students Write Reflections in the Interventional Course Impact Their Performance in the Subsequent Introduction to Real Analysis Course?

Analysis of modality reflections showed that students who wrote high-quality reflections performed better in the interventional course. Perhaps more surprising and more important is the correlation between high-quality (exceptional/acceptable) reflections and the respective students' performance in Introduction to Real Analysis, where students were not asked to use the modalities. Not only did students writing high-quality reflections reliably achieve higher grades in the introductory analysis course, but also the performance gap between students who wrote high- and low-quality reflections increased relative to the performance gap in the interventional course. Combined with the lack of correlation between writing high-quality reflections and students' grades in the prerequisite course, success in the CBIP course was not predicted solely by performance in the prerequisite course, and performance in the real analysis course seems to have been impacted by thoughtful self-reflections.

5.3. Other Considerations

In an effort to elicit high-quality reflections from students, the prompt was revised twice during the course of the study to clarify expectations and provide more structure for the students' ability to reflect metacognitively [25,45,68,69]. Despite these efforts, the data (Table 4) do not support an obvious correlation between the quality of reflections and our efforts to clarify our expectations with regard to modality self-reflections. Dyment and O'Connell [70] describe several limiting factors to reflective writing that might explain this. For instance, over the course of the three semesters of our study, neither the instructor nor the TA provided personalized feedback on students' individual reflections. The instructor, however, did attempt to provide extra structure to the prompts and also frequently provided examples of each of the modalities during class. However, these efforts did not appear to impact students' reflections. In another attempt to improve student reflections, the instructor also clarified expectations by providing examples of high-quality reflections and conveyed the positive impact of self-reflective writing exercises on course outcomes. Subsequent to this study, to provide additional support to students for reflective writing, we added a short, small-group discussion on the modalities at the beginning of the discussion section. We predicted that students will benefit from this additional, deliberate learning opportunity.

As the semester progressed, students showed increased understanding of the modalities, as evidenced by the decrease in the difference in ratings between graders and students for all four modalities. While the correlation results for the structural and creative modalities were statistically significant, the trends for mechanical and critical modalities did not meet the criteria for statistical significance (Table 3). The correlational findings for the structural and creative modalities show that grader–student differences in rating decrease as the semester progresses, which suggests student conceptual growth in these modalities over the semester. We speculate that the evaluative demands of the critical modality may have been the most challenging for the students to master, even by the end of the term. On the other hand, the mechanical modality is less cognitively demanding than the other modalities. Thus, many students may have found the mechanical modality manageable from very early in the term, leaving little room for growth over time. We posit that the structural and creative modalities are more cognitively challenging than the mechanical modality, but not as challenging as the critical modality. Thus, it is in the structural and creative modalities where we would expect students to make the greatest gains in awareness of their own abilities over the semester.

While our results indicate that students who submit high-quality reflections are more successful proof writers, our research methodology was not designed to prove causation.

One may speculate that the convergence between grader ratings and student self-ratings may partly be attributable to the students adjusting to the TA's expectations. Given that the TA is an expert, the natural conclusion is that the student is becoming better at proof writing. It may be the case that students who are naturally good at writing reflections are also good at writing proofs. Since the prerequisite class is more calculational, students with strong proof-writing abilities would not necessarily stand out. While this may contribute to some of the results, subsequent detailed analysis found in [18] shows that the greatest impact the interventional course had was on the students who received a B or a C in the prerequisite course. Thus, while these alternative explanations are plausible, one may reasonably conclude that the reflective writing and the interventional class positively impact student performance.

We further speculate that our findings were impacted both by the global COVID-19 pandemic, which emerged during the second of the three semesters of this study, and by data-driven changes in instruction. Data analysis from the first semester resulted in changing the self-reflection prompt, as well as the Likert scale for the student rating of the modalities. During the second semester that the course was offered, instruction abruptly switched from in-person to online due to the onset of the COVID-19 pandemic. Students were impacted by external factors such as the lack of reliable internet connections, flagging motivation, and challenges of time management [71], which may have influenced their performance in the course in a variety of ways. We also noted challenges surrounding active learning during discussions that continued into the following semester, despite a variety of adaptations that the instructor tried to improve the situation. The class was redesigned for the third semester to improve outcomes in the online environment. Lectures were recorded, and students were required to submit lecture notes to ensure they had viewed the recording. Student attendance was required at both a synchronous question session and a synchronous discussion section. We do not believe that these challenges and modifications to the course influenced our findings in any significant way. On the contrary, despite the challenges of teaching and learning during a pandemic, our findings suggest the significant impact of high-quality self-reflections on student performance in both the FIB and CBIP courses across both in-person and online instructional environments.

6. Conclusions

Our analysis of students' written reflections structured around thinking modalities of proof writing showed that students who wrote high-quality reflections performed better in both the interventional course and the subsequent introduction to analysis course. Furthermore, student performance in the prerequisite course did not predict student performance in the interventional and introduction to analysis courses. It was not just that diligent students performed well both with reflective writing and in the math content of their courses. If that were the case, we would expect their performance in the prerequisite course to correlate with their categorization as writing high-quality reflections, which it did not. Thus, we conclude that repeated exposure to guided self-reflection using the lens of the modalities supports growth in the students' awareness of their own abilities pertaining to proof writing. In particular, the modalities provided students with a framework for discerning what they understood and what they needed help with to understand. Our results support the potential power of repeated, modality-based self-reflection as a strategy to improve students' ability to write better proofs, and thus impact outcomes in future proof-based courses.

While the current work highlights the impact of reflective writing to support students' metacognition around proof writing, future work could focus on the best way to guide students toward more effective reflective writing, thus improving both their self-awareness and outcomes in future courses. For example, future work could investigate the role of the frequency and intensity of reflective writing activities on developing proof-writing abilities. Additionally, our work did not directly measure how the students used the modalities while writing their proofs. Further research into students' metacognitive processes—specifically think-aloud-type methods that reveal students' thought processes as they write proofs—could yield additional insights into common misconceptions and misconstruals that would help instructors tailor their lectures, activities, and assignments accordingly. Lastly, supporting instructors and TAs in understanding theory and practice around reflective writing could enable them to better help students develop reflective writing skills.

Author Contributions: Conceptualization , K.K. and K.H.; methodology, K.K. and K.H.; software, T.H.W.; validation, T.H.W. formal analysis, K.H., T.H.W., and K.K.; investigation, K.K. and K.H.; resources, K.H.; data curation, T.H.W.; writing—original draft preparation, T.H.W., K.K., and K.H.; writing—review and editing, T.H.W., K.K., and K.H.; visualization, T.H.W.; supervision, K.H.; project administration, K.H.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded internally by a Hrabowski Innovation Fund grant at UMBC.

Institutional Review Board Statement: This study was classified by the University of Maryland Baltimore County's Office of Research Protections and Compliance as exempt (IRB protocol Y20KH13003), 25 July 2019.

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

Data Availability Statement: Deidentified, aggregate data may be shared upon individual request to comply with the study's IRB protocol.

Acknowledgments: The authors thank Kalman Nanes and JustinWebster for their input in the design and implementation of the interventional course, as well as many subsequent discussions. Additionally, special thanks go to Linda Hodges and Jennifer Harrison in the Faculty Development Center who provided crucial insight into the design of the interventional course and corresponding assessment methods. We are indebted to Linda Hodges and Keith Weber who provided very helpful feedback on the draft manuscript. We are also grateful to Jeremy Rubin and Teddy Weinberg, teaching assistants for the course, who were essential to the data collection for this study.

Conflicts of Interest: The authors declare no 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.

Abbreviations

The following abbreviations are used in this manuscript:

FIP	Fundamental Introduction to Proof;
CBIP	Content-Based Introduction to Proof;
WTL	Writing-to-learn;
WID	Writing in the disciplines;
SLO	Student learning objective.

Appendix A. Student Learning Objectives

Table A1. Student learning objectives (SLOs) for the FIP Course.

SLO 1	Construct basic proofs of if-then statements about integers and sets.
SLO 2	Evaluate the truth or falsity of given statements; defend this decision by providing justifications or counterexamples as appropriate.
SLO 3	Manipulate and negate simple and compound mathematical statements using propositional logic and truth tables.
SLO 4	Quantify (and negate) precise mathematical statements with proficiency in mathematical statements and propositions.
SLO 5	Utilize common proof techniques such as induction, proof by contraposition, and proof by contradiction; recognize the need for these strategies in given problems.
SLO 6	Apply skills of mathematical reasoning, as listed above, to topics including functions, probability, number theory, and group theory.
SLO 7	Evaluate the validity of a given mathematical argument.
SLO 8	Demonstrate correct and precise use of mathematical language.

Appendix B. Modality Rubrics

Appendix B.1. Fall 2019

Table A2. Modality rubric for Semester 1.

	Superior	Proficient	Acceptable	Poor	No applicable
Mechanical	Student shows flawless and insightful use of definitions and logical structures, and formal manipulation of symbols	Student shows appropriate use of definitions and logical structures, and formal manipulation of symbols, with only small mistakes.	Student has gaps in their use of definitions and logical structures, and in formal manipulation of symbols, but still largely achieves associated goals.	Student's poor use of definitions, logical structures, and formal manipulation of symbols prevents them from achieving goals.	This assignment does not make significant use of mechanical factors.
Structural [1]	Student demonstrates flawless and insightful ability to view the whole statement/proof in terms of the comprising parts.	Student demonstrates the ability to view the whole statement/proof in terms of the comprising parts, with only minor errors.	Student demonstrates gaps in their ability to view the whole statement/proof in terms of the comprising parts, but still largely achieves associated goals.	Student demonstrates inability to view the whole statement/proof in terms of the comprising parts, preventing them from achieving goals.	This assignment does not make significant use of instantiative factors.
Creative	Student makes insightful connections between concepts to correctly ascertain the crux of the statement/proof.	With only minor errors, student makes appropriate connections between concepts to correctly ascertain the crux of the statement/proof.	Student shows gaps in their connections between concepts. They ascertain the crux of the statement/proof with difficulty or sometimes miss the point.	Student largely does not make connections between concepts. They cannot ascertain the crux of the statement/proof or cannot complete proofs because of this inability.	This assignment does not make significant use of creative factors.
Critical	Student demonstrates superior insight in ascertaining the truth or falsity of the statement/proof, and in verifying a sequence of logical steps or producing a viable counterexample, as appropriate.	With only minor errors, student demonstrates proficiency in ascertaining the truth or falsity of the statement/proof, and in verifying a sequence of logical steps or producing a viable counterexample, as appropriate.	Student demonstrates gaps in their ability to ascertain the truth or falsity of the statement/proof. Student struggles to verify a sequence of logical steps or to produce a viable counterexample, as appropriate, but does largely achieve associated goals.	Student largely cannot ascertain the truth or falsity of the statement/proof. Student's inability to verify a sequence of logical steps or to produce a viable counterexample, as appropriate, is a significant impediment to their achievement.	This assignment does not make significant use of critical factors.
Latex	Student's LaTeX typesetting adds significantly to the quality of the written work.	Student's work is typeset in LaTeX proficiently and with only minor errors. Typesetting does not detract from the flow of the reading.	Student's LaTeX typesetting contains significant errors, and detracts significantly from the reader's experience.	Student's LaTeX typesetting is poor enough to make reading the work difficult or impossible. Or, student did not typeset the submission using LaTeX despite a requirement to do so.	This assignment was not to be typeset using LaTeX.

[1] Although students were presented with a category labeled instantiative, what was actually described to the students in the scope of this study is commonly referred to as structural by literature.

Please reflect on your performance on homework 1. Please comment on substantial gaps or significant improvements on your performance with respect to the modalities.

Appendix B.2. Spring 2020

Table A3. Mechanical modality rubric for Semester 2.

Self-Rating	Description of Performance
7	I consistently and correctly use definitions, logical structures, and formal manipulation of symbols.
6	Between a 5 and 7.
5	I appropriately use definitions, logical structures, and formal manipulation of symbols, with small or occasional mistakes.
4	Between a 3 and 5.
3	I have some ability to use definitions, logical structures, and formal manipulation of symbols. I could use more support/feedback to meet associated goals.
2	Between a 1 and 3.
1	I need more exposure to the definitions, logical structures, and in formal manipulation of symbols. I need more practice developing my skills in this area.
0	This assignment does not make significant use of mechanical factors

Table A4. Structural modality rubric for Semester 2.

Self-Rating	Description of Performance
7	I consistently and correctly view the whole statement/proof in terms of the comprising parts and/or ideas.
6	Between a 5 and 7.
5	I appropriately view the whole statement/proof in terms of the comprising parts and/or ideas, with only minor or occasional errors.
4	Between a 3 and 5.
3	I have some ability to view the whole statement/proof in terms of the comprising parts and/or ideas. I could use more support/feedback to meet associated goals.
2	Between a 1 and 3.
1	I need more exposure to the whole statement/proof in terms of the comprising parts and/or ideas. I need more practice developing my skills in this area.
0	This assignment does not make significant use of structural factors

Table A5. Creative modality rubric for Semester 2.

Self-Rating	Description of Performance
7	I consistently and correctly make connections between concepts. Or, I correctly approach the crux of the statement/proof in a unique or novel way.
6	Between a 5 and 7.
5	I appropriately ascertain the truth or falsity of the statement/proof and can recognize a correct and complete proof or counterexample, with only minor or occasional errors.
4	Between a 3 and 5.
3	I have some ability to ascertain the truth or falsity of the statement/proof. It is still difficult to recognize a correct and complete proof or counterexample, and I could use more support/feedback.

Table A5. *Cont.*

Self-Rating	Description of Performance
2	Between a 1 and 3.
1	I need more exposure to verifying the truth or falsity of the statement/proof and need more practice recognizing correct and complete proofs or counterexamples.
0	This assignment does not make significant use of mechanical factors

Table A6. Critical modality rubric for Semester 2.

Self-Rating	Description of Performance
7	I consistently and correctly ascertain the truth or falsity of the statement/proof, and can recognize a correct and complete proof or counterexample.
6	Between a 5 and 7.
5	I appropriately ascertain the truth or falsity of the statement/proof and can recognize a correct and complete proof or counterexample, with only minor or occasional errors.
4	Between a 3 and 5.
3	I have some ability to ascertain the truth or falsity of the statement/proof. It is still difficult to recognize a correct and complete proof or counterexample, and I could use more support/feedback.
2	Between a 1 and 3.
1	I need more exposure to verifying the truth or falsity of the statement/proof and need more practice recognizing correct and complete proofs or counterexamples.
0	This assignment does not make significant use of mechanical factors

Table A7. Latex for Semester 2.

Self-Rating	Description of Performance
7	I consistently and correctly use LaTeX to clearly and aesthetically format typeset.
6	Between a 5 and 7.
5	I typeset in LaTeX appropriately, with only minor or occasional errors.
4	Between a 3 and 5.
3	I have some ability to use LaTeX typesetting and I could use more support/feedback. My LaTeX contains errors that might detract significantly from the reader's experience.
2	Between a 1 and 3.
1	I need more exposure to and practice with LaTeX typesetting. It is difficult to read my work, or, I did not typeset the submission using LaTeX.
0	This assignment does not make significant use of mechanical factors

Please comment on why you chose each of the ratings above.

Appendix B.3. Fall 2020

These rubrics used the same Likert scale as Spring 2020, except for a change to the prompt for commenting on why the student chose their self-ratings. After the ratings for each modality, students were prompted as follows:

> Using the format "claim → evidence → reasoning", justify your chosen ranking for [this modality]. Be sure to include specific examples for evidence to support your claim, and carefully describe your reasoning for how your evidence supports your claim.

References

1. Stylianides, G.J.; Stylianides, A.J. Research based interventions in the area of proof: The past, the present, and the future. *Educ. Stud. Math.* **2017**, *96*, 119–127. [CrossRef]
2. Moore, R.C. Making the transition to formal proof. *Educ. Stud. Math.* **1994**, *27*, 249–266. [CrossRef]

3. David, E.J.; Zazkis, D. Characterizing introduction to proof courses: A survey of US R1 and R2 course syllabi. *Int. J. Math. Educ. Sci. Technol.* **2020**, *51*, 388–404. [CrossRef]
4. Marty, R.H. Getting to Eureka!: Higher Order Reasoning in Math. *Coll. Teach.* **1991**, *39*, 3–6. [CrossRef]
5. Savic, M. Does Content Matter in an Introduction-to-Proof Course? *J. Humanist. Math.* **2017**, *7*, 149–160. [CrossRef]
6. Alcock, L.; Hodds, M.; Roy, S.; Inglis, M. Investigating and improving undergraduate proof comprehension. *Not. AMS* **2015**, *62*, 742–752. [CrossRef]
7. Davies, B.; Alcock, L.; Jones, I. What do mathematicians mean by proof? A comparative-judgement study of students' and mathematicians' views. *J. Math. Behav.* **2021**, *61*, 100824. [CrossRef]
8. Inglis, M.; Alcock, L. Expert and novice approaches to reading mathematical proofs. *J. Res. Math. Educ.* **2012**, *43*, 358–390. [CrossRef]
9. Inglis, M.; Mejia-Ramos, J.P.; Weber, K.; Alcock, L. On mathematicians' different standards when evaluating elementary proofs. *Top. Cogn. Sci.* **2013**, *5*, 270–282. [CrossRef]
10. Ko, Y.Y.; Knuth, E.J. Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *J. Math. Behav.* **2013**, *32*, 20–35. [CrossRef]
11. Lee, K. Students' proof schemes for mathematical proving and disproving of propositions. *J. Math. Behav.* **2016**, *41*, 26–44. [CrossRef]
12. Lew, K.; Mejía Ramos, J.P. Linguistic conventions of mathematical proof writing across pedagogical contexts. *Educ. Stud. Math.* **2020**, *103*, 43–62. [CrossRef]
13. Melhuish, K.; Fukawa-Connelly, T.; Dawkins, P.C.; Woods, C.; Weber, K. Collegiate mathematics teaching in proof-based courses: What we now know and what we have yet to learn. *J. Math. Behav.* **2022**, *67*, 100986. [CrossRef]
14. Weber, K. How mathematicians determine if an argument is a valid proof. *J. Res. Math. Educ.* **2008**, *39*, 431–459. [CrossRef]
15. Alcock, L. Mathematicians' perspectives on the teaching and learning of proof. *Res. Coll. Math. Educ.* **2010**, *VII*, 63–92.
16. Bleiler-Baxter, S.K.; Pair, J.D. Engaging students in roles of proof. *J. Math. Behav.* **2017**, *47*, 16–34. [CrossRef]
17. Hodds, M.; Alcock, L.; Inglis, M. Self-explanation training improves proof comprehension. *J. Res. Math. Educ.* **2014**, *45*, 62–101. [CrossRef]
18. Hoffman, K.; Williams, T.; Webster, J.T.; Harrison, J.M.; Nanes, K. Assessing the Impacts of An Interventional Proof-Writing Course. *submitted*.
19. Britton, J.; Burgess, T.; Martin, N.; McLeod, A.; Rosen, H. The Development of Writing Abilities (11–18). *Rev. Brittonism* **1975**, *9*, 7–16.
20. Bruner, J.S. The ontogenesis of speech acts. *J. Child Lang.* **1975**, *2*, 1–19. [CrossRef]
21. Emig, J. Writing as a Mode of Learning. *Coll. Compos. Commun.* **1977**, *28*, 122–128. [CrossRef]
22. Fulwiler, T.; Young, A. *Language Connections: Writing and Reading Across the Curriculum*; National Council of Teachers of English: Urbana, IL, USA, 1982.
23. Langer, J.A.; Applebee, A.N. Learning to Write: Learning to Think. *Educ. Horizons* **1985**, *64*, 36–38.
24. Boland, S. How I started using writing across the curriculum and ended up taking algebra again: A review of useful works on writing across the curriculum. *WAC Ourna* **1989**, *8*, 111–122. [CrossRef]
25. Bangert-Drowns, R.L.; Hurley, M.M.; Wilkinson, B. The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Rev. Educ. Res.* **2004**, *74*, 29–58. [CrossRef]
26. Cross, D.I. Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. *Int. J. Sci. Math. Educ.* **2009**, *7*, 905–930. [CrossRef]
27. Greer, A.R. Mathematical Communication: A Study of the Impact Expository Writing in the Mathematics Curriculum Has on Student Achievement. Ph.D. Thesis, Capella University, Minneapolis, MN, USA, 2010.
28. Idris, N. Enhancing students' understanding in calculus trough writing. *Int. Electron. J. Math. Educ.* **2009**, *4*, 36–55. [CrossRef]
29. Kenney, R.; Shoffner, M.; Norris, D. Reflecting to learn mathematics: Supporting pre-service teachers' pedagogical content knowledge with reflection on writing prompts in mathematics education. *Reflective Pract.* **2013**, *14*, 787–800. [CrossRef]
30. Menz, P.; Xin, C. Making Students' metacognitive knowledge visible through reflective writing in a mathematics-for-teachers course. *Collect. Essays Learn. Teach.* **2016**, *9*, 155–166. [CrossRef]
31. Britton, J. Writing-to-learn and learning to write. In *Selected Essays of James Britton*; Boynton/Cook Publishers: Portsmouth, NH, USA, 1982; pp. 94–111.
32. Klein, P.D. Reopening inquiry into cognitive processes in writing-to-learn. *Educ. Psychol. Rev.* **1999**, *11*, 203–270. [CrossRef]
33. Young, R.; Sullivan, P. Why write? A reconsideration. In *Essays on Classical Rhetoric and Modern Discourse*; Southern Illinois Press: Carbondale, IL, USA, 1984; pp. 215–225.
34. Bereiter, C.; Scardamalia, M. *The Psychology of Written Composition*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1987.
35. Flower, L.; Hayes, J.R. The dynamics of composing: Making plans and juggling constraints. In *Cognitive Processes in Writing*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1980; pp. 31–50. [CrossRef]
36. Graham, S.; Kiuhara, S.A.; MacKay, M. The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Rev. Educ. Res.* **2020**, *90*, 179–226. [CrossRef]
37. Dewey, J. *How We Think*; D. C. Heath and Co.: Boston, MA, USA, 1910.

38. Silver, N. Reflective pedagogies and the metacognitive turn in college teaching. In *Using Reflection and Metacognition to Improve Student Learning*; Stylus Publishing: Sterling, VA, USA, 2013; pp. 1–17.
39. Di Stefano, G.; Gino, F.; Pisano, G.P.; Staats, B. *Learning by Thinking: Overcoming the Bias for Action Through Reflection*; Harvard Business School: Cambridge, MA, USA, 2015.
40. Moon, J.A. *Learning Journals: A Handbook for Reflective Practice and Professional Development*; Routledge: New York, NY, USA, 2006.
41. Yancey, K. *A Rhetoric of Reflection*; University Press of Colorado: Denver, CO, USA, 2016.
42. Borasi, R.; Rose, B.J. Journal writing and mathematics instruction. *Educ. Stud. Math.* **1989**, *20*, 347–365. [CrossRef]
43. Guce, I.K. Investigating College Students' Views on Mathematics Learning through Reflective Journal Writing. *Int. J. Eval. Res. Educ.* **2017**, *6*, 38–44. [CrossRef]
44. Thropp, J. Using reflective journaling to promote achievement in graduate statistics coursework. *Int. J. Learn. Teach. Educ. Res.* **2017**, *16*, 120–134.
45. Dunlap, J.C. Using guided reflective journaling activities to capture students' changing perceptions. *TechTrends* **2006**, *50*, 20–26. [CrossRef]
46. Grossman, R. Structures for facilitating student reflection. *Coll. Teach.* **2009**, *57*, 15–22. [CrossRef]
47. McDonald, C.R. Toward Defining a Social Reflective Pedagogy for ePortfolios. In *A Rhetoric of Reflection*; University Press of Colorado: Denver, CO, USA, 2016; pp. 203–226.
48. Sommers, J. Problematizing Reflection. In *A Rhetoric of Reflection*; University Press of Colorado: Denver, CO, USA, 2016, pp. 271–287.
49. Steven, D.; Cooper, J. *Journal Keeping: How to Use Reflective Writing for Learning, Teaching, Professional Insight and Positive Change*; Routledge: London, UK, 2009.
50. Vygotsky, L.S.; Cole, M. *Mind in Society: Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978.
51. Samkoff, A.; Weber, K. Lessons learned from an instructional intervention on proof comprehension. *J. Math. Behav.* **2015**, *39*, 28–50. [CrossRef]
52. Edwards, B.S.; Ward, M.B. Surprises from mathematics education research: Student (mis) use of mathematical definitions. *Am. Math. Mon.* **2004**, *111*, 411–424. [CrossRef]
53. Vinner, S. The role of definitions in the teaching and learning of mathematics. In *Advanced Mathematical Thinking*; Springer: Berlin/Heidelberg, Germany, 1991; pp. 65–81.
54. Dubinsky, E.; Elterman, F.; Gong, C. The student's construction of quantification. *Learn. Math.* **1988**, *8*, 44–51.
55. Weber, K.; Mejia-Ramos, J.P. Why and how mathematicians read proofs: An exploratory study. *Educ. Stud. Math.* **2011**, *76*, 329–344. [CrossRef]
56. Fuller, E.; Weber, K.; Mejia-Ramos, J.P.; Rhoads, K.; Samkoff, A. Comprehending structured proofs. *J. Int. De Estud. Em Educ. MatemÁTica* **2014**, *7*, 84–102.
57. Raman, M.; Sandefur, J.; Birky, G.; Campbell, C.; Somers, K. Is that a proof? Using video to teach and learn how to prove at the university level. In *Proof and Proving in Mathematics Education: The 19th ICMI Study*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2, pp. 154–159.
58. Selden, A.; Selden, J. Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *J. Res. Math. Educ.* **2003**, *34*, 4–36. [CrossRef]
59. Evans, T.; Mejía-Ramos, J.P.; Inglis, M. Do mathematicians and undergraduates agree about explanation quality? *Educ. Stud. Math.* **2022**, *111*, 445–467. [CrossRef]
60. Mejia-Ramos, J.P.; Weber, K. Why and how mathematicians read proofs: Further evidence from a survey study. *Educ. Stud. Math.* **2014**, *85*, 161–173. [CrossRef]
61. Stylianides, A.J.; Komatsu, K.; Weber, K.; Stylianides, G.J. Teaching and learning authentic mathematics: The case of proving. In *Handbook of Cognitive Mathematics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 727–761.
62. Weber, K. Mathematics majors' perceptions of conviction, validity, and proof. *Math. Think. Learn.* **2010**, *12*, 306–336. [CrossRef]
63. Weber, K.; Mejia-Ramos, J.P. Mathematics majors' beliefs about proof reading. *Int. J. Math. Educ. Sci. Technol.* **2014**, *45*, 89–103. [CrossRef]
64. Alcock, L.; Weber, K. Proof validation in real analysis: Inferring and checking warrants. *J. Math. Behav.* **2005**, *24*, 125–134. [CrossRef]
65. Scheinerman, E.A. *Mathematics: A Discrete Introduction*; Brooks/Cole: Pacific Grove, CA, USA, 2012.
66. Coffey, A.; Atkinson, P. *Making Sense of Qualitative Data: Complementary Research Strategies*; Sage Publications, Inc.: Thousand Oaks, CA, USA, 1996.
67. Saldaña, J. The coding manual for qualitative researchers. *Am. J. Qual. Res.* **2021**, *6*, 232–237.
68. Harris, M. Scaffolding reflective journal writing–Negotiating power, play and position. *Nurse Educ. Today* **2008**, *28*, 314–326. [CrossRef]
69. Power, J.B. Has this begun to change the way they think? Moving undergraduate learners' level of reflection from where it is to where it needs to be. *Teach. High. Educ.* **2016**, *21*, 235–248. [CrossRef]
70. Dyment, J.E.; O'connell, T.S. The quality of reflection in student journals: A review of limiting and enabling factors. *Innov. High. Educ.* **2010**, *35*, 233–244. [CrossRef]

71. Almendingen, K.; Morseth, M.S.; Gjølstad, E.; Brevik, A.; Tørris, C. Student's experiences with online teaching following COVID-19 lockdown: A mixed methods explorative study. *PLoS ONE* **2021**, *16*, e0250378. [CrossRef]

education sciences

MDPI

Article

Concrete–Representational–Abstract (CRA) Instructional Approach in an Algebra I Inclusion Class: Knowledge Retention Versus Students' Perception

Sherri K. Prosser [1,*] and Stephen F. Bismarck [2]

[1] Department of Educational Specialties, Austin Peay State University, Clarksville, TN 37044, USA
[2] Department of Education, University of South Carolina Upstate, Spartanburg, SC 29303, USA; sbismar@uscupstate.edu
* Correspondence: prossers@apsu.edu; Tel.: +1-386-314-3015

Abstract: Mathematical manipulatives and the concrete–representational–abstract (CRA) instructional approach are common in elementary classrooms, but their use declines significantly by high school. This paper describes a mixed methods study focused on knowledge retention and perceptions of students in a high school Algebra I inclusion class after a lesson on square roots using a novel algebra manipulative. Twenty-five students in a high school Algebra I inclusion class engaged in an interactive lesson on square roots paired with the manipulative to support their conceptual understanding. Participants completed a pretest, a post-treatment questionnaire, and a delayed post-test. The two-sample *t* test showed a significant difference in students' pretest–post-test scores. However, conventional content analysis of the questionnaires showed that most students did not believe the CRA instructional approach supported their learning. Implications include increased use of manipulatives to teach abstract algebraic topics to support students' conceptual understanding and destigmatizing the use of manipulatives in secondary mathematics classrooms.

Keywords: algebra manipulative; inclusion class; concrete–representational–abstract approach; high school algebra; secondary mathematics; student perceptions

1. Introduction

"Our goal must be to develop the talents of all to their fullest. Attaining that goal requires that we expect and assist all students to work to the limits of their capabilities" [1] (p. 12).

The quotation above is taken from the seminal report *A Nation at Risk* and still applies regarding the need for educators to implement instructional practices that assist all students in reaching their fullest potential [1]. As of 2017–2018, the most recent data available, approximately 14% of all public school children in the United States are educated under the Individuals with Disabilities Act [2]. Students with a specific learning disability (i.e., learning disability or LD) constituted the largest percentage of students with disabilities, at 33.6% [2]. The call for equitable treatment of and access for all students has echoed throughout the years in mathematics education reform [3–5].

When examining high school mathematics classrooms that include students with LDs, mathematics teachers focus on providing accommodations to students with LDs that range from additional time to complete assignments to alternative homework assignments and to the use of calculators [6]. These accommodations can be attributed to interventions that target procedural mimicry and recall of facts [7]. A growing body of research studies examining the effectiveness of using hands-on and interactive materials with high school students with LDs provide consistent support for instructional practices that go beyond the accommodations and interventions referenced above [8,9]. Within the mathematics education community, hands-on and interactive materials often include what are referred to as "manipulatives".

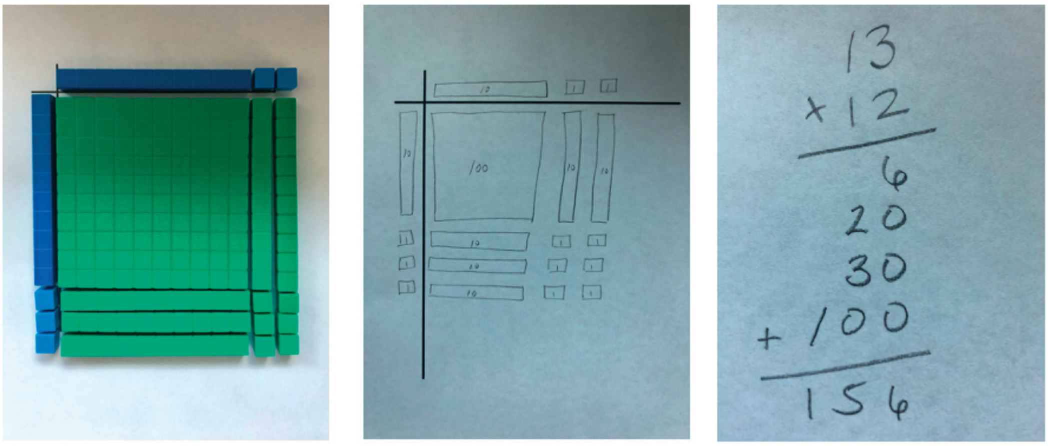
Manipulatives are objects (either virtual or concrete) used to represent abstract mathematical ideas concretely [10]. The conceptual grounding for using manipulatives originates from aspects of constructivist theory that connect students' concrete perceptions and experiences of the world and abstract thinking [11]. The constructivist perspective provides a promising approach as students with LDs have difficulty generalizing learned material and conceptualizing abstract algebraic concepts and tasks [7,12,13]. Manipulatives are concrete objects that students can arrange, partition, and group in ways that assist them in abstract thinking associated with specific mathematical concepts [14].

With the increased availability of electronic devices (e.g., computers, tablets, and interactive whiteboards) in classrooms, the use of virtual manipulatives has also increased. Virtual manipulatives are digital representations of concrete objects. Studies have compared the effectiveness of concrete manipulatives to virtual manipulatives and found them to be equally effective [15–17]. Specifically, Westenskow and Moyer-Packenham [17] examined the use of both concrete and virtual manipulatives for students with mathematical learning disabilities and found that both types of manipulatives provided evidence of statistically significant knowledge gain on a variety of fraction concepts; concrete manipulatives were favored approximately half of the time, and virtual manipulatives favored the other half.

However, it should be noted that using manipulatives does not elicit the automatic learning of mathematical concepts. Ball's [18] article on using mathematical manipulatives with elementary school students puts this notion of mathematical understanding in perspective: "[U]nderstanding does not travel through the fingertips and up the arm. Although concrete materials can offer students context and tools for making sense of the content, mathematical ideas really do not reside in cardboard and plastic materials" (p. 47). Manipulatives need to be implemented in the classroom using appropriate instructional practices. After a 72 h rigorous professional development program specific to a constructivist approach that included manipulatives, for example, teachers of mathematics expressed a belief that students learn abstract topics best when engaged with hands-on activities but noted classroom management (e.g., distractions due to manipulatives) as a barrier to implementation of this approach [19].

One instructional approach that uses concrete models, such as manipulatives, is the concrete–representational–abstract (CRA) approach, which is sometimes referred to as the graduated instructional sequence or the concrete–pictorial–abstract approach. The three parts of the CRA instructional approach build upon each other. The CRA sequence begins with students using the manipulative as they work on a task (i.e., the concrete stage). Once students master the concrete stage, they can create a pictorial display of a completed task with the manipulative (i.e., the representational stage). In the last stage (i.e., the abstract stage), students use numerical or algebraic symbols to facilitate abstract reasoning [20]. For example, students could use base-ten blocks to display a multidigit multiplication problem at the concrete stage, a drawing of the base-ten blocks to comprise the representational stage, and Arabic numerals and symbols to show the same problem during the abstract stage (see Figure 1).

In the early 1980s, Singapore's Ministry of Education began advocating for the implementation of the CRA instructional approach [21]. The strategy has become a staple of teaching mathematics in Singapore and is the grounding behind the country's success on several international mathematics achievement assessments [22]. The CRA sequence has been used while teaching with manipulatives for decades in the United States and has been shown to be an effective approach in mathematics classrooms that include students with LDs [9,23]. Unfortunately, using manipulatives and, therefore, the CRA approach has not been implemented regularly in high school mathematics classrooms [24], likely due to the misconception of the effectiveness of manipulatives with older students [25] and that manipulatives use is distracting [19]. Implementing instructional practices that assist all students includes examining manipulatives that target high school concepts through concrete and visual representations using the CRA approach. Therefore, this article will examine the implementation of the CRA approach with an Algebra I inclusion class, focusing on students' knowledge retention and perception of their learning.



(**a**) (**b**) (**c**)

Figure 1. A multidigit multiplication problem, 12 × 13, is displayed using each stage of the CRA approach: (**a**) Panel 1 shows base-ten blocks, which is the concrete stage; (**b**) Panel 2 shows a diagram, which is the representational stage; and (**c**) Panel 3 shows symbols that model the problem, which is the abstract stage.

2. Perspectives on Manipulatives Use and Need for the Study

A common view among researchers is that instructional practices involving manipulatives are beneficial for young students but unnecessary for older students [25,26]. Using manipulatives in the mathematics classroom has typically been associated with a child's stage of cognitive development [27]. Cognitive development theories suggest that there are developmental stages in children's ability to think and learn as they age [28–30]. For this paper, we focus on the progression from concrete to formal or abstract operations [30,31]. In this progression, children go from using concrete materials to assist them in learning mathematics to the ability to more fully rationalize abstraction while learning mathematics [32]. As a result of this developmental-based progression, hands-on materials and manipulatives are predominantly used to help students learn mathematics in elementary grades [25,26].

Results from a survey conducted by Swan and Marshall [24] reported that using such hands-on materials is almost entirely abandoned by the time students reach high school. Even in middle schools, for example, only 17% of mathematics teachers use manipulatives *frequently* or *very frequently* [33]. It is important to note that cognitive development theory suggests that adolescents develop the ability to more fully rationalize abstraction, but not that concrete materials hinder them in the development of knowledge [34]. The overgeneralization of this theory has contributed to a view of children's cognitive development as inflexible and dichotomous [34].

Research suggests that nearly two-thirds of 17-year-olds (i.e., high school juniors and seniors) have yet to move past the concrete level of thinking and will struggle to formulate abstract thought patterns using purely symbolic representations and making generalizations with little context [32,35]. Furthermore, national groups involved in mathematics education advocate for all students to be actively engaged in the learning of mathematics at all grade levels [4,5,36]: "Students at all grade levels can benefit from the use of physical and virtual manipulative materials to provide visual models of a range of mathematical ideas" [5] (p. 82).

Limited research exists on implementing the CRA approach in high school mathematics courses (e.g., algebra, geometry, trigonometry, and calculus). Bouck and Park [37] conducted a systematic literature review of studies between 1975 and 2017 on using manipulatives to support students with LDs; 29 articles explored using manipulatives through the CRA approach. Of those 29 articles, 5 were identified as targeting secondary mathematics students [13,38–41], 4 of which found that all students increased their mathematical knowledge, skills, or both. Maccini and Hughes [39] found that five of the six student participants improved on all mathematical tasks. This small set of studies presents a promising approach to teaching mathematics to high school students with LDs and requires further investigation.

The low number of studies on high school students' manipulatives use may also be a function of the small number of manipulatives available that target abstract concepts typically taught in the high school mathematics curriculum. Linking concrete materials to abstract representations has presented a significant challenge for educational research [42]. Inherent in this challenge is the dual-representation hypothesis. The dual-representation hypothesis occurs when "symbols are simultaneously objects in their own right and representations of something else" [42] (p. 156). For example, when using algebra tiles, the variable x is represented by the length of a rectangle. The fact that the rectangle has a fixed length may make it harder for the student to focus on the representation of the length, as x is typically presented as an unknown value. This dual representation presents a need to examine and evaluate different manipulatives that target high school concepts through concrete and visual representations while minimizing unnecessary complexities and dual representations.

Indeed, one response-to-intervention recommendation for elementary and middle school students is using concrete manipulatives when visual representations are insufficient for student understanding of the abstract [43]. In fact, the systematic use of manipulatives and visual representations in 13 randomized controlled trials showed moderate evidence for improving students' conceptual understanding [43]. However, others have argued for the appropriateness of manipulative use when introducing new mathematics topics to enhance conceptual understanding and problem solving [44–46]. After an updated review of mathematics interventions for secondary students with learning disabilities, Maccini et al. [46] called for future studies to be conducted within general education classrooms to enhance generalizability and to address middle or high school mathematics topics instead of remedial topics. The focus of this mixed methods study is to determine the knowledge acquisition and retention of knowledge by students after a CRA lesson about an abstract algebra concept. The following research questions guided this study:

1. To what extent will students in a high school Algebra I inclusion class retain the knowledge of simplifying square roots after being instructed using the CRA instructional approach?
2. How will students in a high school Algebra I inclusion class describe the effectiveness of using a mathematical manipulative to learn about a specific mathematical procedure?

3. Materials and Methods

The data highlighted in this article were from a larger study involving 4 teachers and 212 students within 10 college preparation mathematics classes (five Algebra I and five Geometry) at a suburban high school in the southeastern United States. The larger study was a pretest-delayed post-test control group experimental design with three randomly selected algebra classes and three randomly selected geometry classes receiving instruction using the CRA approach (treatment). The remaining two algebra classes and two geometry classes received traditional, didactic instruction (control). Each student in the treatment group received an anonymous open-response questionnaire to complete after the treatment lesson. Approximately one month after the lesson, the post-test was administered to students in both the control and treatment groups.

During the analysis of the pretest–post-test data, the second author discovered that one of the Algebra I classes from the treatment group showed a statistically significant difference that was much greater than any of the other classes. For example, the analysis of two Algebra I treatment classes produced p values of 0.017 and 0.0031, but the other class had a p value less than 0.0001. The researcher contacted the teacher of record to discuss the outlier findings and was informed that the class was an inclusion class. Because this information was not known prior to the study, there was no control group specific to the inclusion class. The statistically significant results in the larger study led to this deeper investigation into the qualitative responses of the students in the inclusion class. This type of retrospective analysis has been referred to as "unmotivated looking"—a term coined by Sacks [47] to describe qualitative analysis when the results are expected to have practical application [48]. For example, educational researchers have applied unmotivated looking to (re)analyze transcripts after unexpected but salient findings in their initial analysis [49].

This study applied a complementarity mixed methods design, which "seeks elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method" [50] (p. 259). Unlike the more commonly used triangulation designs, in which two methods assess the same aspect of a phenomenon, qualitative and quantitative methods in a complementarity design assess different but overlapping facets of a phenomenon [50].

3.1. Context and Participants

The mid-sized high school enrolled 911 students; 53% were deemed proficient in mathematics as measured by state standardized test scores. The school receives Title I funding, with 50% of the population considered economically marginalized [51]. The student racial/ethnic identification is 78% White, 12% African American, 5% Hispanic, and 5% Asian or American Indian.

College preparation and honors were the only two levels of Algebra I classes offered at this school. The class of interest for this article, the Algebra I inclusion class, comprised 25 students, of which at least 40% were identified as having a high-incidence disability (e.g., specific learning disability, speech impairment, language impairment, or other health impairment). The teacher of record was not at liberty to disclose students' specific diagnoses.

3.2. Lesson

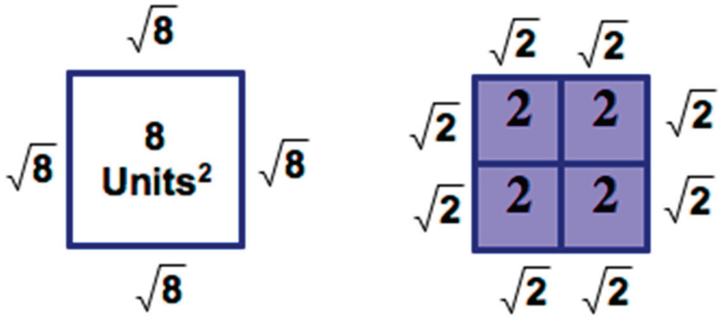
During the lesson, the second author facilitated the entire lesson and took field notes while students worked in small groups. The teacher of record was present in the classroom for the entirety of the lesson. The teacher sat in the back of the classroom, away from the students, and observed the lesson.

The students were taught one lesson on simplifying square roots by the researcher (not the teacher of record) over one 90 min class block using the CRA instructional approach. The lesson began with a 10 min discussion about square roots and their connection to side lengths of squares and ended with introducing the students to the geometric definition of a square root, the length of the side of a square with a given area. Students were placed in pairs, pre-assigned by the teacher of record, and provided with instructions regarding the manipulative and the activity sheet. Students were then instructed to begin working on the first task. The researcher acted as a facilitator, monitoring and answering technical questions regarding the manipulative. After each task, the researcher asked groups to present their solutions and provided time for whole group discussion and connections before moving on to the next task. The first three tasks focused on developing conceptual understanding through the use of the manipulative (i.e., concrete stage), the next two tasks transitioned to drawing representations of the manipulative (i.e., representational stage), and the remaining three tasks provided opportunities to connect the concrete and representational stages to the numerical representation (i.e., abstract stage). This was performed by completing a table, making conjectures, and analyzing the numerical progression, which is traditionally associated with simplifying square roots. Neither group of students received additional instruction on the concept.

The Manipulative

As both concrete and virtual manipulatives have been found to be equally effective [17], the second author in this study used a concrete version of the manipulative; a virtual version is being developed. The manipulative, which was piloted and refined prior to the larger study [52], has since been cited in multiple state standards documents [53–55] and has been used in mathematics professional development grant projects [56]. To initiate the CRA approach, the second author created concrete (i.e., tactile) square manipulatives by printing, laminating, and cutting out squares with whole number areas ranging from 2 cm^2 through 10 cm^2.

Students can select square tiles of a particular size and arrange them in an array to create squares of larger areas, such as those shown in Figure 2. This process allows students to physically model finding perfect square factors rather than watching the teacher model the process. The second author created an activity sheet that provided students with different-sized squares to partition using the manipulative. This activity sheet provided students with opportunities to think and write about the partitions, generalize, and connect the visual representation to the corresponding numerical process.



(**a**) (**b**)

Figure 2. These diagrams are representations that show the equivalence of (**a**) $\sqrt{8}$ to (**b**) $2\sqrt{2}$.

The manipulative minimizes dual representation because the representation of the manipulatives (squares) is directly related to the task (simplifying square roots). The manipulatives are represented as squares of specific whole number areas (in cm). It is critical that all of the squares represent accurate measurements for this model to work. To complete the task, students do not need to interpret these squares in any other context, size, or representation. In fact, it is essential that students see the manipulatives as squares with given areas that correspond to their side lengths. The properties of the square manipulatives are important to their use and, therefore, do not present dual representation. Minimizing dual representation with this type of design has been hypothesized to facilitate transfer and retention [57].

3.3. Instrumentation

Two paper-and-pencil instruments were used for this study: a pretest–post-test and an anonymous open-response questionnaire. The pretest–post-test consisted of three procedural items and two relational items [58]. The procedural items asked students to simplify a given square root. The relational items focused on student explanation, the connection between perfect square factors, and the simplification process. The first relational item asked students to explain why a given square root could not be simplified. The second relational item asked students to identify an error in the simplification process, explain the error, and correct the error. The total score of the assessment was five points (one point for each item). All the questions targeted knowledge needed when simplifying square roots and not knowledge specific to the manipulative or activity.

The questionnaire consisted of two items, each followed by several lines for students to write their responses. The first question asked the students how the activity was helpful to them and to provide support and explanation for their response. The second question asked students whether they thought the activity would help them remember how to simplify square roots and why.

3.4. Procedure

3.4.1. Data Collection

Data were collected using the pretest, questionnaire, and delayed post-test. The five-question pretest was administered by the teacher of record the day before the intervention lesson, and the post-test was administered approximately one month after the lesson had been taught. This length of time between the lesson and the post-test was chosen to measure knowledge retention [59]. All 25 students completed both the pretest and post-test. The open-response questionnaire was provided to the students after the lesson by the teacher of record. Students were given approximately 10 min to complete the questionnaire, and all 25 students responded to each question.

3.4.2. Quantitative Data Analysis

Pretest–post-test scores for each treatment group were analyzed to determine if students (N = 25) showed a significant score increase using matched pair t tests. Normality was confirmed for the group using histograms and Q-Q plots in the computer program SAS. The t test examined if there was a significant gain from the pretest to the post-test, making the test one-sided. The alpha value used to determine significance was 0.01.

3.4.3. Qualitative Data Analysis

The open-response questionnaire was analyzed using conventional content analysis [60]. Conventional content analysis is often used when the goal of a study is to describe a phenomenon where limited information is available related to the phenomenon under study, and open-ended responses are available to form a basis for theory regarding the phenomenon rather than moving toward an existing theory [60]. Analysis, therefore, began by creating word clouds for each set of responses to look for emergent trends in the data that might assist with coding schemes. Word clouds have been found to be a useful tool in determining trends and coding schemes for qualitative data [61,62]. Categories were generated from the data, therefore, in a manner consistent with inductive analysis [60].

Using steps outlined by Hsieh and Shannon [60], the responses for both open-ended items were read, and any words or phrases that suggested a reaction related to the perception or efficacy of the manipulative were highlighted. The highlighted portions were then reviewed to identify common ideas among the highlighted portions. The common ideas were reviewed to identify, categorize, and label themes. Next, the responses were reviewed to bring related categories together and ensure there are categories for response data that do not fit the previously established categories. Lastly, the final list of categories was evaluated and arranged into a hierarchical structure based on how often responses occurred within the category.

4. Results

4.1. Knowledge Acquisition and Retention

A total of 112 treatment students had both pre- and post-test scores (see Table 1). The matched pair t tests indicated significant differences in all three Algebra 1 classes and in the Geometry A and C classes. Inspection of the worksheets from the Geometry B class determined that three students had multiplication errors on the procedural items, which caused their post-test scores to be lower than their pretest scores.

To examine the effect size for each class, Cohen's d was calculated. The effect size for the Geometry B class provided evidence of a small effect (0.202). The effect sizes for the Algebra A, Algebra B, Geometry A, and Geometry C classes provided evidence of a medium effect (ranging from 0.526 to 0.842). The Algebra C (inclusion) class provided evidence of a very large effect (1.378).

The results support that instruction using the CRA approach with a mathematical manipulative had a statistically significant increase in knowledge retention. The quantitative analysis provides evidence of significant knowledge gain after using the manipulative. These results support the hypothesis that knowledge transfer and retention can be facilitated by a manipulative that minimizes dual representation [57]. These results also support discussions that hands-on experiences assist students with LDs in their understanding of how numerical and abstract concepts operate at a concrete level [20,63,64].

Table 1. Matched pair *t* test and Cohen's d for the pretest and delayed post-test (N = 95).

Class	*n*	*M* diff.	*SD*	*t*	*p*	Cohen's *d*
Algebra A	22	−1.272	1.9623	−3.04	0.0031	0.701
Algebra B	22	−0.9773	2.0206	−2.27	0.017	0.526
Algebra C (Inclusion)	25	−1.88	1.8044	−5.21	<0.0001	1.378
Geometry A	10	−0.8	1.2517	−2.02	0.037	0.842
Geometry B	18	−0.5278	1.48	−1.51	0.0743	0.202
Geometry C	15	−0.8333	1.5079	−2.14	0.025	0.744

4.2. Perception of Manipulative Effectiveness

The initial word cloud coding process indicated that students' responses focused on the value of the manipulative (e.g., helped, easier, understand, and better) and the attributes of the manipulative/activity (e.g., visual, way, and squares). Three simple categories were defined to examine how the students perceived the effectiveness of the lesson in developing an understanding of simplifying square roots: positive, neutral, and negative responses. The first category was identified as positive responses and included clear indications that the student found the manipulative helpful, enjoyable, easy, or simple were coded as having a positive response. The second category was identified as neutral responses and included indications that the student found the manipulative somewhat useful, that there was some confusion, or that the student needed more practice. The third category was identified as a negative response and included indications that the student did not find the manipulative useful or that they found the activity/method overly confusing.

Among all six treatment classes (*n* = 115), the majority of students (80%) made positive comments, whereas the remaining students were split among neutral (11.3%) and negative comments (8.7%). See Table 2 for complete data.

Table 2. Student perceptions of manipulative effectiveness (N = 115).

Class	Positive		Neutral		Negative	
	n	%	*n*	%	*n*	%
Algebra A	14	87.5	2	12.5	0	0
Algebra B	20	87	2	8.7	1	4.3
Algebra C (Inclusion)	9	36	8	32	8	32
Geometry A	13	100	0	0	0	0
Geometry B	17	94.4	1	5.6	0	0
Geometry C	19	95	0	0	1	5

Within the inclusion class, however, nine students (36%) had responses coded as positive, eight students (32%) had responses coded as neutral, and eight students (32%) had responses coded as negative. The responses coded as positive included phrases such as "visual understanding" and "hands-on", but there were fewer positive responses than the other responses. The majority (64%) of the students in the inclusion class were coded as having either neutral or negative perceptions of the lesson. Further analysis of the neutral and negative responses was conducted to find sub-themes. The predominant sub-themes from these responses were a lack of confidence and an aversion to being challenged. The responses coded as a lack of confidence showed a negative outlook on learning and mathematics. One student responded with "I don't remember nothing", while another responded with "Honestly it made me confused. I didn't really understand it, so show it to the honors class, they'll understand". The responses coded as having an aversion to being challenged all referenced being confused or that the process was too long and had too many steps. One student responded, "No because it's too many steps", while another responded, "Not really. It takes time for me to know the steps and the process. Factors and square roots make it more confusing for me with all the steps".

The qualitative analysis provides some insight into students' perceptions of how this manipulative was used to deliver effective instruction. The Algebra I inclusion class showed mixed perceptions of the effectiveness of using a mathematical manipulative. The first level of coding found that the responses were almost evenly distributed among the three themes. The sub-themes provided insight into why the students had either a neutral or negative response to using the manipulatives. The manipulative challenged the students to see connections between the geometric representation and a numerical process for simplifying square roots (i.e., identifying perfect square factors). The researchers believe that the students were viewing the lesson in a direct instruction context, typical in mathematics instruction, where everything that is performed in the lesson is modeled by the teacher and then reproduced by the student, step by step. It seems that it was unclear to these students that using the manipulative was to gain a better understanding of simplifying square roots rather than to replicate the entire activity on their own.

When it comes to students' development of knowledge, the majority of the students stated that they were either unsure or did not believe that they would gain and retain knowledge from the lesson. On the contrary, the quantitative analysis provided evidence of significant gains. Perhaps the implementation of CRA, an unfamiliar instructional approach, impacted their perception of knowledge gain and retention. The lesson was designed so that students had to make sense of the process of simplifying square roots visually and progress to making connections to the numerical method while the teacher facilitated the activity. Many students perceived this process as confusing, with several stating that they needed repeated practice before they would retain the knowledge (e.g., "If we continue to practice, then yes! If not, no!").

A need for repeated practice is a hallmark of direct instruction. Based on questionnaire responses, students in the inclusion class seemed to believe that mathematics learning occurs through teacher-led instruction followed by repeated practice. Often, students perceive confusion during a direct instruction lesson to be associated with not learning [65,66]. Perhaps the extensive prior use of direct instruction with students with LDs made them less comfortable with the experience of using manipulatives to develop relational understanding. Providing these students with more information about the CRA instructional approach and how it compares to direct instruction at the beginning of the lesson may have avoided some of the confusion and misunderstood expectations of what students were learning. Implementing the CRA instructional approach regularly may also help students with LDs see the value of this approach. Further research is needed to test these hypotheses.

5. Discussion

Providing high school mathematics students with LDs with an effective instructional approach to develop connections between concrete representations and abstract procedures is within mathematics teachers' reach. Unlike studies presented by Maccini et al. [46], this study provides support that the CRA instructional approach, along with a manipulative that minimizes dual representation, can be effective in both knowledge acquisition and retention in a high school setting. It is important to highlight, however, that the majority of students expressed that they did not believe that this approach was helpful. The students in the larger study, in contrast, had predominately positive statements about the efficacy of the CRA instructional approach.

This study adds to the limited research on using manipulatives in secondary school mathematics classrooms and suggests that the CRA instructional approach needs to be explored further with other abstract topics. As several topics in the typical high school curriculum relate to the properties of squares, this manipulative has the potential to further assist students with abstract thinking across mathematics courses. Two such topics are proportional reasoning related to the area and side length of squares and the converse of the Pythagorean theorem. Further expansion of manipulatives and the CRA approach in secondary school mathematics classrooms also aligns with successful practices demonstrated by Singapore.

This approach allows students to focus on the properties of the manipulative and support their progress to abstract thinking. As many abstract mathematical concepts are grounded in geometric representations, other similar manipulatives should be developed and explored. As further studies emerge that support using manipulatives for abstract thinking, the CRA instructional approach in high school classrooms should increase. In turn, all high school students will have multiple opportunities to develop their relational understanding by making clear connections between visual representations and abstract procedures.

There are several limitations to this study that should be highlighted and discussed. The first limitation was the lack of a control group for the inclusion class. Data from a control group could have compared the CRA approach to traditional instruction. Since the identification of the inclusion class happened after data analysis, data from a control group could not be collected. Including a control group in a future study could provide more clarity regarding student retention when using the CRA approach. Additionally, because the open-ended questionnaire was anonymous, we were unable to link specific students' perceptions of the manipulative lesson with their post-test scores to determine any patterns of student improvement and their perceptions. Future research should create a participant number for each student so these data can be linked.

This study was also limited in what could be determined related to the student's perception of mathematics and the learning of mathematics. The open-response questionnaire provided the researchers with insight into the students' perceptions of the instructional approach and the manipulative as they related to learning a mathematics topic, but not the students' perceptions of themselves as a learner of mathematics. The limited data did, however, provide glimpses into students' perceptions of the purpose of the activity. Future research may indicate the role of the instructor in metacognitive modeling (e.g., "think alouds" [67]) to support students' conceptual understanding of the mathematics content, in general, and the CRA activity, specifically. Creating a classroom culture in which students not only identify confusions but discuss them will also increase metacognition, which can sometimes be neglected when the focus is on a hands-on or interactive lesson [67]. Normalizing these confusions could influence students' willingness to try new activities or persevere when a lesson seems particularly challenging or unhelpful, as was the case with some participants.

Finally, follow-up questionnaires or semistructured interviews could provide further insight into high school inclusion students' self-efficacy as a learner of mathematics after partaking in lessons using the CRA approach, particularly as the students in the present study had substantial misperceptions of their learning with this lesson. Teachers and researchers could conduct clinical interviews, which would elicit information related to students' problem-solving processes that are difficult to obtain through other methods [68,69]. Collecting more qualitative data would allow for a more robust analysis and provide awareness of how different instructional approaches impact students' self-efficacy as learners of mathematics. This type of information could be helpful to teachers, researchers, and policymakers as they consider implementing different instructional approaches throughout the mathematics curriculum.

Author Contributions: Conceptualization, S.F.B.; methodology, S.F.B. and S.K.P.; validation, S.F.B. and S.K.P.; formal analysis, S.F.B.; investigation, S.F.B.; resources, S.F.B.; writing—original draft preparation, S.K.P.; writing—review and editing, S.K.P.; visualization, S.K.P.; supervision, S.F.B.; project administration, S.F.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of South Carolina (protocol code Pro00043889 and 8 September 2015).

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to acknowledge and thank the teachers at Chapman High School in Spartanburg, South Carolina, for opening their classrooms and providing students with the opportunity to explore different ways to think about and learn mathematics.

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

References

1. United States National Commission on Excellence in Education. A Nation at Risk: The Imperative for Educational Reform. 1983. Available online: http://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf (accessed on 8 May 2023).
2. U.S. Department of Education. The Condition of Education: Students with Disabilities. 2020. Available online: https://nces.ed.gov/pubs2020/2020144.pdf (accessed on 8 May 2023).
3. Association of Mathematics Teacher Educators. Position: Equity in Mathematics Teacher Preparation. 2015. Available online: https://amte.net/sites/default/files/amte_equityposistionstatement_sept2015.pdf (accessed on 6 May 2023).
4. National Council of Teachers of Mathematics. *Principles and Standards for School Mathematics*; National Council of Teachers of Mathematics: Reston, VA, USA, 2000; pp. 1–402.
5. National Council of Teachers of Mathematics. *Principles to Actions: Ensuring Mathematics Success for All*; National Council of Teachers of Mathematics: Reston, VA, USA, 2014; pp. 1–139.
6. Bone, E.K.; Bouck, E.C. Evaluating calculators as accommodations for secondary students with disabilities. *Learn. Disabil. Multidiscip. J.* **2018**, *23*, 35–49. [CrossRef]
7. Miller, S.P.; Hudson, P.J. Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learn. Disabil. Res. Pract.* **2007**, *22*, 47–57. [CrossRef]
8. Jitendra, A.K.; Lein, A.E.; Im, S.H.; Alghamdi, A.A.; Hefte, S.B.; Mouanoutoua, J. Mathematical interventions for secondary students with learning disabilities and mathematics difficulties: A meta-analysis. *Except. Child.* **2017**, *64*, 21–43. [CrossRef]
9. Marita, S.; Hord, C. Review of mathematics interventions for secondary students with learning disabilities. *Learn. Disabil. Q.* **2017**, *40*, 29–40. [CrossRef]
10. Moyer, P.S. Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educ. Stud. Math.* **2001**, *47*, 175–197. [CrossRef]
11. Thompson, P.W. Notations, principles, and constraints: Contributions to the effective use of concrete manipulatives in elementary mathematics. *J. Res. Math. Educ.* **1992**, *23*, 123–147. [CrossRef]
12. Gagnon, J.C.; Maccini, P. Preparing students with disabilities for algebra. *Teach. Except. Child.* **2001**, *34*, 8–15. [CrossRef]
13. Strickland, T.K.; Maccini, P. The effects of the concrete-representational-abstract integration strategy on the ability of students with learning disabilities to multiply linear expressions within area problems. *Remedial Spec. Educ.* **2012**, *34*, 134–153. [CrossRef]
14. Clements, D. 'Concrete' manipulatives, concrete ideas. *Contemp. Issues Early Child.* **1999**, *1*, 45–60. [CrossRef]
15. Burns, B.A.; Hamm, E.M. A comparison of concrete and virtual manipulative use in third- and fourth-grade mathematics. *Sch. Sci. Math.* **2011**, *111*, 256–261. [CrossRef]
16. Moyer-Packenham, P.; Baker, J.M.; Westenskow, A.; Rodzon, K.; Anderson, K.; Shumway, J.; Jordan, K. A study comparing virtual manipulations with other instructional treatments in third- and fourth-grade classrooms. *J. Educ.* **2013**, *193*, 25–39. [CrossRef]
17. Westenskow, A.; Moyer-Packenham, P. Using an iceberg intervention model to understand equivalent fraction learning when students with mathematical learning difficulties use different manipulatives. *Int. J. Technol. Math. Educ.* **2016**, *23*, 45–62. [CrossRef]
18. Ball, D. Magical hopes: Manipulatives and the reform of math education. *Am. Educ.* **1992**, *6*, 14–47.
19. Corkin, D.S.; Coleman, S.L.; Ekmekci, A. Navigating the challenges of student-centered mathematics teaching in an urban context. *Urban. Rev.* **2019**, *51*, 370–403. [CrossRef]
20. Anstrom, T. *Supporting Students in Mathematics through the Use of Manipulatives (Research Brief)*; Center for Implementing Technology in Education: Washington, DC, USA, 2006.
21. Leong, Y.H.; Ho, W.K.; Cheng, L.P. Concrete-pictorial-abstract: Surveying its origins and charting its future. *Math. Educ.* **2015**, *16*, 1–18.
22. Mullis, I.V.S.; Martin, M.O.; Goh, S.; Cotter, K. (Eds.) TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. 2016. Available online: http://timssandpirls.bc.edu/timss2015/encyclopedia/ (accessed on 6 May 2023).
23. Witzel, B.S. Using CRA to teach algebra to students with math difficulties in inclusive settings. *Learn. Disabil. Contemp. J.* **2005**, *3*, 49–60.
24. Swan, P.; Marshall, L. Revisiting mathematics manipulative materials. *Aust. Prim. Math. Classr.* **2010**, *15*, 13–19.

25. Carbonneau, K.J.; Marley, S.C.; Selig, J.P. A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *J. Educ. Psychol.* **2013**, *105*, 380–400. [CrossRef]
26. Sowell, E.J. Effects of manipulative materials in mathematics instruction. *J. Res. Math. Educ.* **1989**, *20*, 498–505. [CrossRef]
27. McNeil, N.; Jarvin, L. When theories don't add up: Disentangling the manipulatives debate. *Theory Pract.* **2007**, *46*, 309–316. [CrossRef]
28. Bruner, J.S. On learning mathematics. *Math. Teach.* **1960**, *53*, 610–619. [CrossRef]
29. Bruner, J.S.; Goodnow, J.J.; Austin, G.A. *A Study of Thinking*, 2nd ed.; Transaction Publishers: New Brunswick, NJ, USA, 1986; pp. 1–350. [CrossRef]
30. Piaget, J. *The Origins of Intelligence in Children*, 2nd ed.; International Universities Press: Madison, CT, USA, 1952; pp. 1–419. [CrossRef]
31. Lawson, A.E.; Wollman, W.T. Encouraging the transition from concrete to formal cognitive functioning: An experiment. *J. Res. Sci. Teach.* **2003**, *40*, S33–S50. [CrossRef]
32. Ojose, B. Applying Piaget's theory of cognitive development to mathematics instruction. *Math. Educ.* **2008**, *18*, 26–30.
33. McKinney, S.; Frazier, W. Embracing the principles and standards for school mathematics: An inquiry into the pedagogical and instructional practices of mathematics teachers in high-poverty middle schools. *Clear. House A J. Educ. Strateg. Iss. Ideas* **2008**, *81*, 201–210. [CrossRef]
34. Balter, L.; Tamis-LeMonda, C.S. (Eds.) *Child Psychology: A Handbook of Contemporary Issues*; Psychology Press: London, UK, 2016; pp. 1–522.
35. Orlich, D.C. Education reform and limits to student achievement. *Phi Delta Kappan* **2000**, *81*, 468–472.
36. National Governors Association Center for Best Practices, Council of Chief State School Officers. Common Core State Standards for Mathematics. 2022. Available online: https://learning.ccsso.org/wp-content/uploads/2022/11/ADA-Compliant-Math-Standards.pdf (accessed on 15 January 2023).
37. Bouck, E.C.; Park, J. A systematic review of the literature on mathematics manipulatives to support students with disabilities. *Educ. Treat. Child.* **2018**, *41*, 65–106. [CrossRef]
38. Cass, M.; Cates, D.; Smith, M.; Jackson, C. Effects of manipulative instruction on solving area and perimeter problems by students with learning disabilities. *Learn. Disabil. Res. Pract.* **2003**, *18*, 112–120. [CrossRef]
39. Maccini, P.; Hughes, C.A. Effects of a problem-solving strategy on the introductory algebra performance of secondary students with learning disabilities. *Learn. Disabil. Res. Pract.* **2000**, *15*, 10–21. [CrossRef]
40. Maccini, P.; Ruhl, K.L. Effects of a graduated sequence on the algebraic subtraction of integers by secondary students with learning disabilities. *Educ. Treat. Child.* **2000**, *23*, 465–489.
41. Scheuermann, A.M.; Deshler, D.D.; Schumaker, J.B. The effects of the explicit inquiry routine on the performance of students with learning disabilities on one-variable equations. *Learn. Disabil. Q.* **2009**, *32*, 103–120. [CrossRef]
42. Uttal, D.H.; O'Doherty, K.; Newland, R.; Hand, L.L.; DeLoache, J. Dual representation and the linking of concrete and symbolic representations. *Child. Develop Perspect.* **2009**, *3*, 156–159. [CrossRef]
43. Gersten, R.; Beckmann, S.; Clarke, B.; Foegen, A.; Marsh, L.; Star, J.R.; Witzel, B. *Assisting Students Struggling with Mathematics: Response to Intervention (RTI) for Elementary and Middle Schools. NCEE 2009-4060*; National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education: Washington, DC, USA, 2009.
44. Kelly, C.A. Using manipulatives in mathematical problem solving: A performance-based analysis. *MT Math. Enthus.* **2006**, *3*, 184–193. [CrossRef]
45. Larkin, K. Mathematics education and manipulatives: Which, when, how? *Aust. Prim. Math. Classr.* **2016**, *21*, 12–17.
46. Maccini, P.; Mulcahy, C.A.; Wilson, M.G. A follow-up of mathematics intervention for secondary students with learning disabilities. *Learn. Disabil. Res. Pract.* **2007**, *22*, 58–74. [CrossRef]
47. Sacks, H. Notes on methodology. In *Structures of Social Action: Studies in Conversation Analysis*; Heritage, J., Atkinson, J.M., Eds.; Cambridge University Press: Cambridge, UK, 1984; pp. 2–27.
48. O'Reilly, M.; Kiyimba, N.; Lester, J.N.; Muskett, T. Reflective interventionist conversation analysis. *Discourse Comm.* **2020**, *14*, 619–634. [CrossRef]
49. Weiner, J.M.; Burton, L.J. The double bind for women: Exploring the gendered nature of turnaround leadership in a principal preparation program. *Harv. Educ. Rev.* **2016**, *86*, 339–365. [CrossRef]
50. Greene, J.C.; Caracelli, V.J.; Graham, W.F. Toward a conceptual framework for mixed-method evaluation designs. *Educ. Eval. Policy Anal.* **1989**, *11*, 255–274. [CrossRef]
51. U.S. News and World Report. U.S. News & World Report Releases the 2016 Best High Schools Rankings. 19 August 2016. Available online: https://www.usnews.com/info/blogs/press-room/articles/2016-04-19/us-news-releases-the-2016-best-high-schools-rankings (accessed on 27 August 2022).
52. Schultz, K.T.; Bismarck, S.F. Radical thoughts on simplifying square roots. *Math. Teach. Middle Sch.* **2013**, *19*, 222–228. [CrossRef]
53. Georgia Department of Education. What Do Mathematics Standards Look Like in the Classroom? *2020.* Available online: https://www.georgiastandards.org/ (accessed on 27 August 2022).
54. New York State Education Department. NYS Next Generation Mathematics Learning Standards Unpacking Documents. 2019. Available online: http://www.nysed.gov/ (accessed on 27 August 2022).

55. Ohio Department of Education. Ohio's Model Curriculum with Instructional Supports: Mathematics. 2018. Available online: http://education.ohio.gov/ (accessed on 27 August 2022).
56. Bowling Green State University. NWO/COSMOS: Grant Projects: Middle Grades Resources. n.d. Available online: https://www.bgsu.edu/nwo/current-grant-projects/c2am2p-middle-grades/resources/ (accessed on 27 August 2022).
57. Sloutsky, V.M.; Kaminski, J.A.; Heckler, A.F. The advantage of simple symbols for learning and transfer. *Psychon. Bull. Rev.* **2005**, *12*, 508–513. [CrossRef]
58. Skemp, R. Relational understanding and instrumental understanding. *Math. Teach. Middle Sch.* **2006**, *12*, 88–95. [CrossRef]
59. Brown, G.T.L.; Irving, E.; Keegan, P. *An Introduction to Educational Assessment, Measurement, and Evaluation: Improving the Quality of Teacher-Based Assessment*; Dunmore Publishing: Aukland, New Zealand, 2014; pp. 1–194.
60. Hsieh, H.-F.; Shannon, S.E. Three approaches to qualitative content analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288. [CrossRef]
61. Kurz, T.L. Insight into Perceptions of Mathematics Using Wordle. In Proceedings of the Society for Information Technology & Teacher Education International Conference, Las Vegas, NV, USA, 2–6 March 2015; Rutledge, D., Slykhuis, D., Eds.; Association for the Advancement of Computing in Education: Waynesville, NC, USA, 2015; pp. 2702–2704.
62. McNaught, C.; Lam, P. Using Wordle as a supplementary research tool. *Qual. Rep.* **2010**, *15*, 630–643. [CrossRef]
63. Devlin, K. Finding Your Inner Mathematician. 29 September 2000. Available online: https://www.chronicle.com/article/finding-your-inner-mathematician/ (accessed on 15 January 2023).
64. Maccini, P.; Gagnon, J.C. Best practices for teaching mathematics to secondary students with special needs. *Focus Except. Child.* **2000**, *32*, 1–22. [CrossRef]
65. Wentzel, K.R.; Brophy, J.E. *Motivating Students to Learn*, 4th ed.; Routledge: Oxfordshire, UK, 2013; pp. 1–312.
66. D'Mello, S.; Lehman, B.; Pekrun, R.; Graesser, A. Confusion can be beneficial for learning. *Learn. Instr.* **2014**, *29*, 153–170. [CrossRef]
67. Tanner, K.D. Promoting student metacognition. *CBE Life Sci. Educ.* **2012**, *11*, 113–120. [CrossRef] [PubMed]
68. Forbringer, L.L.; Fuchs, W.W. *RTI in Math: Evidence-Based Interventions for Struggling Students*; Routledge: Oxfordshire, UK, 2014; pp. 1–304.
69. Long, M.J.; Ben-Hur, M. Informing learning through the clinical interview. *Arith. Teach.* **1991**, *38*, 44–46. [CrossRef]

education sciences

MDPI

Article

Exploratory Study on Geometric Learning of Students with Blindness in Mainstream Classrooms: Teachers' Perspectives Using the Van Hiele Theory

Hisae Miyauchi [1,*] and Robinson Thamburaj [2]

1 Institute of Human Sciences, University of Tsukuba, Tsukuba 305-8577, Japan
2 Department of Mathematics, Madras Christian College, Chennai 600059, India
* Correspondence: hmiyauch@human.tsukuba.ac.jp

Abstract: Ensuring mathematics education for all learners, including students with blindness learning in mainstream classrooms, is crucial. This exploratory research aims to clarify the characteristics of geometric learning among students with blindness and to identify the factors contributing to the challenges faced by this population. The Van Hiele theory of geometric thought served as a reference framework. Qualitative data were gathered through group interviews with specialists in the field of education for students with blindness and analyzed using inductive analysis. Participants affirmed that students with blindness progress through Van Hiele levels of geometric thought in a manner similar to sighted students, suggesting that much of the learning can take place alongside sighted peers in mainstream classrooms. However, they also highlighted the unique challenges these students face in reaching level 0, a level where students recognize shapes without a formal understanding of their properties or attributes. Among the reasons for these challenges were that for these particular students, subskills, such as bimanual exploration, hand coordination, and cognitive integration, are required to reach level 0. The study also identified the necessity for specialists in visual impairment education to guide students using appropriate tasks and learning materials that reflect the characteristics of haptic perception. Since level 0 serves as a gateway to both basic and advanced geometry, the findings underscore the importance of providing differentiated support that targets these subskills early in students' schooling. To ensure meaningful geometry instruction, mainstream teachers are encouraged to collaborate with specialists in visual impairment education, who can guide the selection of appropriate learning tools and support the development of the subskills.

Keywords: geometry; students with blindness; subskills; Van Hiele theory

1. Introduction

Geometry, a branch of mathematics addressing spatial sense and geometric reasoning (Howse & Howse, 2014), is a fundamental component of education. Compared to other mathematical topics like numbers, algebra, measurement, and data analysis, geometry occupies a considerable portion of the curriculum at every educational level (Trimurtini et al., 2022).

Blindness is a condition within the spectrum of visual impairment, characterized by a decrease in the ability to see to a degree that challenges daily living or access to learning and cannot be corrected with glasses or contact lenses (World Health Organization, 2019). Students with blindness predominantly use braille and tactile materials for learning. Their needs are distinct from those who are sighted or those who have lost their vision later in life, as these groups can rely on visual imagery, whereas congenitally blind individuals depend on haptic imagery—imagery perceived through touch. Students with blindness, whose impairment is solely in vision, are expected to learn the same mathematics curriculum as their sighted peers and achieve equivalent outcomes. Therefore, providing adequate geometry education to students with blindness is equally important as it is for students without blindness.

In recent years, there has been a growing trend to include students with disabilities, including blindness and other diverse needs, in mainstream rather than specialized schools (e.g., schools for the blind). This shift toward inclusive education is influenced by international and national frameworks, such as the Convention on the Rights of Persons with Disabilities (United Nations, 2006) as a former example. These frameworks assert that all students, including those with blindness, have the right to be included and receive adequate education alongside their sighted peers in their local communities or mainstream schools. This right to inclusive education has opened new avenues for individuals with disabilities and prompted mainstream schools to become more flexible and innovative in addressing the needs of diverse learners, many of whom had previously been overlooked (Dalgaard et al., 2022; Szumski et al., 2017). However, while inclusive education is promising in theory, in reality, students with specific disabilities, such as blindness, face many barriers in mainstream schools. A key example of this is their exclusion from higher-level "visual subjects"—subjects that rely heavily on visual input or practical applications, such as mathematics, science, and physical education (Miyauchi, 2020). Despite their abilities, students with blindness are often denied access to these subjects, highlighting a significant shortfall in inclusive education.

The Van Hiele theory of geometric thought, which this research used as a reference, is a well-known framework explaining how students learn geometry (Crowley, 1987). It consists of five levels of understanding, beginning with level 0 (visualization), where students recognize shapes by their visual appearance without a formal understanding of their properties or attributes. At level 1 (analysis), students notice different shapes by recognizing attributes but do not yet understand their relevance. This is followed by level 2 (abstraction), level 3 (deduction), and level 4 (rigor), where students progressively identify and analyze geometric forms, moving toward higher levels of abstraction (Crowley, 1987; Naufal et al., 2021). Past studies confirmed that students transition gradually from one level to the next, often moving back and forth between stages without sudden jumps or skipping levels (Duroisin & Demeuse, 2015).

To the authors' knowledge, Argyropoulos (2002) is the only study that explores Van Hiele's theory in relation to students with blindness. This study suggests that Van Hiele's theory is a suitable framework for examining the geometrical thinking process of these students while also highlighting the unique challenges they face, particularly at the earlier levels. However, given the scarcity of research on this topic, further investigation is needed.

Exploring how students with blindness learn geometry and the challenges they face using a framework designed for sighted learners is crucial for advancing inclusive education in the following ways. First, this process, we believe, will help reveal both the similarities and differences in how children with and without visual impairments learn. Recognizing these similarities and differences is essential, as societal perceptions of human differences can be both positive and negative—what some researchers term the dilemmas of difference (Norwich, 2008; Paulsrud, 2024). Acknowledging and addressing a child's unique learning needs may lead to physical separation (or segregation), potentially resulting in feelings of exclusion and a lack of acceptance among peers. However, failing to recognize and respond to these differences can limit students' access to resources and specialist services, hindering their equitable participation in education and society. Hence, finding the right balance between differences and similarities in how children learn is crucial.

Second, this process helps us understand what effective differentiated teaching within a mainstream classroom may look like—one that strikes the balance between differences and similarities in how children learn. Differentiated teaching is a method employed by teachers that involves assessing and monitoring students' learning readiness and processes (Vaughn et al., 2022/2023). Its purpose is to extend the knowledge and skills of each individual child; hence, this approach includes small-group or individualized instruction and the use of tailored instructional tools, rather than rigid whole-class instruction with the same tools for all students. Incorporating differentiated teaching into mainstream classroom instruction has been identified as a key strategy for enhancing inclusion (Soan & Monsen, 2023).

This research uses the Van Hiele theory of geometric thought to explore how students with blindness learn geometry. This approach provides valuable insights into achieving a balance between recognizing commonalities in the learning processes of students with blindness and sighted students and identifying the specific content that should be reflected in differentiated teaching. Such an understanding can guide mainstream teachers in effectively collaborating with specialists, supporting and enhancing the learning experiences of students with blindness in classrooms.

This study aims to clarify the characteristics of geometry learning among students with blindness in comparison to their sighted peers, by interviewing specialists in the education of students with blindness using the Van Hiele theory of geometric thought as a reference. By highlighting both similarities and differences, this research seeks to identify areas where differentiated teaching is needed to foster effective inclusive education. Furthermore, it aims to provide practical insights for mainstream teachers to enhance the learning experiences of students with blindness.

The specific research questions addressed are as follows:

1. Does the Van Hiele theory align with how teachers and specialists observe their students with blindness learn geometry?
2. What specific challenges do students with blindness face in learning geometry?
3. What are the origins of these unique challenges?

2. Materials and Methods

2.1. Study Design

Given the limited research on geometry education for students with blindness, this study adopted an exploratory design. The research focused on qualitative group interviews with a purposive sample of Japanese individuals who possessed extensive knowledge of teaching both students with and without blindness. Qualitative research methods were chosen as they are appropriate to expand knowledge on unexplored areas. Group interviews were chosen owing to their ability to foster dynamic discussions, allowing participants to build on each other's responses and generate rich data through group interaction (Krueger & Casey, 2014).

2.2. Study Setting

2.2.1. Study Context

With the Japanese government promoting an "inclusive education system", whereby students with disabilities are enrolled in mainstream classrooms to the extent possible (Ministry of Education, Culture, Sports, Science & Technology, 2013), more children with visual impairment are learning in mainstream settings. However, within visual impairment, students who uses braille (hence, students with blindness) tend to continue to enroll in schools for the blind, as these schools offer more specialized support. Japan is a unitary state where the central government, specifically the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), governs the national curriculum known as the Course of Study. All textbooks used in schools are authorized by MEXT based on the Course of Study. Students with blindness are required to follow the same Course of Study as their sighted peers. Hence, braille textbooks for these students are created by modifying textbooks used in mainstream schools.

2.2.2. Recruitment and Sampling

Purposive sampling was employed to recruit participants with extensive teaching knowledge who could provide in-depth insights into the similarities and differences in geometry learning between sighted students and students with blindness. The selection was based on two main criteria. First, participants had to be either teachers or lecturers. Teachers were required to hold teaching certificates in mathematics or other relevant subjects (e.g., geography) for both mainstream schools and schools for the blind, with at least 20 years of teaching experience. Lecturers, on the other hand, needed to have a PhD in the field and demonstrate expertise through teaching and research on the education of students with visual impairments. Second, to ensure participants had extensive knowledge of the similarities and differences in how students with blindness learn compared to sighted students, all participants—whether teachers or lecturers—had to have served as MEXT-appointed editorial board members in adapting school textbooks into Braille versions for students with blindness. In Japan, braille modification is conducted strictly to preserve the content of mainstream school textbooks, altering only the aspects that must be adapted due to vision impairment. As a result, editorial board members are required to have expertise, not only in how blind students learn, but also in how sighted students learn, ensuring that the adapted materials remain as equivalent as possible. This selection process ensured that participants aligned with the purpose of this research. Lastly, participants who were blind themselves were intentionally included if they met the above criteria. This approach ensured that the study incorporated the lived experiences of blind individuals, who learned through tactile perception.

A total of five teachers and experts in the field participated in this study. Of these, four were teachers (three math teachers and one geography teacher) and had an average of 27.75 (SD: 6.0) years of experience teaching students with blindness. Among the four teachers, one was female and three were male. The three male teachers were congenitally blind. All four teachers held dual teaching certificates: one for lower and upper secondary school (subject areas of mathematics or social studies, geography, and history) and another teaching certificate for special needs education (with a focus on visual impairment). All were actively involved in editing braille textbooks authorized by the government in addition to training other educators and authoring books on the education of students with visual impairment. One was a part-time university lecturer who was congenitally blind and possessed expertise in tactile arts and graphics for students with blindness. Participant details are described in Table 1.

Table 1. Demographics of the individuals interviewed.

Participant	Sex	Age	Occupation	Obtained Degree/Licensure	Number of Years Teaching	Visual Impairment Status
A	Male	60s	Lower and upper secondary math teacher at school for the blind, retired	Teaching certificate for lower and upper secondary school mathematics/teaching certificate for special needs education (visual impairment)	36	Blind
B	Male	50s	Lower and upper secondary math teacher at school for the blind	Teaching certificate for lower and upper secondary school mathematics/teaching certificate for special needs education (visual impairment)	28	Blind
C	Female	50s	Lower and upper secondary math teacher at school for the blind	Teaching certificate for lower and upper secondary school mathematics/teaching certificate for special needs education (visual impairment)	25	Sighted
D	Male	50s	Lower and upper secondary social studies, geography, and history teacher at school for the blind	Teaching certificate for lower and upper secondary school social studies, geography, and history/teaching certificate for special needs education (visual impairment)	22	Blind
E	Female	60s	Parttime university lecturer	Doctor of Philosophy in art	NA	Blind

2.2.3. Data Collection and Ethical Considerations

A total of four in-person group semi-structured group interviews, with all participants present were conducted. The number of group interviews was determined based on data saturation. Each interview lasted approximately two hours.

A semi-structured interview protocol was developed specifically for this study based on the research question. The protocol consisted of five open ended questions: "What are your thoughts on the Van Hiele theory? In what ways does Van Hiele theory align with or differ from students with blindness' geometric learning?"; "What are your thoughts on the levels? Does learning typically proceed through these levels?"; "What challenges do students with blindness face in learning geometry?"; "How do you address these challenges as a teacher?"; and "What factors do you think contribute to these challenges?" During the initial meeting, the Van Hiele theory was introduced and explained to participants, along with definitions of each level. Participants were encouraged to express their thoughts freely so that the conversation flowed naturally. The lead author facilitated the interviews and used the open-ended questions as a guide to ensure discussions remained relevant and did not stray from the research focus.

Prior to the interviews, participants were provided with a consent form, which the first author read aloud, and written consent was obtained. All interviews were audio-recorded, transcribed verbatim, and anonymized during transcription. This research was approved by the Research Ethics Committee of the first author's affiliated university (No. Tsuku2023-232A).

2.2.4. Analysis

The data were analyzed inductively with the help of NVivo software, following the guidance of Braun and Clarke (2006). This approach involved reading and re-reading the data to identify initial ideas, systematically coding the entire dataset, and collating these initial codes into potential themes. The first author, who has PhD training in qualitative methods, performed this process. The themes were then reviewed with the co-author to ensure their applicability across the entire dataset. Subsequently, the themes were refined, named, and representative extracts were identified as exemplars for each theme. This iterative process continued until thematic saturation was achieved.

3. Results

3.1. Results from the Analysis

The following four themes emerged from the analysis: "similar geometric thought processes as sighted individuals"; "challenges exclusive to students with blindness at the visualization level"; "visualization level requiring multiple tactics"; and "the need for specialists to guide students with blindness with appropriate tasks and learning materials". Each theme is summarized in Table 2, along with illustrative data. The following sections provide a detailed explanation of each theme.

Table 2. Themes and illustrative data.

Theme	Illustrative Data
Similar geometric thoughts processes as sighted individuals	"Yes, this (levels described by Van Hiele) makes sense. I agree with the levels described by Van Hiele" (Participant B). "The process (described by Van Hiele makes sense" (Participants A, C, D, E).
Challenges exclusive to students with blindness at the visualization level	"Once the blind students can pass this level, the other levels will be just like for the sighted" (Participant B). "Level 0 is where you will see the challenges among students because they have no vision to rely on" (Participants A, C).
Visualization level requiring multiple tactics	"I have seen many children who can 'touch' but cannot understand what they are touching" (Participant C). "Good hand movement, along with the skills to obtain and integrate tactile details in their head to understand the overall shape or form of an object, is necessary... because you have no vision to rely on" (Participants A, B). "Whether or not the blind child is interested in touching matters a lot" (Participants B, D, E).
The need for specialists to guide students with blindness with appropriate tasks and learning materials	"Our hands naturally curve, making it easier to perceive 3D shapes. However, with 2D shapes, one must intentionally control the fingers, adjusting both pressure and direction to accurately perceive the shape" (Participant A). "When touching a contour line, the child needs to be able to distinguish curved line from a straight line, which is hard for blind children who cannot rely on vision...we usually have blind students practice using shapes with curbed line, like a circle... and straight line, using like a polygon, first so they understand the difference" (Participants A, B, C).

3.1.1. Similar Geometric Thought Process as Sighted Individuals

All participants agreed that students with blindness follow the geometric thought process outlined by Van Hiele. Furthermore, they confirmed that although students may go back and forth between levels, there is a distinct thinking/learning sequence. Therefore, level 0 serves as the foundational level necessary for students to progress to levels 1, 2, and beyond, echoing past studies focused on sighted students.

3.1.2. Challenges Exclusive to Students with Blindness at the Visualization Level

While all participants agreed that students with blindness follow the thinking process suggested by Van Hiele, they particularly emphasized the importance of the level 0, the visualization level. Unlike sighted students, who typically progress through this level with ease, some students with blindness encounter significant obstacles, preventing them from advancing beyond visualization and causing delays in understanding more advanced geometry. Participants remarked, "Once the blind students can pass this (the visualization) level, the other levels will be just like for the sighted. Of course, some students will have difficulty transitioning from the analysis level to the abstraction level, but that is similar to sighted students, and the reason would not solely be due to vision" (Participant B).

3.1.3. Visualization Level Requiring Multiple Tactics

All participants highlighted the significant differences and complexities that students with blindness face at level 0, particularly concerning visualizing shapes through touch. One teacher emphasized, "I have seen students who can 'touch' but cannot understand what they are touching. If you are sighted, you see and you understand what you are seeing" (Participant C). Another teacher, who is blind, added, "The act of touching does not equate to the act of seeing" (Participant A).

According to participants, visualizing shapes through touch requires a range of complex skills and strategies, even if the shape is a simple, basic geometric form. These challenges highlight the intricate haptic system that students with blindness must navigate because of their impairment. For instance, participants noted that for children with blindness to visualize geometric forms, the ability to effectively use both hands to capture the object as a whole—and not just a portion—is foundational. While sighted students typically perceive an entire object at once unless parts are covered, students with blindness may touch only a portion of an object and may misidentify it. One teacher shared, "A blind child who lacks basic hand exploration skills may touch only a portion of a rectangle, like the upper right corner, and mistakenly identify it as a triangle" (Participant C).

Effective hand strategies are also crucial. Participants emphasized the need for techniques where one finger remains on a fixed point while the other hand explores the overall shape, aiding the child in building a complete mental image. Equally important is the ability for students to distinguish fundamental features such as straight lines, curved lines, and angles after mentally assembling the image. Two participants explained: "We first practice distinguishing between curved and straight lines before introducing shapes like circles and polygons. This helps students understand the differences within these shapes when touching" (Participants B and C).

Furthermore, strategies for retaining and integrating the images obtained through touch were highlighted. Participants noted, "Verbalizing is essential for categorizing the information touched, allowing students to piece together the complex image in their minds" (Participants A, B, and C).

Lastly, intrinsic motivation also played a key role, highlighting the active nature of tactile exploration compared to the passive nature of visual perception. Participants expressed, "Whether or not the blind child is interested in touching matters a lot" (Participants B, D, and E). Another participant elaborated, "When using vision, external information comes automatically and unintentionally. When you are blind, information only comes in when you touch actively and intentionally" (Participant A).

3.1.4. The Need for Specialists to Guide Students with Blindness with Appropriate Tasks and Learning Materials

The size of the object and whether it is two-dimensional or three-dimensional also appeared to significantly affect the ability of children with blindness to reach level 0. Larger objects that exceed the size of a child's palm presented additional challenges, requiring more advanced dual-hand manipulation skills.

All participants agreed that touching two-dimensional shapes requires more technique than touching three-dimensional shapes. One teacher explained: "Our hands naturally curve, making it easier to perceive 3D shapes. However, with 2D shapes, one must intentionally control the fingers, adjusting both pressure and direction to accurately perceive the shape" (Participant A).

All participants, who also taught students with blindness daily, mentioned how they supported them in developing these skills and tactics. Specifically, they provided adequate two-dimensional and three-dimensional models and offered verbal cues on how to tactically touch and retain tactilely perceived information in their minds.

4. Discussion

This exploratory research aimed to clarify the characteristics of geometric learning among students with blindness and to identify the factors contributing to the challenges faced by this population. Through interviews with specialists in the education of students with blindness and using the Van Hiele theory of geometric thought as a framework, it sought to provide practical insights for mainstream schoolteachers on how they could effectively support and enhance the learning experiences of students with blindness.

Participants confirmed their agreement with the structure and processes outlined in the Van Hiele theory, aligning with previous research by Argyropoulos (2002). The findings suggest that the geometry learning process for students with blindness follows a closely aligned pattern to that of their sighted peers, indicating that much of the learning can occur alongside sighted peers in mainstream classrooms.

However, participants also noted specific challenges for students with blindness at level 0, the visualization level. While the Van Hiele theory consists of five levels, participants' statements primarily focused on level 0, suggesting that this stage requires particular attention for students with blindness. This finding is in line with Argyropoulos (2002), whose experimental study with students with blindness identified that while many struggled to reach level 1, some faced challenges even in achieving level 0. Although the methodologies differ, both studies highlight the significant difficulties that students with blindness encounter at the visualization level. In contrast, research on sighted students has shown that while many struggles to progress beyond level 3 or are in a transitional stage between Levels 0 and 1, they generally do not face challenges in reaching level 0 itself (Ma et al., 2015; Škrbec & Čadež, 2015; Trimurtini et al., 2022). Thus, this study further underscores that students with blindness encounter unique difficulties that sighted students do not.

Furthermore, participants identified bimanual exploration, hand coordination, and cognitive integration as important skills for students with blindness to reach level 0. As these skills are specific to those who rely on tactile perception, this highlights the need for differentiated teaching that focuses on developing these skills. This underscores the importance of implementing differentiated teaching which addresses these particular skills, early, as reaching the level 0 is a gateway for learning basic and more advanced geometry alongside the sighted peers in the mainstream school settings.

Finally, the choice between 2D and 3D representations was identified as an important factor in supporting students with blindness in reaching the visualization level. Tools that are too large or overly simplified 2D representations may place excessive cognitive demands on students, potentially hindering their progress. This underscores the importance of not only implementing differentiated teaching early in students' schooling but also carefully selecting learning materials and adopting teaching strategies that are tailored to the unique characteristics of tactile perception.

Although the skills highlighted above have not previously been discussed in direct connection with geometry learning using the Van Hiele theory, past research emphasized their significance for students with blindness. For instance, Kershman (1977) identified subskills necessary for tactile discrimination, including active touch, fine hand movements, and whole-hand explorations. Additionally, in research on haptic perception, Lederman and Klatzky (2009) defined how haptic perception relies on multiple sensory inputs—such as cutaneous and kinesthetic input—along with working memory to retain images formed through tactile exploration. For example, when geometric figures are small enough to fit under a fingertip, they can be identified primarily through kinesthetic input. However, larger objects requiring the use of both hands (palms and fingers) demand additional reliance on cutaneous input and working memory, making the visualization process more complex.

From the above findings, it is evident that students who rely on haptic perception encounter unique challenges in reaching the visualization level of the Van Hiele theory, challenges that sighted students do not face. Several factors specific to students with blindness significantly impact their ability to "visualize" geometric shapes due to the nature of haptic perception, which mainstream teachers should be aware of. The role of specialists, such as teachers of students with visual impairments who understands the principles of haptic perception and can select appropriate learning tools is critical. Teachers with this specialized knowledge should collaborate closely with mainstream teachers and teaching assistants who may directly be involved in supporting children with blindness in the classroom to ensure students of this population can fully access and receive meaningful geometry instruction in mainstream classrooms.

Lastly, this research is not without limitations. As an exploratory study, it relied primarily on insights from a specific group of individuals in Japan. Future research should include a more diverse range of participants from various regions and cultures to enhance the broader applicability of these findings.

5. Conclusions

This study highlights that students with blindness encounter unique challenges that sighted students do not. To address these challenges, the study proposes effectively using differentiated teaching focusing on specific skills needed to reach level 0, such as bimanual exploration, hand coordination, and cognitive integration. Furthermore, special attention should be given to both the content of instruction and the learning tools used, ensuring they align with the characteristics of haptic perception. Since level 0 serves as a gateway to both basic and advanced geometry, it is crucial that differentiated teaching is provided early in students' school lives. This support should involve collaboration with specialists in the education of students with visual impairments, who can guide the selection of appropriate tools to facilitate skill development.

Author Contributions: Conceptualization, investigation, data analysis, and validation, H.M. and R.T.; writing—original draft preparation, H.M.; writing—review and editing, H.M. and R.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI, grant number 23KK0038.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethical Committee of University of Tsukuba (protocol code Tsuku2023-232A; date of approval: 5 March 2024).

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

Data Availability Statement: The data that support the findings of this study are not openly available due to their containing information that could compromise the privacy of research participants. Data can be requested from the corresponding author. The corresponding author will conduct a confidentiality review before releasing the data.

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

References

Argyropoulos, V. S. (2002). Tactual shape perception in relation to the understanding of geometrical concepts by blind students. *British Journal of Visual Impairment*, *20*(1), 7–16. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

Crowley, M. (1987). The van Hiele model of the development of geometric thought. In M. Lindquiest (Ed.), *Learning and teaching geometry, K–12* (pp. 1–16). National Council of Teachers of Mathematics.

Dalgaard, N. T., Bondebjerg, A., Viinholt, B. C. A., & Filges, T. (2022). The effects of inclusion on academic achievement, socioemotional development, and well-being of children with special educational needs. *Campbell Systematic Reviews*, *18*(4), 1291. [CrossRef]

Duroisin, N., & Demeuse, M. (2015). What role for developmental theories in mathematics study programmes in French-speaking Belgium? An analysis of the geometry curriculum's aspects, framed by Van Hiele's model. *Cogent Education*, *2*(1), 1049846. [CrossRef]

Howse, T. D., & Howse, M. E. (2014). Linking the Van Hiele theory to instruction. *Teaching Children Mathematics*, *21*(5), 304–313. [CrossRef]

Kershman, S. M. (1977). A hierarchy of tasks in the development of tactual discrimination: Part two. *Education of the Visually Handicapped*, *8*(4), 107–115.

Krueger, R. A., & Casey, M. A. (2014). *Focus groups: A practical guide for applied research*. SAGE Publications.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*(7), 1439–1459. [CrossRef]

Ma, H. L., Lee, D. C., Lin, S. H., & Wu, D. B. (2015). A study of Van Hiele of geometric thinking among 1st through 6th graders. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 1181–1196. [CrossRef]

Ministry of Education, Culture, Sports, Science and Technology. (2013). Outline of report on the promotion of special needs education for developing an inclusive education system leading to the creation of cohesive society. *NISE Bulletin*, *12*, 22–27.

Miyauchi, H. (2020). A systematic review on inclusive education of students with visual impairment. *Education Sciences*, *10*(11), 346. [CrossRef]

Naufal, M. A., Abdullah, A. H., Osman, S., Abu, M. S., Ihsan, H., & Rondiyah. (2021). Reviewing the Van Hiele model and the application of metacognition on geometric thinking. *International Journal of Evaluation and Research in Education*, *10*(2), 597–605. [CrossRef]

Norwich, B. (2008). Dilemmas of difference, inclusion and disability: International perspectives on placement. *European Journal of Special Needs Education*, *23*(4), 287–304. [CrossRef]

Paulsrud, D. (2024). Resolving dilemmas: Swedish special educators and subject teachers' perspectives on their enactment of inclusive education. *Journal of Education Policy*, *39*(2), 303–320. [CrossRef]

Soan, S., & Monsen, J. (2023). *Inclusive education theory and policy: Moving from special educational needs to equity*. Open University Press.

Szumski, G., Smogorzewska, J., & Karwowski, M. (2017). Academic achievement of students without special educational needs in inclusive classrooms: A meta-analysis. *Educational Research Review*, *21*, 33–54. [CrossRef]

Škrbec, M., & Čadež, T. H. (2015). Identifying and fostering higher levels of geometric thinking. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(3), 601–617. [CrossRef]

Trimurtini, Waluya, S. B., Sukestiyarno, Y. L., & Kharisudin, I. (2022). A systematic review on geometric thinking: A review research between 2017–2021. *Eurasian Society of Educational Research*, *11*(3), 1535–1552. [CrossRef]

United Nations. (2006). Convention on the rights of persons with disabilities. *Treaty Series*, *2515*, 3. Available online: https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-articles (accessed on 1 February 2025).

Vaughn, S., Alsolami, A., & Swanson, E. (2023). Differentiating instruction for students who are blind and with low vision. *Teaching Exceptional Children*, *55*(4), 244–250, (Original work published 2022). [CrossRef]

World Health Organization. (2019). *World report on vision*. World Health Organization. Available online: https://iris.who.int/bitstream/handle/10665/328717/9789241516570-eng.pdf?sequence=18 (accessed on 1 February 2025).

education sciences

Article

Building Mathematics Learning through Inquiry Using Student-Generated Data: Lessons Learned from Plan-Do-Study-Act Cycles

Christopher R. Rakes [1,*], Angela Wesneski [2] and Rebecca Laws [3]

[1] Department of Education, University of Maryland Baltimore County, Baltimore, MD 21250, USA
[2] Curtis Bay Elementary, Baltimore City Public Schools, Baltimore, MD 21225, USA; arwesneski@bcps.k12.md.us
[3] Baltimore City College High School, Baltimore City Public Schools, Baltimore, MD 21218, USA; rilaws@bcps.k12.md.us
* Correspondence: rakes@umbc.edu

Abstract: This paper describes how plan-do-study-act cycles engaged a classroom mentor teacher and student teacher in a professional collaboration that resulted in two inquiry activities for high-school geometry classes. The PDSA cycles were carried out in four high school geometry classes, each with 30 to 35 students, in a mid-Atlantic urban school district in the U.S. The four geometry classes were co-taught by the second and third authors of this paper. The data consisted of classroom documents (e.g., activity prompts, tasks), classroom observations, student feedback about activities, and monthly PDSA reports. The PDSA cycles had a direct effect on the professional learning of the teachers. The resultant classroom activities used a data collection approach to engaging students in inquiry to learn about trigonometry functions and density. Student learning behaviors were noticeably improved during these activities compared with traditional mathematics instruction. We concluded that the data collection sequence provided an accessible entry point for students to begin scientific inquiry in mathematics. The process opened the conceptual space for students to develop curiosity about mathematical phenomena and to explore their own research questions. The use of culturally relevant topics was especially compelling to students, and the open-ended nature of these exploratory activities allowed students to see mathematics through their own cultural lenses.

Keywords: mathematics education; inquiry; student-generated data; improvement science; teacher classroom research

1. Introduction

Teacher preparation programs are considered one of the most effective leverage points for long-term improvement in teacher performance and retention of productive teachers [1–3]. Yet, the reform-based practices promoted by universities seldom find their way into the secondary mathematics classroom, limiting the ability of these programs to transform the field. Several contributing factors have been identified to explain this discrepancy, such as lack of reform-teaching models, greater intellectual demands on teachers, and resistance to change [4]. Gainsburg [4] also noted that the demands of reform-based teaching are especially burdensome for new teachers.

Although research is scarce on how to effectively prepare new mathematics teachers [5,6], many aspects of effective professional development (PD) have been well studied (e.g., Desimone [7]; Loucks-Horsely et al. [8]). Structuring teacher preparation as initial professional development, consistent with Bangel et al. [9] and Pollock et al. [10], allows the preparation program to benefit from existing knowledge about effective PD. By "effective," we mean that the experience supports the development of teachers as professionals and results in significant improvements in classroom practice [11,12]. The inclusion of teachers in

PD design and the application of professional learning promotes their growth as professionals [13] and situates PD experiences within particular school and classroom contexts [14]. Other characteristics of effective PD are an emphasis on student learning and classroom practice, a focus on specific academic content, and sustained opportunities for teachers to collaborate and provide peer feedback [7,8,15]. The Professional Development: Research, Implementation, and Evaluation framework ("PrimeD;" [16,17]) was designed to synthesize research and theory about effective PD. In the present study, PrimeD was applied to a teacher preparation program to support the professional learning of new and experienced teachers simultaneously. Through iterative cycles of whole-group activities and classroom implementation, the connection between professional learning and classroom practice is made explicit. Plan-do-study-act (PDSA) cycles [18] provide an organizational structure to classroom implementation.

In this article, we present a case study of a teacher candidate and classroom mentor teacher (hereafter "mentor") who, through plan-do-study-act (PDSA) cycles [18], developed a series of reform-based lesson activities throughout the full-time student teaching semester. The overarching questions driving the project were:

1. How do PDSA cycles support pedagogical innovation in the classroom?
2. How can reform-based teaching be transferred from theoretical ideas to classroom practice during full-time student teaching?

The candidate and mentor were participants in a teacher preparation program guided by PrimeD and developed a series of reform-based lesson activities during the full-time student teaching semester. The experiences of the candidate and mentor provide insights into the dynamics and ramifications of framing teacher preparation as professional development through PrimeD.

2. Background

PrimeD structures professional learning through four phases: design, implementation, evaluation, and research. In Design Phase I, participants map out a challenge space that includes a mission, vision, goals, targets, and strategies. In Implementation Phase II, participants form a networked improvement community (NIC) and meet regularly as a group. Change ideas developed during NIC meetings are taken to the classroom using plan-do-study-act (PDSA) cycles. Participants return to the whole-group meetings with results from their PDSA cycles. Phase III Evaluation consists of both formative and summative feedback. In Research Phase IV, research about the PD program is conducted, and findings from PDSA cycles are generalized across contexts.

Using PrimeD to structure teacher preparation is a unique and comprehensive approach for examining how to translate learning from a preparation program into actual teaching practices in the field. A lack of coherence between theory and practice may explain why some teachers do not use the strategies learned in their preparation program in their classrooms [4,19]. The implementation of PrimeD [16,17] to structure teacher preparation directly addresses such incoherence by explicitly connecting a well-defined, commonly-agreed-upon challenge space to pedagogical strategies that are used in coursework and field experience settings and refined through an iterative improvement process.

2.1. The PrimeD Framework: A PD Framework for Teacher Preparation

The PrimeD framework was initially developed through a systematic review of the literature [20] and through the evaluation of a state-wide PD program [17]. PrimeD applies the principles of improvement science to professional learning [21–23]. The use of PrimeD situates teacher preparation as PD, consistent with Bangel et al. [9] and Pollock et al. [10]. PrimeD organizes PD into four phases that work in a cyclic nature and occur iteratively throughout a PD program (Figure 1).

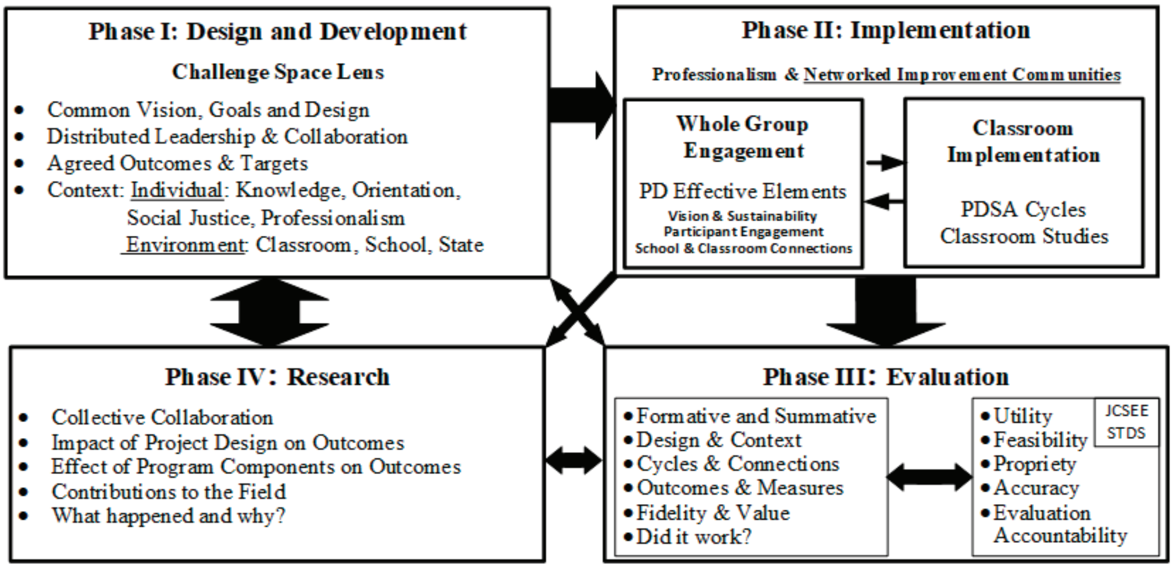


Figure 1. Condensed Model of the PrimeD Framework (adapted from [16]).

2.1.1. Phase I Design and Development

Phase I is foundational to every other phase in PrimeD. It goes beyond simply planning PD (e.g., courses, seminars, and field experiences in a teacher preparation program). Stakeholders come together to map out a challenge space—an explicit description of needs, vision, goals, targets, and strategies for meeting the challenges being addressed by and faced within the program [21]. In teacher preparation, participants include university faculty, classroom teachers, field experience supervisors, and teacher candidates. The challenge space is more than a list of obstacles or difficulties; it embodies the program's call to action to improve professional practice (e.g., classroom teaching, professional learning, and leadership) and expresses a pragmatic vision of the potential for systemic and systematic change [16]. Each course, class session, and field experience should be purposeful and intentionally aligned with the challenge space. But perhaps more importantly in teacher preparation, structural supports are needed to bind course and field experiences together into a coherent system wholly focused on achieving particular goals and outcomes defined by the challenge space [4,19,21].

2.1.2. Phase II Implementation

A program using PrimeD as its framework intends to engage teachers and teacher candidates as professional partners. The role of PD providers is to engage participants collaboratively with research and tools to support professional decision-making. As Datnow and Stringfield [24] noted:

> The fundamental difference between an amateur and a professional in any field is not one of intelligence or willingness to work hard. Rather, it is that professionals are trained at accessing their own research field, and therefore are much less likely to spend time repeating the others' prior mistakes. Educational reforms seem to have a less-than-glorious tradition of replicating major aspects of previous failed efforts. (p. 197)

Network improvement communities (NICs) and plan-do-study-act (PDSA) cycles [18,21] are the primary components of PrimeD Phase II and provide participants with opportunities to direct their own professional learning and apply their learning to the classroom. An NIC focuses on a problem of practice and develops change ideas to address that problem in the field through PDSA cycles. A problem of practice addresses obstacles to learning in the classroom that are focused on instructional practices and are actionable, observable, and measurable. PDSA cycles are intended to be rapid, small-scale changes that build over time into measurable improvement at scale [21]. For example, a teacher may change the way new topics are introduced and may refine the strategy each class period of the day. One advantage of pursuing PDSA cycles in groups is that the same trial can be tried out by multiple teachers in multiple settings to provide a more comprehensive test of the strategies studied.

2.1.3. Phase III Evaluation

As professionals, teachers participate in establishing what is best practice [25]. Engaging candidates in evaluation lays a foundation for professionalism throughout their careers. Evaluation cycles in PrimeD include feedback mechanisms to the challenge space (arrow from Phase III to Phase I in Figure 1).

Participants engage in regular self-evaluation through the PDSA cycles and peer evaluation through small- and large-group presentations at NIC meetings. Facilitators observe discussions at the NIC meetings as a formative assessment. Two to three local and non-local evaluators observe NIC meetings and provide monthly feedback about the quality of the NIC meetings and alignment to PrimeD. This feedback is used by NIC planning teams to guide subsequent meetings. The planning teams consist of faculty and representatives from the participant groups (e.g., mentors and candidates).

2.1.4. Phase IV Research

Teachers regularly carry out action research in their classrooms [26] and seek out research that is directly applicable to the classroom [27]. Teachers may at the same time think of "research" as a hands-off activity with little connection to the classroom [18]. Methods such as design-based research are especially useful to support partnerships between researchers and practitioners with a goal of generating outcomes that are both practical and contribute to theory [26]. PrimeD recognizes that viewing research as a seamless component of PD adds access, richness, and complexity to the process and has been shown to improve professional learning outcomes for teachers (e.g., [28–31]).

Teachers ideally conduct research as a normal function of their practice; that is, they test and evaluate their approach to teaching every day, seeking causal explanations for outcomes they observe. But these types of efforts are often contextually limited. The connection between implementation (Phase II) and research (Phase IV) activities (one-way arrow in Figure 1) situates classroom research activities as a first step toward generalizing results to be useful for a larger audience. While implementing PD innovations in Phase II's PDSA cycles, teachers create research questions from their classrooms. Results are generalized in Phase IV, when they are shared with the larger group to be tried and vetted to determine what works and does not work for desired outcomes under various conditions and why [32]. The NIC may use a variety of approaches and designs to generalize results beyond specific classroom contexts.

The inclusion of Phase IV in teacher preparation indicates an intention to prepare candidates to engage in professional research as teachers. Through the NIC and PDSA cycles, candidates observe mentor activities, ask questions, engage collaboratively, and develop the necessary foundations for contributing to the knowledge base. Mentors, supervisors, and faculty help to hone candidates' professional judgment as they draw conclusions about their classroom research.

2.2. Reformed Teaching, Inquiry, and Constructivism

The mathematics teaching field has recognized for centuries the need to reform traditional teaching techniques to improve learners' conceptual and relational understanding, critical thinking, and reasoning (e.g., [33,34]). Traditional epistemology in U.S. mathematics classrooms views the teacher as an authority who conveys knowledge to students, who are largely viewed as blank slates [35,36]. Constructivism views learning as the construction of meaning by the learner rather than the passive reception of knowledge [37].

Piaget described the process of knowledge acquisition through a constructivist perspective. When students encounter new information that fits into their existing conceptual framework, the new information is assimilated (not requiring reconstruction of students' schemas/conceptual frameworks). For example, when a student believes that when two numbers are multiplied the product is always larger than the original two numbers, and every multiplication example they encounter results in larger products, their conceptual framework will be reinforced, leaving intact their belief about multiplicative structures. Sometimes, however, the information is recognized as not aligning to their current schema, requiring accommodation, in which case a restructuring of the schema is required to resolve the cognitive dissonance [38]. Teachers can encourage accommodation in mathematics by choosing tasks and activities within the range of their students' assimilation abilities but which have elements that introduce some degree of cognitive dissonance [38]. From the above example, students who encounter multiplication examples that result in products smaller than the original numbers must accommodate the new information when it does not fit their current understanding.

The constructivist perspective requires substantial shifts in traditional educational practice, such as decentering teacher authority, valuing social contexts, and emphasizing students' natural curiosity [37]. Reformed teaching is founded upon constructivist epistemology, including lesson pedagogy and a classroom culture that supports change [39]. Reformed teaching is typically inquiry-based, meaning that students engage in exploration and experimentation prior to a formal presentation. The National Research Council [40] summarized scientific inquiry through eight practices:

1. Asking questions;
2. Developing and using models;
3. Planning and carrying out investigations;
4. Analyzing and interpreting data;
5. Using mathematics and computational thinking;
6. Constructing explanations;
7. Engaging in argumentation from evidence;
8. Obtaining, evaluating, and communicating information. (p. 42)

The term "practice" is used to emphasize that students must simultaneously coordinate knowledge and skill [40]. The expectation for inquiry-based teaching is that students will themselves engage in the practices and not merely learn about them secondhand. By treating mathematics as a scientific endeavor, teachers promote the building of abstract knowledge from simpler, concrete experiences, and student explorations precede formal presentations. Students engage in predictions, hypotheses, and estimation as well as designing experiments to test their conjectures. Students engage in constructive criticism of one another's ideas [39]. These pedagogical approaches directly support constructivist views of learning by building new knowledge from pre-existing knowledge in learning communities and tapping into students' natural curiosity.

3. Methods

This classroom study followed PDSA cycles [18], which provided a structure for multiple classroom trials with refinements at each iteration. The trials were carried out in four high school geometry classes, each with 30 to 35 students in a mid-Atlantic urban school district in the U.S. The four geometry classes were co-taught by the second and third authors of this paper. By "co-taught," we mean that both teachers were involved in the design of the lessons. The teacher candidate led the enactment of the lessons with the mentor providing support, observing and taking notes, and providing feedback on both the design and enactment of the lessons.

3.1. PDSA Cycles in the NIC

The teachers in the present study were part of a networked improvement community (NIC). The overarching problem of practice was focused on how to improve mathematics teaching through inquiry. PDSA cycles provided the structure for participants to plan, enact, reflect, and refine a change idea (teaching strategy) that was decided upon during a monthly NIC meeting. The classroom artifacts, data and evidence, and participant reflections were brought back to the subsequent NIC meeting. The NIC then refined the overall strategies as a group based on the participant reports. The PDSA classroom research process mirrors design-based research in that the innovations and techniques evolve through each iteration.

The NIC met monthly throughout an entire school year. Participants developed their problem of practice during the fall semester (September through December) and tried out their initial change ideas. By the beginning of the spring semester (January through May), the change idea had been refined and was ready for more intensive try-outs. Participants completed at least one PDSA form each month, which represented a variable number of PDSA cycles. While lessons throughout the school year were affected by the PDSA cycles, the lessons presented in the present study represent the culmination of the teachers' reflections and refinements.

3.2. Data and Measures

Data consisted of classroom documents (e.g., activity prompts, tasks, assessments), classroom observations, student feedback about activities, and monthly PDSA reports. Student views were collected through classroom discussions, informal student interviews, and open-ended survey questions. Both the mentor and candidate took notes and observed student behaviors during lesson activities. Teacher views were collected through interviews and PDSA forms.

The degree to which the candidate's teaching improved in terms of reform-based teaching was measured through the formal observations of a field experience supervisor (not an author) and the mentor and scored on the Reformed Teaching Observation Protocol with equity-based performance descriptors (RTOP-E). With 25 indicators on the RTOP-E, each indicator is rated from 0 (no evidence) to 4 (fully reformed practice) for a possible total of 100 points. Level 2 performances are considered to be more traditional with some reformed elements, and Level 3 performances are considered to be more reformed with some traditional elements.

The RTOP-E was based on the RTOP+ [39,41,42] and explicitly incorporated the equitable teaching practices described in *Catalyzing Change in High School Mathematics* [43]. For example, Row 1 was revised to include students' cultural identity (RTOP-E new text italicized): "The instructional strategies and activities respected students' *cultural identity and* prior knowledge and the preconceptions inherent therein." Performance descriptors were revised to include expectations of equity explicitly, especially at Levels 3 and 4 of the rubric. For example, Row 1, Level 3, on the RTOP-E stated, "The teacher actively solicits student ideas *and cultural experiences*, and discussion of these ideas and *experiences* takes place throughout the lesson, but lesson direction is teacher determined". Level 4 stated, "The teacher actively solicits student ideas *and cultural experiences* and builds the lesson from these ideas *and experiences* as a starting point. The direction of the lesson is shaped by student ideas *and experiences*". The revisions were made by a team of mathematics and STEM faculty, then shared with an expert panel for feedback to enhance content validity. The RTOP-E indicators and performance descriptors were used in monthly NIC meetings to structure conversations about effective pedagogy. These conversations included scoring sample lesson videos on the RTOP-E and supported a common understanding of the measured constructs and performance descriptors (construct validity) and how to score the RTOP-E consistently (inter-rater reliability).

The supervisor and mentor independently scored three lessons, one near the end of the Phase I internship in November, one at the beginning of the Phase II internship in February, and one at the end of the Phase II internship in April. The supervisor and mentor scores were the same for 53/75 scores (70.6%) and were adjacent (a difference of 1) for 14/75 scores (18.7%), meaning that they were in agreement (exact or adjacent) for 89.3% of the indicators. The intraclass correlation (ICC) was 0.704, which was considered good based on Cicchetti's criteria [44]. For simplicity, the supervisor's scores were used in the present analysis.

4. Results

The NIC brought participants (mentors, teacher candidates, field experience supervisors, faculty, and alumni) together monthly to discuss the program challenge space, classroom change ideas, and strategies for implementing a change idea. The challenge space was developed by a team of participants and was organized by teacher knowledge, teacher orientation, teacher practice, and student outcomes (see Appendix A). Participants were invited to help to plan monthly NIC meetings and agree upon a focus within the challenge space. The NIC meetings focused primarily on the teacher practice category of the challenge space, especially using reform-based teaching practices [26] and the RTOP-E as a framework to discuss various challenge space goals.

Teachers in the NIC focused on the problem of practice of how to build connections between new mathematics content and students' pre-existing knowledge and experiences. One change idea that the mentor and teacher candidate explored was the use of an inquiry-based activity process to engage student pre-existing knowledge to build new understanding. As part of the "Plan" for PDSA cycles, and based on Watson [45] and Lamar and Boaler [46], it was hypothesized that a data collection inquiry process would facilitate student engagement in inquiry-based activities such as those described by Anderson et al. [47] and Engle and Conant [48]. It was also hypothesized that this type of engagement would improve learning of mathematics concepts that are typically taught procedurally at the high school level in the U.S. (for example, mathematical formulas and their proofs) [36]. By "engagement," we mean that students attempted at least one lesson activity, task, question, or problem.

4.1. Teaching Mathematics through Inquiry

The approach to inquiry in this setting began with student data collection and pattern analysis as a scaffolded entry to theoretical concepts. Through student discussions and debating of ideas, students were able to engage meaningfully with the material and continue developing their conceptual understanding.

While this process may seem fairly straightforward to those familiar with inquiry, many mathematics curriculum materials in the U.S. are not written in a way that supports student-led inquiry. Traditional mathematics teaching is not inquiry-driven, focusing instead on practicing procedures with a notable absence of mathematical reasoning [36,43,49–51]. In many ways, the teachers were "starting from scratch," determining how to adapt their curriculum to be a more robust learning experience for their students, especially those who struggled. The PDSA cycles provided them with a structure to organize their own learning of how to teach through experimentation, reflection, and adjustment. Table 1 presents an example plan developed during an NIC meeting.

Table 1. Example plan for PDSA cycles.

Prompt	Response
Challenge or goal of this PDSA cycle.	Collaboration with data collection
Context (e.g., grade level, course, topic)	10th grade, geometry, density
Expected duration of this PDSA cycle. (e.g., 10/15 min).	One lesson, modeling data collection will be at the beginning of the lesson
Change idea or strategy for meeting your challenge	Model data collection before the students collect their own data
Prediction(s)/hypotheses (What you think the change idea/strategy will accomplish?)	Modeling the data collection will show students how to complete procedures. Avoid confusion when starting the collaboration and data collection
Evidence to collect	Student work and student ability to complete data collection on their own/with minimal help from the teacher

The inquiry process developed through the PDSA cycles began with student data generation and development of a question about a phenomenon rather than procedures to be memorized, consistent with Anderson et al. [47]. Students gathered data, looked for patterns, drew conclusions, and discussed how to interpret the evidence. Student reflection was followed by reinforcement activities that helped students make connections between their exploration and mathematical procedures. This process addresses the tenets of Engle and Conant's productive disciplinary engagement [48]: using problems to engage students with content, giving students authority to investigate the problems, and facilitating their exploration with relevant resources and support. Lessons that use this process will engage students primarily in the Common Core Mathematics Practice #7, Look for and Make Use of Structure, but may also address Practice #4, Model with Mathematics [51]. The modeling of data collection shown in Table 1 was an important component that was added and refined during the PDSA cycles in response to student feedback. As cycles were completed, the teachers refined the change idea into a general process, shown in Figure 2.

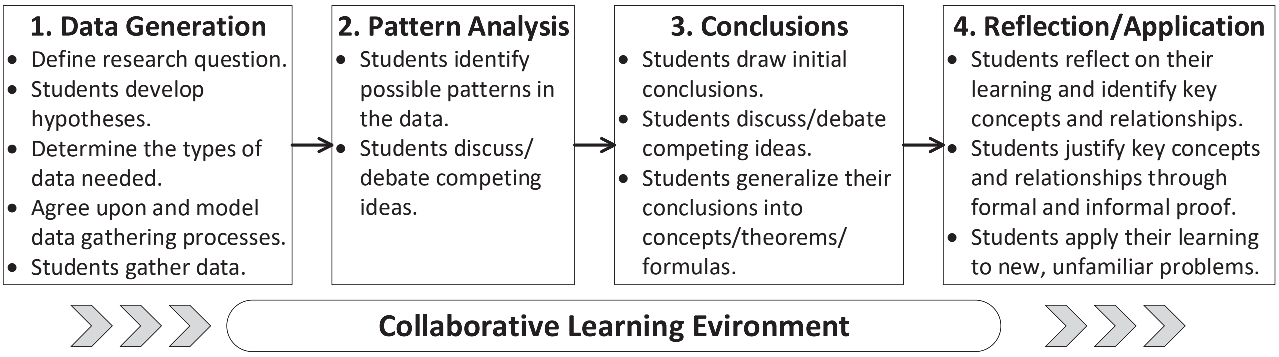


Figure 2. Inquiry learning process in a data collection context developed through PDSA cycles.

Two activities illustrate the inquiry process from Figure 2 and how the PDSA cycle process enhanced the lessons. The first activity, Inquiry into Trigonometric Ratios, focused on the development of deep connections between the various trigonometric functions. The second activity, Population Density, used a culturally relevant approach to developing conceptual understanding of density. Candidate and mentor reflections and notes, student feedback, and independent classroom observation notes were incorporated into the activity descriptions.

4.1.1. Example PDSA Lesson 1: Inquiry into Trigonometric Ratios

In this introductory lesson to trigonometric ratios, the objective was: *Students will be able to explain the relationship between sine and cosine of complementary angles verbally and algebraically*. We (mentor and student teacher) began the lesson by modeling a separate, simpler trigonometric concept with the goal of teaching students how to use the trigonometric functions on an online calculator. Students used the Desmos Graphing Calculator [52], which includes all six trigonometric functions. Student perceptions of the calculator component of the lesson, collected through a classroom survey, were mostly positive, with some students explicitly stating that they "liked using the calculators to solve problems". This whole-class introduction asked students to generate data by choosing angle values between 0 and 90 degrees then filling out the table in Figure 3.

Angle	Sin(A)	Cos(A)	Tan(A)	Sec(A)	Csc(A)	Cot(A)	1/Sec(A)	1/Csc(A)	1/Cot(A)

Figure 3. Introductory table for calculator exploration of trigonometric relationships.

After students completed the table, the teachers asked them to notice and wonder about the values they found [53]. Notice and Wonder is a method of open-ended pattern exploration that encourages students to look for whatever patterns they can find and ask questions about anything confusing. Students were able to identify that $sin(A) = \frac{1}{csc(A)}$, along with the other reciprocal identities. The class used these observations to write equations describing the relationships between all six trigonometric functions.

Once they had completed this introductory activity as a class, students were given another chart to use for data collection (Figure 4). One row of values was provided as an example for students to use as guidance.

Angle A	Sin(A)	Cos(A)	B=90 – A	Sin(B)	Cos(B)
51	0.777	0.629	49	0.629	0.777

Figure 4. Follow-up exploratory table to develop sine and cosine relationships.

We allowed students to choose whether to work independently or collaboratively. Most chose to work collaboratively. Based on task assessment and teacher reflections, some students struggled to understand the notation in the column titles. Many students began the activity by asking us what went in each column rather than interpreting the notation at the top of each column. Instead of simply answering these questions, we asked students to look at the notation and take an "educated guess" as to what we were looking for, then we asked guiding questions until they figured it out, for example, "What does the title of this column tell us to do? What does ''sin(A) mean? What is A?" Such productive struggle was embraced because it ended up helping them to build a stronger understanding of variable meaning and substitution. By the end of the activity, most students were referencing the notation and interpreting what each column required computationally, completing the charts without teacher support.

Students then engaged in another Notice and Wonder activity without teacher guidance [53] as a way to help students move deeper into pattern analysis (Step 2 in Figure 3). The candidate and mentor observed that students readily noticed columns that were identical in value. They also used the column headings to write equations describing the relationship between the sine and cosine of complementary angles, for example, noticing that sin$(A) = \cos(B)$ when m∠A + m∠B = 90°. Based on classroom survey data, students found this part of the lesson intriguing; for example, students stated, "sin/cos = tan was very interesting to learn."

Student feedback on the survey was mostly positive, and students were able to meet the lesson objective. As part of the Act step in the PDSA cycle, we considered ways to improve the lesson going forward. Several students on the survey noted that the numerical analyses were difficult, with statements such as "I didn't like all the numbers", "I didn't like looking at all the numbers and getting mixed up", and "The numbers being different, sometimes it confused me, thinking I was wrong". We realized upon reflection that this data collection method was too separate from the tangible work we had been doing with triangles in class prior to the lesson. We also noted that, in an introductory lesson to trigonometric ratios, greater emphasis needed to be placed on the reference angle. In subsequent lessons, students struggled to transfer their learning from this lesson to the triangle contexts, which supported our analysis of the lesson.

Upon reflection during a PDSA cycle, it was determined that an exploration that includes a visual representation offers a way to enhance these kinds of connections and emphases with students. For example, a Geogebra app such as the one shown in Figure 5 allows students to discover that, regardless of the size or orientation of the triangle, the ratios of the side lengths stay constant.

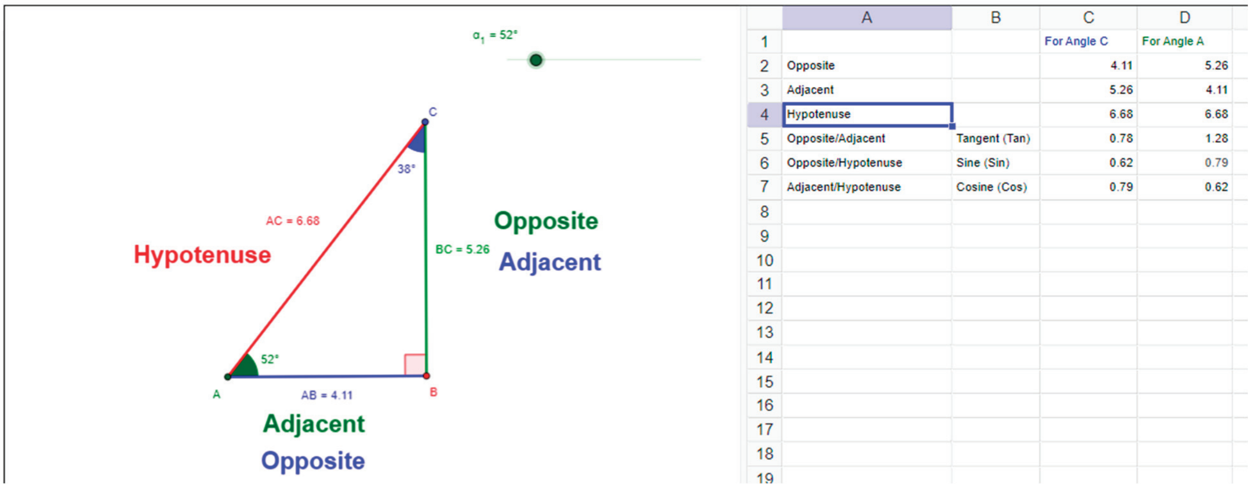


Figure 5. Screenshot of the Trigonometric Ratio Geogebra app [54].

The slider a1 defines the measure of ∠A. Points A and B can move to change the side lengths but not the measures of the angles. Point C is fixed to maintain the right angle at Point B. As Point A and Point B are moved, the side lengths change, but the angle measures and ratios of the side lengths remain constant. With the slider, students can discover that even a slight change to the angle measure changes the ratios of the side lengths. The color coding of the segments and text in the image helps reinforce how the labels of opposite and adjacent are specific to the angle of interest. The app includes sample questions that teachers can use to guide students through the exploration process to discover the one-to-one relationship between angle measures and trigonometric ratios. For example, students can move Point A to several new positions, as shown in Figure 6.

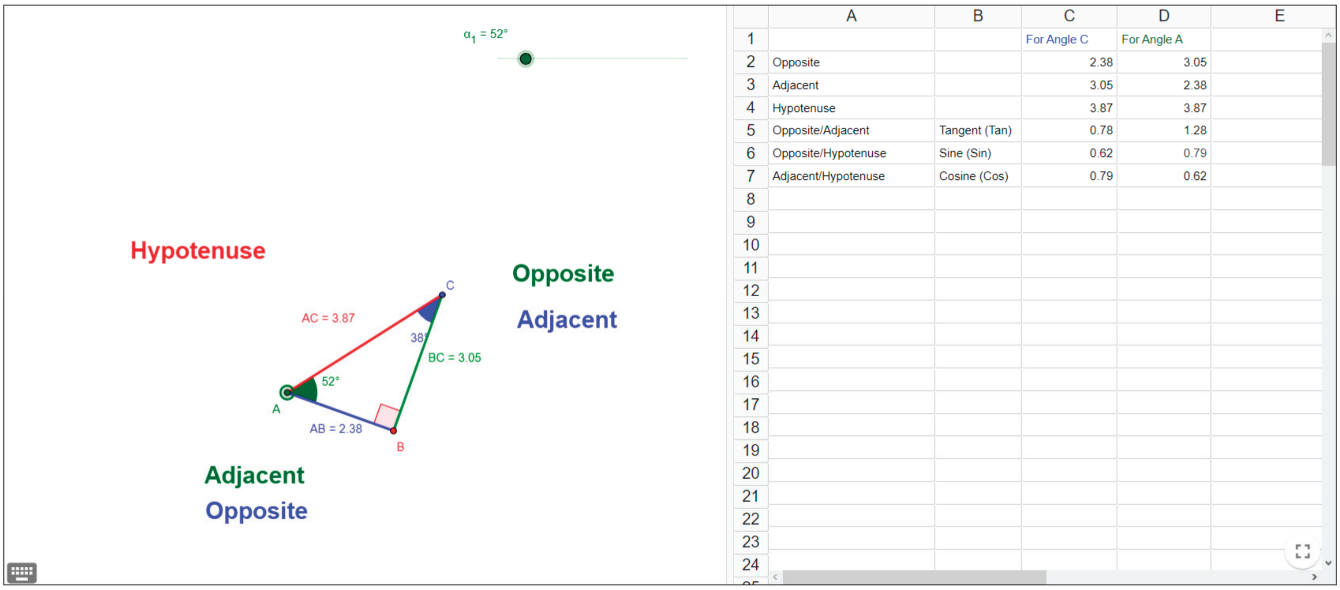


Figure 6. Moving Point A in Figure 4 shows different side lengths but constant trigonometric ratios [54].

By contrast, any change in the angle measure will change the trigonometric ratios, as shown in Figure 7. The sample questions provide a pre-planned experiment. Once students are already familiar with experimentation, the activity can be modified to allow them to design their own experiment.

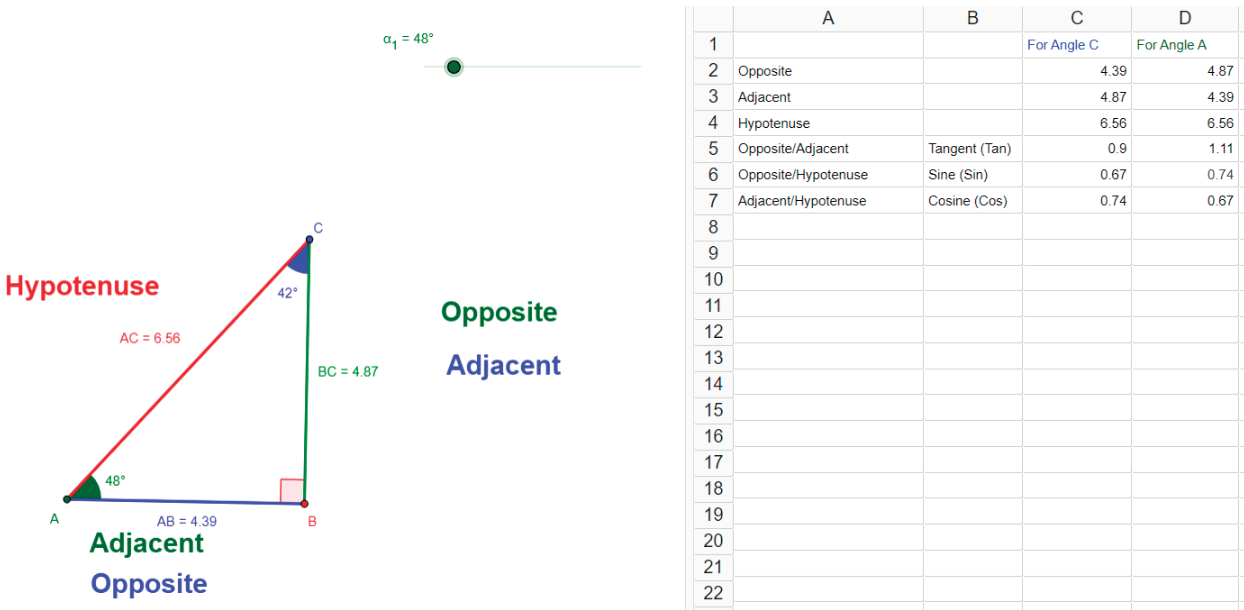


Figure 7. Changing the slider changes the measures of Angle A and Angle C and the trigonometric ratios [54].

Through such an exploration, students are able to make predictions/hypotheses about the triangles and side length ratios and conduct an experiment. This type of applet could be used as an extension of the original exploration but would be more powerful if it were used to create the initial data for the calculator activity (Columns 1–4 in Figure 3). Once students are able to connect the triangle ratios to the values for sine, cosine, and tangent, they will likely be more ready to investigate the secant, cosecant, and cotangent functions (and their relationships to sine, cosine, and tangent) with a calculator.

4.1.2. Example PDSA Lesson 2: Exploration of Population Density

Building from the trigonometry lesson, our PDSA change idea of using student inquiry through data exploration was applied to a geometry lesson about density. Density as a concept is often taught conceptually, usually with a real-world tie to physics that is difficult for some students to grasp when learning about density for the first time. Rather than introducing density through a physics application, we chose to take a more general approach, describing density as simply "an amount of stuff in an amount of space." This approach opened doors for us to explore density in a multitude of ways that included, but was not restricted to, an amount of mass in a specific volume. Using population density as an entry point, students collected data (Step 1 in Figure 2) and used their data to design and conduct a research investigation.

We began the lesson by prompting students to recall previous examples of density that we had worked with in class. We also recalled the framing:

$$\text{Density} = \frac{\text{stuff}}{\text{space}} \tag{1}$$

With some prompting, students were able to generate an equation for population density:

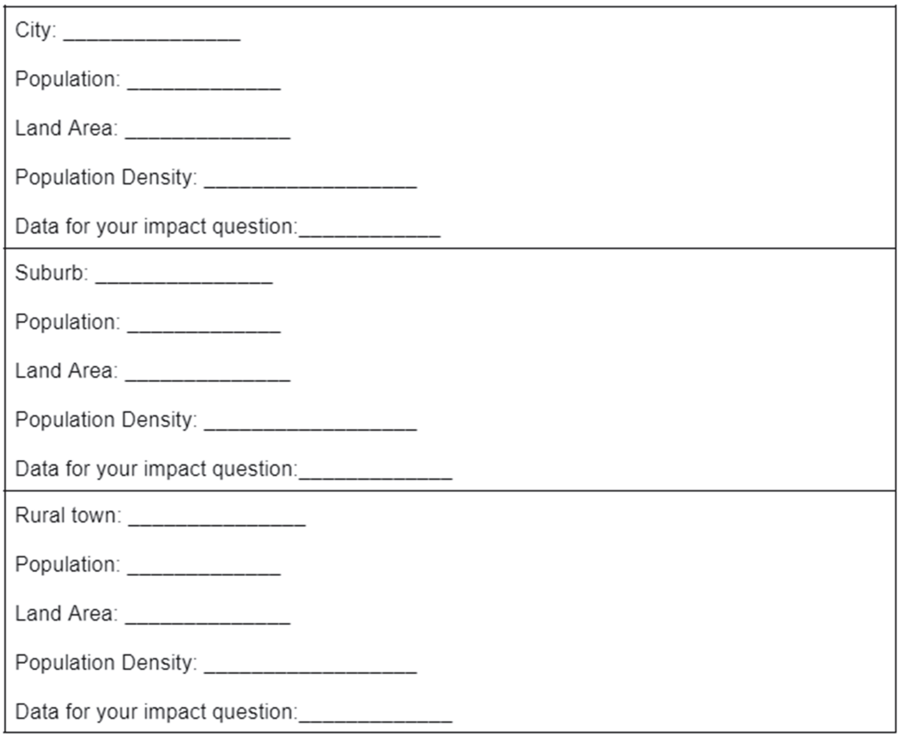
$$\text{Population Density} = \frac{\text{Population}}{\text{SquareMiles}} \tag{2}$$

Based on the PDSA cycle reflections from the trigonometry lesson, we included teacher modeling of the data collection process by finding the population density of Baltimore City, where our school is located. As a class, we used an online search engine to find recent data on the population of Baltimore and the land area of Baltimore, and then we substituted this into our equation to find the population density.

We asked students to reflect individually on why population density might be important to them or other members of the community and pose a question to explore. Having students pose research questions and design their own investigations moved our overall change idea deeper into reformed teaching from the trigonometry lesson, from a class exploration of a phenomenon to a student-led, open-ended inquiry (as described by Sawada et al. [39]). This trajectory was purposeful in the evolution of our change idea: as Blair [55] noted, teachers may restrict an inquiry's activity in the hope of engaging the whole class. Rather than incorporating other inquiry pathways later in the trigonometry lesson (more consistent with Blair's un-planning process), we opted to instead incorporate more inquiry pathways in the density lesson. This approach allowed us to continue enhancing our ability to conduct inquiry-based lessons without falling behind in the required district curriculum.

Most student research questions in this lesson filled the sentence frame "How does population density affect ______________?" Topics chosen by students included police interactions, number of schools, and commute times. Students then researched on the internet to find the population density of three locations. Ideally, these locations spanned different geographic areas, including a city, suburb, and rural area. They also found a data point related to their research question. They used their data to fill out the prompts shown in Figure 8.

Lastly, students completed a reflection question in which they answered their original research question based on the data they collected. Most students reflected that they had enjoyed this lesson more than usual, and some students displayed a deep interest in their research questions. This topic sparked interest from students that had typically had trouble focusing in class. Students enjoyed picking their own research questions as well as collecting their own data. Some students struggled to find data, which led to some discussion on how to research and what questions to type into an online search engine to find the data we are looking for.



City: _______________

Population: _____________

Land Area: ______________

Population Density: __________________

Data for your impact question:____________

Suburb: _______________

Population: _____________

Land Area: ______________

Population Density: __________________

Data for your impact question:_____________

Rural town: _______________

Population: _____________

Land Area: ______________

Population Density: __________________

Data for your impact question:_____________

Figure 8. Example population density student data information form.

4.2. Participation in Lessons Affected by PDSA Change Ideas

The PDSA cycle outcomes reported here were conducted in four high school geometry classes. Student participation in the lessons was generally higher in the PDSA-affected lessons ("PDSA lessons" hereafter) than in comparable lessons before and after. Participation was operationalized as either engagement (at least one activity, task, question, or problem) or full participation (completing all independent work). For comparability, lessons were chosen that required independent work using an online district platform (ImagineMath [56]) intended to increase student accessibility and participation in the lessons. Table 2 provides an example of the numbers and percents of students who participated during one PDSA lesson compared with ImagineMath Lessons, before and after.

Table 2. Numbers and percents of students that participated in PDSA and comparison lessons.

		PDSA Lesson (Inquiry through Trigonometric Ratios)		Comparison Lessons: No (and Percent) That Engaged in Any Work (at Least One Question)		
Class	**No. Students**	**No. (and Percent) That Completed Independent Work**	**No. (and Percent) That Engaged in Any Work (at Least One Question)**	**Lesson 1 before PDSA Lessons**	**Lesson 2 after PDSA Lessons**	**Lesson 3 after PDSA Lessons**
1	22	8 (36.4)	13 (59.1)	2 (9.1)	2 (9.1)	3 (13.6)
2	27	18 (66.7)	21 (77.8)	9 (33.3)	12 (44.4)	18 (66.7)
3	29	15 (51.7)	16 (55.2)	3 (10.3)	10 (34.5)	5 (17.2)
4	31	19 (61.3)	20 (64.5)	3 (9.7)	12 (38.7)	15 (48.4)

Based on task assessments, participation was generally higher in the example PDSA lesson. The candidate and mentor compared participation rates for individual students and found that students with consistently low participation rates had higher participation rates in the PDSA lessons. While there are many potential contributing factors to participation rates in a lesson, student feedback on the classroom survey was also quite positive for the approach of examining data. For example, multiple students stated in a class survey, "I liked picking my own numbers." Students considered the lessons to be more accessible.

4.3. Teacher Candidate Growth in Reformed Teaching

Lessons led by the teacher candidate were scored three times by a field experience supervisor. The supervisor's observations provided an independent measure of her ability to use reform-based teaching methods. The lesson observed at Time 1 was an exploration of rotational symmetry. The lesson observed at Time 2 was an introduction to identifying trigonometric ratios on a right triangle, a precursor to the trigonometric pattern exploration described in Section 4.1.1 above. The lesson observed at Time 3 was the density lesson described in Section 4.1.2 above. As shown in Figure 9, most of the growth occurred during the Phase II Internship, which is the full-time student teaching semester.

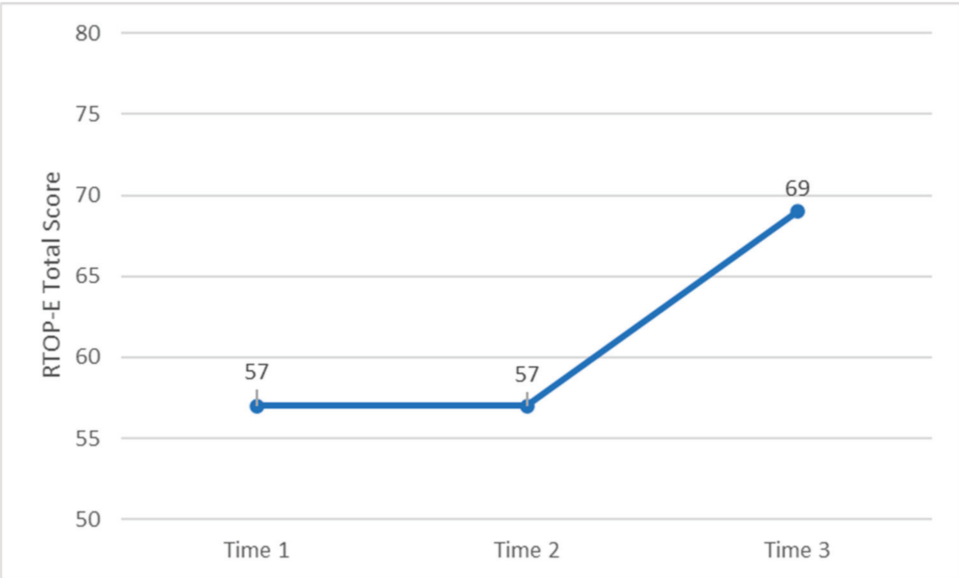


Figure 9. Teacher candidate RTOP-E scores. Time 1 = Phase I internship, mid-November. Time 2 = Beginning of Phase II internship, late January. Time 3 = Second half of Phase II, late March.

This growth trend is consistent with prior cohorts [42] and comparable to the cohort means at each time point. A repeated-measures ANOVA (RM-ANOVA) was used to analyze differences across time. Mauchly's test of sphericity indicated that the assumption of sphericity was not violated, $W(\text{df} = 2) = 0.921$, $p = 0.386$. The RM-ANOVA showed that the growth was significant, $F(2, 48) = 3.197$, $p = 0.05$.

The supervisor noted at Time 1 that the lesson included multiple opportunities for student collaboration and discussion within groups. He noted, "[the candidate] comfortably discussed the activity with various groups and visited all of the groups during the period." Issues to be addressed focused on classroom management issues and equitable access to technology used in the lesson and equitable participation of students in whole-class discussions.

At Time 2, the supervisor noted that the lesson acknowledged students' cultural perspectives. Students were given opportunities to lead discussions. Issues to be addressed focused on suggestions for multiple ways to represent and clarify the trigonometric reference angle.

At Time 3, the supervisor remarked on several strengths of the lesson, especially its cultural relevance: "the ability to make mathematics relevant to students and show how it can be used to plan for living in various environments, urban and suburban. Relating mathematics to different content areas such as urban planning. Allowing students to use research in the development of mathematical concepts." Issues to be addressed included planning for students to write a summative paragraph of their findings and to share their work in class.

5. Discussion and Conclusions

The present study focused on how PDSA cycles support pedagogical innovation in the classroom (Research Question 1) and how reform-based teaching can be transferred from theory into practice during full-time student teaching (Research Question 2). The PDSA cycles provided a structure for improving pedagogy in subsequent lessons. The results showed that the integration of candidate and mentor observations and reflections, student feedback, and independent observation feedback provided the data needed for the candidate to improve the way inquiry was used in the lessons. The RTOP-E scores demonstrated that measurable growth in the candidate's use of reform-based teaching was observed by the supervisor and mentor.

The PDSA cycles for this project focused on engaging students in pre-exploration and collaborative discussions. Student participation was noticeably improved during these activities compared with traditional mathematics instruction. For example, the candidate and mentor observed that students who frequently gave up on exploratory activities instead engaged in productive struggle. The strong connections to their local community in the density project led to student excitement about the mathematics, expressed to the candidate and mentor through the classroom survey and informal class discussions. As noted in Section 4.1.2, students led their own investigations by posing their own questions and designing their own experiments. The supervisor noticed that students led more of the classroom discussions in this lesson compared with prior observed lessons. The willingness of students to take on more responsibilities during the trigonometry activities (e.g., leading classroom discussions) surprised the teachers and spurred them to give more responsibility for the learning to the students in the density lesson (e.g., picking their own research questions).

Based on the improved participation rates, we concluded that the student data collection sequence provided an accessible entry point to begin scientific inquiry in mathematics. The process provided an opportunity for students to develop curiosity about mathematical phenomena and to explore their own research questions. Such open-ended opportunities are sometimes described as "opening the conceptual space" (e.g., Niesser et al. [57]) because they allow students to understand the content in multiple ways and through multiple perspectives rather than through a narrow interpretation provided by a lecturer. Such an approach allows students to analyze conceptions that are partially correct and determine whether such conceptions are valid in various contexts [58]. In the present study, students analyzed data to determine the extent to which the patterns they noticed held true. According to survey results, the use of culturally relevant topics was especially compelling to students, and the open-ended nature of these exploratory activities allowed students to see mathematics through their own cultural lenses.

As shown in Table 1, the PDSA cycles provided a structure in which both mentor and candidate could direct their own professional learning. The NIC meetings provided a monthly forum in which the mentor and candidate explored mathematics pedagogy with a community of educators, planned strategies for improving their classroom practice, and received feedback on their change ideas. Between NIC meetings, the mentor and candidate completed multiple PDSA cycles and wrote up the results and reflections on a PDSA form. Both mentor and candidate found that the PDSA cycles provided structure to their daily reflections, especially the explicit focus on planning to collect data about outcomes resulting from the change idea. We concluded that the PDSA process opened up communication between the teacher candidate and the mentor to mutually support their professional learning as they enacted, studied, and refined their pedagogical change ideas.

The mentor and supervisor both noticed in their observations that the teacher candidate developed stronger classroom communication skills; for example, the ability to clearly convey mathematical ideas to students (e.g., using multiple representations as noted at

Time 2 observation). The mentor found that she became more acclimated to shifting responsibility for thinking to her students, which was also shown in the supervisor feedback. She also noticed an increase in the number of lessons developed by the candidate that focused on discovery, exploration, and inquiry rather than processes and procedures outside the formally observed lessons.

The PrimeD framework guided the teacher preparation program, structuring the program challenge space and directing the process for developing pedagogical change ideas as a professional community and testing those ideas in specific classroom contexts. PrimeD is recommended as one way to structure teacher preparation to facilitate professional learning. The results of the present case study provide encouraging results for teachers to use data collection and inquiry activities to frame mathematics as a vibrant, interesting, and relevant scientific endeavor.

Author Contributions: Conceptualization and Methodology, C.R.R., A.W. and R.L.; Project Administration, C.R.R., A.W. and R.L.; Writing-Original Draft, C.R.R. and A.W.; Writing-Review and Editing, A.W. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation, grant numbers 2013250, 2013397, 2013266, 2013256.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of University of Maryland Baltimore County (protocol code Y17CR24115, approved on 2 July 2020).

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: The authors would like to thank Robert Ronau for his feedback on earlier versions of this manuscript.

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

Appendix A.1. Secondary Mathematics Challenge Space

Development of the challenge space is ongoing and includes input from classroom mentor teachers, mathematics coaches and supervisors, and teacher candidates and program completers. The goals of the program are viewed through four constructs: teacher knowledge, teacher orientation (e.g., attitudes, beliefs, self-efficacy), teacher practice, and student outcomes. Teacher knowledge and orientation influence each other and inform teacher practice. Teacher practice includes reflection on student outcomes, thereby reinforcing or refining teacher knowledge and orientation and informing the program (i.e., a feedback loop).

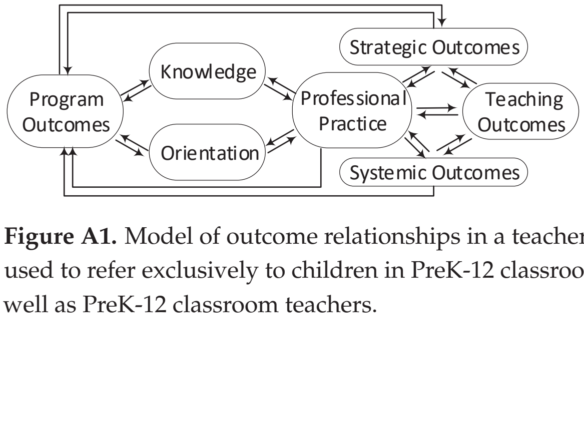


Figure A1. Model of outcome relationships in a teacher education program [42]. Note: "Student" is used to refer exclusively to children in PreK-12 classrooms; "Teacher" refers to teacher candidates as well as PreK-12 classroom teachers.

Appendix A.2. Vision

The mathematics teacher preparation program is designed to help candidates explore, enact, and insist upon equitable teaching practices to support robust mathematics learning communities. Learning communities are defined as collaborative groups that are pursuing common goals for mathematics learning experiences. Learning communities that are safe, humanizing, collaborative, and culturally aware empower participants to direct their engagement in scientific inquiry and the examination of diverse ideas and perspectives. Candidates enter the field ready to improve the quality of mathematics teaching and learning for each and every student in their classrooms, schools, and districts and to become societal change agents in the field.

Appendix A.3. Goals

Teacher Knowledge

- Subject Matter Knowledge. Teachers have a robust knowledge of mathematics, understanding how concepts and procedures are interrelated and how to frame mathematics knowledge in a meaningful way to help students learn (Mathematics Knowledge for Teaching).
- Pedagogical Content Knowledge. Teachers develop robust pedagogical knowledge to support deep mathematics learning in their classrooms, including the use of tools for teaching mathematics (Knowledge for Teaching Mathematics).
- Knowledge of Orientation. Teachers understand and respect the relevance of the affect of each member of a learning community (e.g., attitudes, culture, beliefs, values, confidence, and anxiety) in learning mathematics.
- Knowledge of Discernment. Teachers understand that discernment encompasses the connections between cognition, metacognition, and learning and decision-making processes. Knowledge of discernment includes understanding developmental processes and the socio-emotional and sociocultural components of learning.
- Knowledge of Individual Context. Teachers understand that learning and decision-making processes take place within the context of the intersectionality of social categories.
- Knowledge of Environmental Context. Teachers understand the importance of building an inclusive and equitable environment to support a robust learning community.

Appendix A.4. Teacher Orientation

Orientation plays an important role in how teachers approach the profession individually as well as in collaboration with students, colleagues, schools, and the community. Orientation includes, but is not limited to, constructs such as attitudes, perceptions, self-efficacy, beliefs, confidence, self-concept, motivation, value of mathematics, interest in mathematics, enjoyment of mathematics, enjoyment of teaching, usefulness of mathematics, mathematics goals, professional goals, attributions of success/failure, mathematics anxiety, professional anxiety, professional dispositions, commitment to lifelong learning, and perceptions of power and agency.

These orientations can be about a wide range of topics, including, but not limited to, mathematics, teaching and learning, assessment, students, socio-cultures, families and caregivers, collaboration, the profession, and schools and districts.

Teachers examine orientation as an ongoing part of their growth and learning to ensure that all aspects of the profession are approached through a productive lens. Teachers are willing to change their views when appropriate.

Appendix A.5. Teacher Practice

- Culture. Teachers establish a culture of access and equity through classroom structures and culturally relevant pedagogy to support each and every student in learning and participating in mathematics deeply. These classroom structures empower students to value diverse perspectives by elevating their voices, providing leadership opportunities, and developing a strong learning community. Teachers model vulnerability, viewing mistakes as learning opportunities. Varied approaches are visible and valued.

- Active Engagement. Teachers actively engage students in learning mathematics and/or science with meaning.
- Conceptual Understanding. Teachers explicitly foster, model, and insist upon conceptual understanding and coherence for all learners at all levels as a primary means for promoting procedural understanding in mathematics. Teachers insist that all teaching activities and learning experiences embrace the development of conceptual understanding as the fundamental core of learning and form the foundation for peer discussions.
- Connections. Teachers structure lessons through a phenomena-first approach, recognizing that authentic contexts are the foundation of the lesson and frame the content to be learned. Contexts are not simply enrichment that happens after the "real" lesson if at all.
- Reasoning. Inquiry-based projects are incorporated in every unit. Quantitative reasoning is modeled as scientific inquiry (claim, evidence, rationale).
- Questioning. Questioning is purposefully crafted to foster higher-order thinking and alternative modes of thinking about mathematics. Teachers pose questions of their students and encourage their students to ask deep, rich questions about their mathematical reasoning and that of their peers.
- Assessment. The ability to provide students feedback through formative (ongoing) and summative (reflective) assessment is differentiated from and valued more than grades. Assessments are ongoing, are aligned to standards, and (in)form teacher practice. Teachers understand that assessment can take many forms including formative (ongoing) and summative (reflective) assessment. Teachers incorporate a variety of assessments to ensure that each and every student has an opportunity to express their current understanding, including, but not limited to, observations, student-to-student and student-to-teacher dialogue, projects, performance tasks, interviews, portfolios, presentations, exit slips, and dynamic technology-based activities. Teachers recognize that understanding develops over time and leverage opportunities to reassess throughout the learning process.

Appendix A.6. Student Outcomes

Teachers assess and reflect upon a wide range of student outcomes to inform their practice, such as social and emotional well-being, persistence, goal setting, achievement, thinking/reasoning/explaining, orientation, cognition and meta-cognition, and learning behaviors.

References

1. Darling-Hammond, L. Teacher quality and student achievement: A review of state policy evidence. *Educ. Policy Anal. Arch.* **2000**, *8*, 1. Available online: http://epaa.asu.edu/ojs/article/view/392/515 (accessed on 1 June 2023). [CrossRef]
2. Ginsberg, R.; Rhodes, L.K. University faculty in partner schools. *J. Teach. Educ.* **2003**, *54*, 150–162. [CrossRef]
3. U.S. Department of Education. *Teacher Professional and Career Development*; U.S. Department of Education: Washington, DC, USA, 2018. Available online: https://oese.ed.gov/archived/oii/teacher-professional-and-career-development/ (accessed on 15 July 2023).
4. Gainsburg, J. Why new mathematics teachers do or don't use practices emphasized in their credential program. *J. Math. Teach. Educ.* **2012**, *15*, 359–379. [CrossRef]
5. Morris, A.K.; Hiebert, J. Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *Am. Educ. Res. J.* **2017**, *54*, 524–567. [CrossRef]
6. Taylor, P.M.; Ronau, R. Syllabus study: A structured look at mathematics methods courses. *AMTE Connect.* **2006**, *16*, 12–15. Available online: https://amte.net/sites/all/themes/amte/resources/newsletters/AMTENewsletter_Vol16-1_Fall2006.pdf (accessed on 14 January 2023).
7. Desimone, L.M. Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educ. Res.* **2009**, *38*, 181–199. [CrossRef]

8. Loucks-Horsley, S.; Stiles, E.; Mundry, S.; Love, N.; Hewson, P. *Designing Professional Development for Teachers of Science and Mathematics*, 3rd ed.; Corwin: Thousand Oaks, CA, USA, 2010.
9. Bangel, N.J.; Enersen, D.; Capobianco, B.; Moon, S.M. Professional development of preservice teachers: Teaching in the super Saturday program. *J. Educ. Gift.* **2006**, *29*, 339–361. Available online: https://files.eric.ed.gov/fulltext/EJ746288.pdf (accessed on 14 January 2023). [CrossRef]
10. Pollock, M.; Bocala, C.; Deckman, S.L.; Dickstein-Staub, S. Caricature and hyperbole in preservice teacher professional development for diversity. *Urban Educ.* **2016**, *51*, 629–658. [CrossRef]
11. Garet, M.S.; Porter, A.C.; Desimone, L.; Birman, B.F.; Yoon, K.S. What makes professional development effective? Results from a national sample of teachers. *Am. Educ. Res. J.* **2001**, *38*, 915–945. [CrossRef]
12. Putnam, R.T.; Borko, H. What do new views of knowledge and thinking have to say about research on teacher learning? *Educ. Res.* **2000**, *29*, 4–15. [CrossRef]
13. Jones, K.; O'Brien, J. Introduction: Professional development in teacher education: European perspectives. In *European Perspectives on Professional Development in Teacher Education*; Jones, K., O'Brien, J., Eds.; Routledge: New York, NY, USA, 2014; pp. 1–6.
14. Philippou, S.; Papademetri-Kachrimani, C.; Louca, L. 'The exchange of ideas was mutual, I have to say': Negotiating researcher and teacher 'roles' in an early years educators' professional development programme on inquiry-based mathematics and science learning. *Prof. Dev. Educ.* **2015**, *41*, 382–400. [CrossRef]
15. Sztajn, P. Research commentary: Standards for reporting mathematics professional development in research studies. *J. Res. Math. Educ.* **2011**, *42*, 220–236. [CrossRef]
16. Rakes, C.R.; Bush, S.B.; Mohr-Schroeder, M.J.; Ronau, R.N.; Saderholm, J. Making teacher PD effective using the PrimeD framework. *New Engl. Math. J.* **2017**, *50*, 52–63. Available online: http://hdl.handle.net/11603/14416 (accessed on 15 June 2023).
17. Saderholm, J.; Ronau, R.N.; Rakes, C.R.; Bush, S.B.; Mohr-Schroeder, M. The critical role of a well-articulated conceptual framework to guide professional development: An evaluation of a state-wide two-week program for mathematics and science teachers. *Prof. Dev. Educ.* **2017**, *43*, 789–818. [CrossRef]
18. Martin, W.G.; Gobstein, H. Generating a networked improvement community to improve secondary mathematics teacher preparation: Network leadership, organization, and operation. *J. Teach. Educ.* **2015**, *66*, 482–493. [CrossRef]
19. Wilson, S.M.; Floden, R.E.; Ferrini-Mundy, J. *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations. A Research Report Prepared for the U.S. Department of Education*; Document R-01-3; Center for the Study of Teaching and Policy, University of Washington: Seattle, WA, USA, 2001. Available online: https://www.education.uw.edu/ctp/sites/default/files/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf (accessed on 15 January 2023).
20. Driskell, S.O.; Bush, S.B.; Ronau, R.N.; Niess, M.L.; Rakes, C.R.; Pugalee, D. Mathematics education technology professional development: Changes over several decades. In *Handbook of Research on Transforming Mathematics Teacher Education in the Digital Age*; Niess, M.L., Driskell, S.O., Hollebrands, K.F., Eds.; IGI Global: Hershey, PA, USA, 2016; pp. 107–136.
21. Bryk, A.S.; Gomez, L.M.; Grunow, A. Getting ideas into action: Building networked improvement communities in education. In *Frontiers in Sociology in Education*; Hallinan, M.T., Ed.; Springer: Dordrecht, The Netherlands, 2011; pp. 127–162. Available online: https://link.springer.com/chapter/10.1007/978-94-007-1576-9_7 (accessed on 14 January 2023).
22. Bryk, A.S.; Gomez, L.M.; Grunow, A.; LeMahieu, P. *Learning to Improve. How America's Schools Can Get Better at Getting Better*; Harvard Educational Publishing: Cambridge, MA, USA, 2015.
23. Yeager, D.; Bryk, A.; Muhich, J.; Hausman, H.; Morales, L. *Practical Measurement*; Carnegie Foundation for the Advancement of Teaching: Stanford, CA, USA, 2013.
24. Datnow, A.; Stringfield, S. Working together for reliable school reform. *J. Educ. Stud. Placed Risk* **2000**, *5*, 183–204. [CrossRef]
25. Shulman, L.S. *The Wisdom of Practice: Essays on Teaching, Learning, and Learning to Teach*, 1st ed.; Jossey-Bass: San Francisco, CA, USA, 2004.
26. Anderson, T.; Shattuck, J. Design-based research: A decade of progress in education research? *Educ. Res.* **2012**, *41*, 16–25. [CrossRef]
27. Drill, K.; Miller, S.; Behrstock, E. Teachers' Perspectives on Educational Research (EJ1057457). *Brock Educ.* **2012**, *23*, 3–17. Available online: https://files.eric.ed.gov/fulltext/EJ1057457.pdf (accessed on 14 January 2023).
28. Atay, D. Teacher research for professional development. *ELT J.* **2008**, *62*, 139–147. [CrossRef]
29. Elliott, J. Teachers as researchers: Implications for supervision and for teacher education (ED293831). *Teach. Teach. Educ.* **1990**, *6*, 1–26. Available online: https://files.eric.ed.gov/fulltext/ED293831.pdf (accessed on 14 January 2023). [CrossRef]
30. Savoie-Zajc, L.; Descamps-Bednarz, N. Action research and collaborative research: Their specific contributions to professional development. *Educ. Action Res.* **2007**, *15*, 577–596. [CrossRef]
31. Zeichner, K.M. Teacher research as professional development for P–12 educators in the USA. *Educ. Action Res.* **2003**, *11*, 301–326. [CrossRef]
32. Kirkwood, M.; Christie, D. The role of teacher research in continuing professional development. *Br. J. Educ. Stud.* **2006**, *54*, 429–448. [CrossRef]
33. Brownell, W.A. Psychological considerations in the learning and the teaching of arithmetic. In *The Teaching of Arithmetic*; Yearbook 10; Reeve, W.D., Ed.; National Council of Teachers of Mathematics: Washington, DC, USA, 1935; pp. 1–31. Available online: https://files.eric.ed.gov/fulltext/ED096171.pdf (accessed on 15 January 2023).

34. Colburn, W. *First Lessons: Intellectual Arithmetic, upon the Inductive Method of Instruction*; W. J. Reynolds & Co.: Boston, MA, USA, 1847. Available online: https://archive.org/details/colburnsfirstle00colbgoog/mode/2up (accessed on 15 January 2023).
35. Fey, J.T. Mathematics teaching today: Perspectives from three national surveys. *Math. Teach.* **1979**, *72*, 490–504. [CrossRef]
36. Hiebert, J.; Stigler, J.W.; Jacobs, J.K.; Givvin, K.B.; Garnier, H.; Smith, M.; Hollingsworth, H.; Manaster, A.; Wearne, D.; Gallimore, R. Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educ. Eval. Policy Anal.* **2005**, *27*, 111–132. [CrossRef]
37. Ültanir, E. An epistemological glance at the constructivist approach: Constructivist learning in Dewey, Piaget, and Montessori. *Int. J. Instr.* **2012**, *5*, 195–212.
38. Bybee, R.W.; Sund, R.B. *Piaget for Educators*, 2nd ed.; Charles E. Merrill: Columbus, OH, USA, 1982.
39. Sawada, D.; Piburn, M.D.; Judson, E.; Turley, J.; Falconer, K. Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *Sch. Sci. Math.* **2002**, *102*, 245–253. [CrossRef]
40. National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; Committee on a Conceptual Framework for New K-12 Science Education Standards; Board on Science Education, Division of Behavioral and Social Sciences and Education; The National Academies Press: Washington, DC, USA, 2012. [CrossRef]
41. 41. Williams, T.H.; Singer, J.E.; Krikorian, J.G.; Rakes, C.R.; Ross, J.M. Measuring pedagogy and the integration of engineering design in STEM classrooms. *J. Sci. Educ. Technol.* **2019**, *28*, 179–194. [CrossRef]
42. Rakes, C.R.; Saderholm, J.; Bush, S.B.; Mohr-Schroeder, M.J.; Ronau, R.N.; Stites, M.L. Structuring secondary mathematics teacher preparation through a professional development framework. *Int. J. Res. Sci. Eng.* **2022**, *10*, 194–208. Available online: https://ijres.org/papers/Volume-10/Issue-12/1012194208.pdf (accessed on 15 July 2023).
43. National Council of Teachers of Mathematics. *Catalyzing Change in High School Mathematics. Initiating Critical Conversations*; National Council of Teachers of Mathematics: Reston, VA, USA, 2018.
44. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluation normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [CrossRef]
45. Watson, J.M. Statistics in context. *Math. Teach.* **2000**, *93*, 54–58.
46. Lamar, T.; Boaler, J. The importance and emergence of K12 data science. *Phi Delta Kappan* **2021**, *103*, 49–53. [CrossRef]
47. Anderson, R.D.; Anderson, B.L.; Varanka-Martin, M.A.; Romagnano, L.; Bielenberg, J.; Flory, M.; Mieras, B.; Whitworth, J. *Issues of Curriculum Reform in Science, Mathematics and Higher Order Thinking across the Discipline (ED368064)*; U.S. Department of Education, Office of Educational Research and Improvement, ERIC: Washington, DC, USA, 1994. Available online: https://ia802607.us.archive.org/13/items/ERIC_ED368064/ERIC_ED368064.pdf (accessed on 14 January 2023).
48. Engle, R.A.; Conant, F.R. Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cogn. Instr.* **2002**, *20*, 399–483. [CrossRef]
49. National Council of Teachers of Mathematics. *Catalyzing Change in Middle School Mathematics*; National Council of Teachers of Mathematics: Reston, VA, USA, 2020.
50. Stigler, J.W.; Hiebert, J. Understanding and improving classroom mathematics instructions: An overview of the TIMSS video study. *Phi Delta Kappan* **1997**, *79*, 14–21.
51. Council of Chief State School Officers (CCSSO). Common Core State Standards Initiative: Standards for Mathematical Practice. 2010. Available online: http://www.corestandards.org/Math/Practice/ (accessed on 1 June 2022).
52. Desmos Studio. Desmos Online Calculator. 2022. Available online: https://www.desmos.com (accessed on 12 July 2023).
53. National Council of Teachers of Mathematics. Notice and Wonder. 2023. Available online: https://www.nctm.org/noticeandwonder/ (accessed on 1 August 2023).
54. Rakes, C.R. Trigonometric ratios Geogebra applet. 2021. Available online: https://www.geogebra.org/m/rhs38edz (accessed on 15 August 2023).
55. Blair, A. Planning and unplanning mathematical inquiry. *Math. Teach.* **2020**, *271*, 40–41.
56. Imagine Learning. ImagineMath. 2023. Available online: https://math.imaginelearning.com (accessed on 10 July 2023).
57. Niesser, J.; Abreu, G.; Drane, D.L.; Pedersen, N.P.; Parsons, T.D.; Cleary, A.M. Opening a conceptual space for metamemory experience. *New Ideas Psychol.* **2023**, *69*, 100995. [CrossRef]
58. Sreenivasulu, B.; Subramaniam, R. Mapping the conceptual space formed by students' understaning of coordination number of a transition metal complex: An exploratory study. *Chem. Educ. Res. Pract.* **2019**, *20*, 468–483. [CrossRef]

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Education Sciences Editorial Office
E-mail: education@mdpi.com
www.mdpi.com/journal/education

MDPI

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-4930-7