education sciences

Special Issue Reprint

# Second Language Acquisition and Language Education - Bridging the Interface

Edited by
Martin Howard

mdpi.com/journal/education

MDPI

# Second Language Acquisition and Language Education—Bridging the Interface

# Second Language Acquisition and Language Education—Bridging the Interface

Editor

**Martin Howard**

MDPI

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

*Editor*
Martin Howard
Department of French
University College Cork
Cork
Ireland

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

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

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

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

**ISBN 978-3-7258-1118-2 (Hbk)**
**ISBN 978-3-7258-1117-5 (PDF)**
**doi.org/10.3390/books978-3-7258-1117-5**

# Contents

# About the Editor

**Martin Howard**

Martin Howard is Vice-Head (Global) of the College of Arts, Celtic Studies and Social Sciences at University College Cork. His research is situated within second language acquisition, with particular reference to the acquisition of French, (socio)linguistic variation, the advanced learner and study abroad. He is a past President of the Association for French Language Studies and previously served as Vice-President of the European Second Language Association.

education sciences

MDPI

*Editorial*

# Second Language Acquisition and Language Education—Bidirectional Synergies between Research and Practice

**Martin Howard**

Department of French, University College Cork, T12 K8AF Cork, Ireland; m.howard@ucc.ie

## 1. Introduction

This Special Issue brings together the fields of second language acquisition (SLA) and language education in an attempt to offer a venue for exploring mutual insights into classroom language learning. While there is a natural interface between the two fields, opportunities for a more extensive dialogue are often restricted, with research often holding an independent identity in one field or the other. Notwithstanding any perceived independence, research in both fields carries significant insights for the wide-ranging stakeholders involved in language education, from practitioners to policy makers, as well as learners themselves. These stakeholders reflect that the instructed learning environment is a significant venue of language learning for language learners worldwide. While some previous publications have showcased the potential for a mutually beneficial dialogue, this Special Issue aims to further advance that agenda by presenting a range of studies that exemplify the scope for mutual engagement (for examples of previous publications, see [1–5]). While the scope of the interface between the two fields is extensive, the range of thematic areas covered here is necessarily non-exhaustive, offering contemporary perspectives on different themes of mutual interest across the dual fields. This introduction provides a brief overview of the natural synergies that exist, before presenting a short synopsis of the contributing articles.

## 2. Second Language Acquisition

SLA has been a buoyant field since the 1970s within the wider field of applied linguistics, drawing on a vast range of approaches and methods that offer a multifaceted lens on the processes and outcomes underlying our learning of a language beyond our first language. While a strong historical focus on the learner's language system, interlanguage [6,7], and its developing characteristics has prevailed, the field has necessarily extended its scope to input matters and learner characteristics, reflecting a generally threefold thematic focus. Such thematic remit puts the learner at the heart of SLA research inquiry, offering insight that spans a vast range of questions concerning the development of the learner's second language (L2) repertoire, from processing, comprehension, representation, and knowledge issues to learner production at different developmental stages. The role of crosslinguistic influences, as well as L2 input and interaction and their contribution to such development, constitute significant areas of investigation, as does the significant inter-learner variability that arises in the differential experiences of language learning across individual learners and their levels of attainment. The latter issues concerning individual variability extend to wide-ranging individual, crosslinguistic, psychological, social, and environmental factors that are hypothesised to potentially constrain and shape development. The extensive scope of the field is thus characterised by equally wide-ranging approaches, as exemplified by neurolinguistic, cognitive, psycholinguistic, developmental, social, environmental, and experiential perspectives. Indeed, such an array highlights the necessarily cross-disciplinary and transdisciplinary perspectives adopted and provided. Irrespective of the learner, L2, and learning context, the thematic, methodological, and

**Citation:** Howard, M. Second Language Acquisition and Language Education—Bidirectional Synergies between Research and Practice. *Educ. Sci.* **2024**, *14*, 345. https://doi.org/10.3390/educsci14040345

Received: 4 March 2024
Accepted: 10 March 2024
Published: 25 March 2024

disciplinary scope of the field should carry insight that extends beyond SLA in itself to hold relevance for other stakeholders involved in the language learning enterprise.

The field pertains to all learners of an additional language, be it their first L2 or another learnt contemporaneously or subsequently, at any post-L1 acquisition level and in all learning contexts. Therefore, it is not restricted to an instructed learning environment, although instructed second language acquisition (ISLA) constitutes a significant sub-domain, giving rise to a range of areas of investigation (for a presentation, see [3,8,9]). Learning context has traditionally been considered in terms of a dichotomy between instructed learning in the foreign language classroom and naturalistic learning in the target language community. However, the alternation and complementarity between contexts are evident, such as in the case of naturalistic learners availing of language classes and instructed learners spending periods of time in the target language community during study abroad [10]. Moreover, even without physically venturing into the target language community, instructed learners have increasing opportunities to access the language outside the classroom via different means [11]. Moreover, even within an instructed environment, different types of instruction are available [12], with domestic immersion approaching some characteristics of naturalistic learning to varying degrees [13,14].

While a natural universality may be hypothesised to underpin acquisition processes across individual learners [15], the potential role of the learning context in impacting developmental processes and outcomes calls for an understanding of how contextual factors shape L2 acquisition. While a narrow conceptualisation of processes focuses on internal, cognitive, psycholinguistic perspectives, it is important to be cognisant that processes also pertain to social, cultural, and interactional aspects, among others, of language development and usage. Issues of language contact and exposure, as well as learner identity are fundamental to the fourfold learning context distinction alluded to above, namely naturalistic and instructed learning, as well as study abroad and domestic immersion, with a key question arising as to how instructed and naturalistic learning may differ, or not. The question pertains to the long-standing issue of input provision and opportunities for language use during instructed learning, such as the different manipulations of language input in terms of the type, quantity, quality, frequency, duration, and intensity of exposure and use. Critical examples concern the role of metalinguistic knowledge more generally, and more specifically grammar instruction on an explicit–implicit continuum in the context of focus-on-form versus focus-on-meaning approaches [16,17]. Similarly, against the backdrop of a traditional drip–feed approach [18] characterised by classes allowing for regular but short exposure to the language, other manipulations of instructed learning conditions have showcased the potential of more intensive exposure over longer periods of time. Beyond input and interaction matters, issues of identity may also be fundamentally at play, reflecting the differential status of the learner as a member of different speech communities, be it at home or abroad [19].

## 3. Second Language Acquisition and Language Education

Against the preceding backdrop, as noted, the scope of the interface between SLA and second language education is extensive, extending to a significant range of stakeholders involved in instructed L2 learning. Indeed, instructed language learners constitute a significant cohort of L2 learners, along with their naturalistic counterparts, with SLA-related research studies clearly classifying learner participants in individual studies based on such a distinction. Other distinctions pertain to study-abroad learners, namely instructed learners spending time abroad, and domestic immersion learners whose instruction holds other distinguishing characteristics. Second language education has a long history, with the learning of an additional language(s) being the norm for many students. Over time, the evolving pedagogic approaches prescribed within different educational policies and contexts have reflected changes in the conceptualisation of learning, underpinned by a greater awareness and understanding of language learning processes and outcomes. Language testing and proficiency scales have correspondingly been enhanced over time with the aim of

better capturing learner developing proficiency as it pertains to different areas of language and competences underlying language use and comprehension (for examples of publicly available proficiency tests and scales, see the Common European Framework of Reference for Languages [CEFR [20], American Council on the Teaching of Foreign Languages [ACTFL [21] proficiency guidelines, and the ACTFL [22] Oral Proficiency Interview [OPI]).

Taken together, the two fields of SLA and L2 education have benefited from mutual dialogue to varying degrees, with some critical reflection arising in how relevant findings are used to inform one another. A case in point concerns language testing and proficiency scales, where there has been useful critique of different aspects of different scales and expected competences at different proficiency levels (see [23] for the case of the CEFR). Other examples concern our understanding of the role of different pedagogic treatments, as in the case of focus-on-form and focus-on-meaning approaches [14,16,17], along with task-based language teaching approaches [24] which, together, have been the focus of increasing reflection on the learning benefits that they promote. Taken together, such areas of mutual engagement point to the potential for a sharper reflection on learning expectations, activities, and outcomes for instructed learners.

This Special Issue is intended as an opportunity to continue that engagement agenda, as a means of showcasing some examples of the mutual insights to be had in contemporary research on L2 learning in an instructed context. The different articles presented exemplify the scope of reflection across areas that span input and interaction matters, linguistic skills, and learner factors. In the former case, a number of articles focus on issues in input provision and their relationship with the development of different linguistic skills. They highlight the importance of reflection on the role of different types and manipulations of input, as well as interactional activities. Particularly, at the level of linguistic skills, the articles further highlight the scope of skills involved in L2 usage, extending to areas that may traditionally have received less attention in classroom instruction, as in the case of sociolinguistic competence (for discussion, see [25]). Regarding learner factors, other contributing articles cast their lamp on factors that have previously been less explored, pointing to the scope of such factors at play in language learning, as in the case of learner investment and flow.

The articles present investigations that span different instructional contexts among learners of different languages within different age groups, reflecting the breadth of the instructed educational enterprise. Moreover, they reflect different methodologies, and the rich insights that such different approaches provide, as well as providing perspectives that extend beyond the learner, instructor, and classroom to curricular and programmatic matters. In the following, a brief synopsis of each article is presented.

## 4. Synopsis of the Contributing Articles

Maja Roch, Raffaele Dicataldo, and Maria Chiara Levorato present a study that investigates receptive vocabulary and listening comprehension among Italian–English sequential bilingual children attending an international English school in Italy. Reflecting the importance of listening comprehension in L2 learning, the study offers a quantitative investigation of such skills in learners' L1 and L2, finding that L1 skills are more advanced than in the L2 for both receptive vocabulary and listening comprehension. However, the results for listening are within the monolingual range, pointing to the level of attainment made by the children, although not in the case of vocabulary. A correlation was also found between listening and receptive vocabulary among the younger learners. The study highlights the role of vocabulary understanding in the development of listening.

Numerous articles offer empirical studies on input presentation and interaction issues in an instructed context. The article by Roger Gilabert provides a suitable backdrop, where the author specifically offers reflections on task-based language teaching (TBLT) as a significant area of instructed second language acquisition (ISLA) investigation. The article first relates various premises underlying TBLT to wider issues in SLA and ISLA, thereby offering a range of insights for language practitioners into the interface between

the three interrelated (sub)fields. The author proceeds by considering task-based needs analysis, offering a range of practical insights for the enhancement of task design that draw on theoretical and applied concepts and constructs. The article highlights the synergies that can be beneficially embraced in exploring the interface between the three (sub)fields.

Reflecting the critical importance of learner exposure to the target language, Daniela Avello and Carmen Muñoz offer a study of the impact of captioned video-viewing on the development of listening and reading. The investigation of primary school learners of English in Chile explores the potential for such input provision in an instructed context, where learners' exposure to the language is primarily restricted to a limited number of classes. The comparative study of learners at two levels includes similar control groups of students who completed a range of tests which illuminate their listening skills, reading efficacy, reading speed, and vocabulary knowledge. Longitudinal tracking in pre-, post-, and delayed post-tests points to the developmental gains of the experimental groups at both levels. These gains pertain to enhanced listening skills at each stage of the study, along with enhanced reading efficacy and reading speed. Learner L2 vocabulary knowledge and L1 reading efficacy were found to contribute to their listening scores, while L2 listening skills further contributed to reading efficacy. These findings showcase the potential for captioned video usage in the classroom as a means of providing an input-rich resource for receptive skill development.

Raquel Serrano continues the focus on input exposure in an instructed setting, presenting a study of the role of extensive reading. The study of Spanish/Catalan child learners of English compares the impact of such reading over the course of an academic year through two modes, namely extensive reading only and reading while listening, with a control group included. Quantitative analysis of the increased post-test and delayed post-test vocabulary scores outlines the significant impact of such reading activities, with little difference between the two experimental groups. The study, however, shows some differences over the course of the year, whereby the gains were less extensive in term two compared to term one. While there were no differences between the two experimental groups in term two, the more reduced gains may reflect some differences in the reading activities between both terms, pointing to the importance of considering the arrangements of such activities in the integration of extensive reading within an instructed programme. Notwithstanding, the findings highlight the importance of extensive reading in L2 vocabulary development.

Joan C. Mora and Ingrid Mora-Plaza present a study of pronunciation training using a computerised map task. Their article reviews the difficulties that pronunciation poses to L2 learners, accounting for the ways in which targeted training can facilitate development among instructed learners. Situated in a task-based pronunciation teaching (TBPT) framework, the quantitative study offers a comparison of an experimental group and a control group of Hispanophone university learners in relation to their perception and production of the /i:/-/I/ contrast in English, as in 'feet' vs. 'fit'. While the findings demonstrate the benefits of a TBPT approach, they also illuminate a potential role of task complexity, as well as some issues in learner ability to generalise pronunciation in novel contexts. Overall, the study highlights the importance of focusing on pronunciation skill development during instructed learning.

Cyrille Granget, Cecilia Gunnarsson, Inès Saddour, Clara Solier, Vera Serrau, and Charlotte Alazard continue the focus on pronunciation in a study of the relationship between learner pronunciation and orthography. Their study focuses on nasal vowels among Japanese beginner learners of French as a third language. The learners' realisation of such sounds and their orthographic representation in writing are investigated in a spoken and written task. The findings indicate strong orthographic representation in writing, with the learners demonstrating high accuracy levels even at the outset of the longitudinal study presented. In contrast, the learners' oral production is more fragile, especially in spontaneous speech compared to repeated speech, with various alternatives to the nasal vowel produced, as in the case of a vowel followed by /n/. While some development is evident between the initial stage of the study and the post-test results, the findings indicate

crosslinguistic influences at play, as learners draw on other languages in their multilingual repertoire. Thus, the conclusion highlights the importance of pronunciation training, and the authors offer various pedagogic recommendations as a means of enhancing learner phonetic representation that does not rely on written representation alone.

Aintzane Doiz and David Lasagabaster explore the area of teacher–student interaction in relation to teacher questions in the classroom. Their study is set in an English-medium higher education context, drawing on data recordings of a series of lectures by teachers in the Humanities as a subject discipline that has not received attention in the area. Using a taxonomy of different question types, quantitative analysis shows how some question types dominate, with considerable inter-teacher variability also reflecting different questioning styles. When confirmation checks dominated as the question type, the authors observe how they do not serve their intended pedagogic goal, but rather are often used as fillers. Overall, the sense of student engagement through questioning seems to be restricted, as reflected by limited student responses. The authors consider the findings in terms of the interactional potential of the questions which dominate the classroom discourse, calling for greater teacher training in the area, in tandem with customised training to reflect the individual questioning styles of teachers.

Marit Myhre Bredesen and Kari-Anne B. Naess also present an article that investigates teacher questions among kindergarten children during a digital picture book reading task. Their case study focuses on 'quiet' children, representing a learner cohort who have not previously been the subject of extensive investigation on such a topic. The study considers different question types, from closed to open, where the questions are considered from the perspective of the learner response level that they elicit in learners' L2 Norwegian. The findings are based on both quantitative and qualitative analyses, pointing to different frequencies and details of responses across and within question types, with open-ended questions in particular not found to elicit extensive responses compared to the other question types. The findings have implications for teacher engagement with 'quiet' learners within the age cohort considered, as well as beyond.

Nadia Mifka-Profozic builds on interactions in a classroom, studying learners' interactions in pairs and groups, and especially exploring the features of priming and alignment in contributing to the development of learner interactional competence. The author uses conversation analysis to consider how such features arise within interactional routines and their potential for language learning. The features relate to how learners may notice and use linguistic features previously arising in an interactional encounter. The study participants are university learners of English in Croatia who performed two tasks in pairs or small groups. The qualitative and quantitative findings indicate differences between the two group types. The study illuminates the potential of priming and alignment in the acquisition of interactional competence, with pedagogical implications for their integration in classroom interactional tasks presented.

Lidia Mañoso-Pacheco and Roberto Sánchez-Cabrero focus on bilingual education programmes in Spain as an example of content and language integrated learning (CLIL) education. In particular, they present quantitative findings relating to pre-service teachers' attitudes towards such an education model, reflecting a critical cohort of stakeholders as future teachers for such programmes. The study draws on a questionnaire which allows the authors to consider correlations between a range of variables and attitudinal perspectives that emerge. While the bilingual programme is generally valued, key findings relate to the impact of self-perceived English proficiency in shaping participants' attitudes, often conditioned by the participants' own background schooling, as well as concerns around learner development on subject content in such an education context. The findings have implications for the future development of such programmes in a Spanish context, with regard to the relationship between Spanish and English as a dominant global lingua franca of our times.

Miguel Hernández Hernández and Jesús Izquierdo analyse teacher attitudes in a study that investigates their relationship with the adoption of curricular developments. The

study is situated in a rural education environment in Mexico, complementing a previous predominant focus of such issues in an urban context. The study draws on both quantitative and qualitative analysis of questionnaire and interview data to illuminate the challenges faced by generalist teachers who are called on to offer instruction on English in tandem with other subjects. These challenges are situated against the backdrop of curricular changes relating to English in a Mexican context. The study reports a weak correlation between teacher attitudes and the adoption of such changes, showcasing the critical importance of teacher engagement and support to teachers in curricular reform and implementation.

Focusing on sociolinguistic competence, Katherine Rehner and Ivan Lasan offer a study in the context of French as a second language in Ontario, Canada. Using different data types, they explore learner retrospective reflections on the fulfilment of their sociolinguistic needs during language learning. These findings are complemented with teacher reflections on how they are addressing sociolinguistic development in their classrooms in the context of integrating the Common European Framework of Reference. The study highlights mismatches between learners' sociolinguistic needs in their use of the language and the underdeveloped sociolinguistic skills they believe to hold. Optimistically, however, the teacher reports demonstrate the changes they have made to their pedagogic approach. The article builds on the extensive body of research on the acquisition of sociolinguistic competence in a second language to highlight the pedagogic potential of sociolinguistic development.

Leonor Dauzón-Ledesma and Jesús Izquierdo explore the area of learner investment in their language learning. Following a review of the concept of investment, the authors present the development of a quantitative questionnaire for learners in a foreign language institutional learning context, as opposed to the target language community, reflecting that the dimensions of investment may be distinct in one learning environment compared to the other. The questionnaire includes different items on a Likert scale which tap into different dimensions of investment, namely motivation, necessity, engagement, and agency. The remainder of the article presents findings stemming from a study which used the questionnaire among university learners whose English language learning is mandatory within their institution in Mexico. While the learners demonstrated strong motivation and a utilitarian perception of English language learning, learner engagement and agency were more reduced. In providing an understanding of learner investment on mandatory language learning programmes, the findings carry implications for language education policy in relation to such programmes.

Finally, in their article, Jean-Marc Dewaele, Alfaf Albakistani, and Iman Kamal Ahmed explore the concept of 'flow' during student learning in online remote English-as-a-foreign-language classrooms as they arose during the pandemic circumstances. The authors review the concept of flow, referring to 'an optimal psychophysical state' that may arise during engagement in a learning situation. The authors draw on questionnaire data to offer a quantitative comparative analysis with in-person classroom learning, finding that flow is significantly enhanced under the latter conditions. Their study also explores the role of some learner internal and external variables, finding that a larger range of variables impact flow during in-person classroom learning compared to online learning. While offering some reassurance on the potential for flow to be supported in an online remote teaching environment, the study highlights the importance of in-person classroom dynamics in engaging learner flow.

**Acknowledgments:** I am grateful to the authors who have contributed their work to this Special Issue.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Bardel, C.; Hedman, C.; Rejman, K.; Zetterholm, E. *Exploring Language Education. Global and Local Perspectives*; Stockholm University Press: Stockholm, Sweden, 2022.
2. Cook, V. *Second Language Learning and Language Teaching*; Routledge: London, UK, 2016.

3. Spada, N. SLA research and L2 pedagogy: An uneasy relationship. In *Exploring Language Education: Global and Local Perspectives*; Bardel, C., Hedman, C., Rejman, K., Zetterholm, E., Eds.; Stockholm University Press: Stockholm, Sweden, 2022; pp. 13–37.
4. Trévisiol-Okamura, P.; Komur-Thilloy, G. *Discours, Acquisition et Didactique des Langues. Les Termes d'un Dialogue*; Orizons: Paris, France, 2011.
5. Watorek, M.; Rast, R.; Arslangul, A. *Premières Etapes dans l'Acquisition des Langues Etrangères. Dialogue entre Acquisition et Didactique des Langues*; Presses de l'INACLO: Paris, France, 2021.
6. Selinker, L. Interlanguage. *Int. Rev. Appl. Linguist. Lang. Teach.* **1972**, *10*, 209–231. [CrossRef]
7. Han, Z.; Tarone, E. *Interlanguage. Forty Years Later*; John Benjamins: Amsterdam, The Netherlands, 2014.
8. Housen, A.; Pierrard, M. *Investigations in Instructed Second Language Acquisition*; Mouton deGruyter: Berlin, Germany, 2005.
9. Lightbown, P.; Spada, N. *How Languages Are Learned*; Oxford University Press: Oxford, UK, 2021.
10. Howard, M. *Study Abroad, Second Language Acquisition and Interculturality*; Multilingual Matters: Bristol, UK, 2019.
11. Reinders, H.; Lai, C.; Sundqvist, P. *The Routledge Handbook of Language Learning and Teaching beyond the Classroom*; Routledge: London, UK, 2022.
12. Rebuschat, P. *Implicit and Explicit Learning of Languages*; John Benjamins: Amsterdam, The Netherlands, 2015.
13. Johnson, K.; Swain, M. *Immersion Education. International Perspectives*; Cambridge University Press: Cambridge, UK, 1997.
14. Lightbown, P. *Focus on Content-Based Language Teaching*; Oxford University Press: Oxford, UK, 2013.
15. Hulstijn, J. Different speakers but same Language Acquisition Device. *Linguist. Approaches Biling.* **2012**, *2*, 273–276. [CrossRef]
16. Doughty, C.; Williams, J. *Focus on Form in Classroom Instruction*; Cambridge University Press: Cambridge, UK, 1998.
17. Hulstijn, J. Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Stud. Second. Lang. Acquis.* **2005**, *27*, 129–140. [CrossRef]
18. Muñoz, C. *Intensive Exposure Experiences in Second Language Learning*; Multilingual Matters: Bristol, UK, 2012.
19. Pavlenko, A.; Blackledge, A. *Negotiation of Identities in Multilingual Contexts*; Multilingual Matters: Bristol, UK, 2004.
20. Council of Europe. *CEFR Companion Volume*; Council of Europe: Strasbourg, France, 2020.
21. American Council for the Teaching of Foreign Languages. ACTLF Proficiency Guidelines. 2012. Available online: https://www.actfl.org/educator-resources/actfl-proficiency-guidelines (accessed on 3 March 2024).
22. American Council for the Teaching of Foreign Languages. Oral Proficiency Interview. n.d. Available online: https://www.actfl.org/assessments/postsecondary-assessments/opi (accessed on 3 March 2024).
23. Hulstijn, J. The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *Mod. Lang. J.* **2007**, *91*, 663–667. [CrossRef]
24. Nunan, D. *Task-Based Language Teaching*; Cambridge University Press: Cambridge, UK, 2004.
25. Rehner, K.; Mougeon, R. The effect of educational input on the development of sociolinguistic competence by French immersion students: The case of expressions of consequence in spoken French. *J. Educ. Thought* **2003**, *37*, 259–281.

education sciences

MDPI

*Article*

# Receptive Vocabulary and Listening Narrative Comprehension of Italian–English Bilingual Children between 5 to 7 Years

**Maja Roch *, Raffaele Dicataldo and Maria Chiara Levorato**

Department of Developmental Psychology and Socialization, University of Padua, 35131 Padua, Italy; raffaele.dicataldo@unipd.it (R.D.); chiara.levorato@unipd.it (M.C.L.)
* Correspondence: maja.roch@unipd.it

**Abstract:** Vocabulary is the key component of listening narrative comprehension, but its contribution has been scarcely investigated in bilingual children. This study aimed to examine (a) listening narrative comprehension and receptive vocabulary in L1 (Italian) and L2 (English) in preschool and first grade children; (b) the specific contribution of receptive vocabulary to listening narrative comprehension in both languages. Participants were 30 preschoolers and 32 first graders, who are all Italian children attending an international school in English. In both languages, receptive vocabulary was assessed through PPVT-R and listening narrative comprehension through TOR 3-8. The results showed that listening narrative comprehension was age appropriate in both languages but higher in L1. Lower vocabulary in L1 than L2 was found, and this difference is higher for preschoolers than for first grades; finally, two regressions performed on listening narrative comprehension in each language showed that only vocabulary in the same language accounted for listening narrative comprehension. Children obtain higher performance in L1; however, after a few years of L2 exposure in the educational context, language skills fall within the normal range, with some weakness in vocabulary. Vocabulary contribution to listening narrative comprehension is similar in both languages and specific for each.

**Keywords:** receptive vocabulary; listening narrative comprehension; bilinguals; school readiness

**Citation:** Roch, M.; Dicataldo, R.; Levorato, M.C. Receptive Vocabulary and Listening Narrative Comprehension of Italian–English Bilingual Children between 5 to 7 Years. *Educ. Sci.* **2023**, *13*, 780. https://doi.org/10.3390/educsci13080780

Academic Editor: Martin Howard

Received: 4 May 2023
Revised: 25 July 2023
Accepted: 28 July 2023
Published: 1 August 2023



## 1. Introduction

Emergent literacy includes a set of interrelated linguistic skills, knowledge and attitudes identified as developmental precursors to conventional forms of reading and writing, pivotal for later school readiness and academic achievement [1]. Among these skills, listening comprehension—the ability to understand spoken language, crucial for successful communication—has a significant role in emergent literacy acquisition and later reading comprehension [2]. Listening narrative comprehension is a constructive and integrative process in which the interpretation of vocabulary, sentences and explicit and implicit information results in a coherent mental representation of the text [3]. In this study, we adopted the multicomponent model of text comprehension according to which narrative comprehension involves several languages and cognitive skills that interact dynamically and reciprocally allowing the obtainment of a coherent mental representation of the narrative meaning [4].

Although the importance of listening narrative comprehension from a preschool age has been recognized as the best predictor of subsequent reading comprehension [5], most studies adopting a multicomponent approach focus on reading comprehension in school-aged children [6,7]. However, recently this model has been partially tested for listening narrative comprehension in preschool children [8–13]. A growing body of research has investigated how monolingual children, from preschool to school age, use different language and cognitive skills in listening narrative comprehension [10–12,14–16]. However, research has not yet established how linguistic skills predict listening comprehension concurrently and longitudinally in children who speak more than one language [17]. Our study had three

distinct research questions. First, we explored performance in vocabulary and listening narrative comprehension in both languages of Italian preschool and school-aged children attending an international English school. Second, we wanted to investigate potential interrelations between L1 and L2 levels and the cross-linguistic transfer of linguistic skills. Third, we explored to what extent L1 and L2 vocabulary contribute to listening narrative comprehension in both languages in preschoolers and first graders speaking two languages.

Developmentally, the transition to primary school represents a particularly challenging period for the refinement of emergent literacy, as children enter a new environment where they must learn to understand oral and written texts, handle more cognitively complex tasks that require and build on good basic skills, both cognitive and linguistic. Preschoolers, through their daily experiences, encounter opportunities to develop oral language skills, gain knowledge about the forms and functions of written language and practice their emerging literacy skills. Once they move to school, children encounter opportunities to develop reading skills and practice; through specific exercises, their decoding skills have the potential to affect later reading comprehension [18]. Several components of listening narrative comprehension develop during preschool age and become more and more efficient once children enter primary school and advance through formal education [16]. Certainly, any weakness or developmental delay in core oral language skills may act as a bottleneck and constrain the ability to engage in higher-level comprehension processes [18]. Involvement in prereading activities may produce differences in the relation between comprehension and language skills, such as vocabulary, in the transition from preschool to school, particularly for children exposed to more than one language, who often lack the second language and preliteracy skills needed to best adapt to school demands.

### *1.1. The Role of Vocabulary in Listening Narrative Comprehension in Monolinguals and Bilinguals*

Understanding single words and their structural relationship within a sentence is the essential first step for understanding the meaning of a text. Vocabulary and morphosyntax have repeatedly been associated with language comprehension [8,12,14,17]. Recent research has highlighted the importance of vocabulary in text comprehension, showing that it represents the core ability and one of the best predictors for narrative comprehension from kindergarten to school in whatever modality a text is presented [8,9]. Previous studies, see ref. [19], showed that receptive word knowledge in preschoolers accounted for 4% of unique variance in reading comprehension when they were in third grade. Several longitudinal studies found that measures of receptive word knowledge directly predicted reading and listening comprehension, over and above the autoregressor and other components, in primary school children [20–24]. Thus, it may be argued that vocabulary appears as a core language ability for successful listening narrative comprehension in preschool and school-aged children [13]. Additionally, Meara argued that vocabulary size is the fundamental competence for acquiring lexical competence and emphasized that children with larger vocabularies are more proficient language users, understand more and produce more complex oral narratives than children with smaller vocabularies [25]. This evidence emphasizes the importance of vocabulary in listening narrative comprehension that may be limited in children who have a smaller vocabulary, as is often the case with children exposed to more than one language [26].

In the last few decades, society has become increasingly multilingual and worldwide the number of children developing in multilingual contexts has grown exponentially [27]. In parallel, there is growing research investigating how multilingualism shapes the linguistic developmental trajectory during preschool and how bilingual children face literacy and schooling [28]. The strongest effect of bilingual exposure on language development concerns vocabulary growth; thus, the investigation of vocabulary development in bilingual children received great attention in the previous literature. Bilingual children typically have lower scores than monolinguals on measures of both receptive [29–31] and expressive vocabulary [32–34] in at least one, but frequently in all, of the spoken languages. Additionally, they show a slower rate of vocabulary development in both languages compared to

monolingual peers [35]. The vocabulary of bilingual children is smaller compared to that of monolingual peers both in preschool [36,37] and school-aged children [29], even after three consecutive years of exposure [38]. This evidence causes concerns for the academic outcomes of bilingual children. In fact, poor vocabulary knowledge limits the ability to understand oral and written narratives, which is required to have full access to the curriculum, and consequently hinders the progress of spoken and written language [39]. On the one hand, the knowledge of words is crucial for understanding the meaning of the whole narrative. On the other hand, exposure to texts represents the main source for the acquisition of new words [40,41]. The relationship between these skills is reciprocal: the better children understand the narrative, the greater the opportunity to learn new vocabulary; on the other hand, increased vocabulary knowledge results in a greater chance that the narrative is understood [23]. However, research has not yet established how vocabulary predicts listening comprehension concurrently and longitudinally in children who speak more languages.

In recent years, narratives have been used for assessing bilingual language development during preschool and for establishing the relationship between bilingual exposure and language development [42–44]. Measures of narrative competence, which include both narrative production (storytelling and/or retelling) and comprehension assessment, allow the examination of a wide range of linguistic abilities as well as cognitive and pragmatic skills and provide rich data on children's multiple linguistic abilities, including story structure, structural complexity, internal state language, cohesion, morpho-syntax, lexical diversity and productivity [45]. For narrative production, a general result that emerges is that bilingual children in preschool years show similar narrative competence as far as macrostructure is concerned, whereas they tend to struggle with microstructure of narratives [46], attributable to lower exposure to each of their languages [47]. Numerous studies have investigated the narrative production of bilingual children showing that the ability to produce narratives in both L1 and L2 develops and improves from preschool to school age [48–50].

With regard to listening narrative comprehension, Rodina [51] found that bilinguals' ability to understand a story is equally developed in both languages. Bohnacker [45], investigating narrative comprehension in bilinguals aged 5–7, found that for the children's comprehension of macrostructural elements, the 5-year-olds scored lower but still relatively close to the 6- to 7-year-olds, with large variation within the group. For neither age group was there any difference between the two languages. The results show a large gap between story production and story comprehension for both age groups: comprehension was clearly ahead of production. In summary, the literature shows that even bilinguals with limited linguistic competence are able to comprehend narratives as far as macrostructure is concerned, albeit showing poor vocabulary and morphosyntactic comprehension [26,48,49,52]. Following on from these findings, it seems that the global structure of narratives develops as a function of age and not of language [53], while the ability to use adequate vocabulary for comprehending narratives develops as a function of increased exposure and linguistic input [26]. The question of how much vocabulary is necessary to adequately understand an oral narrative for bilingual preschool and school-aged children remains open.

To date, research with children speaking more than one language focusing on the contribution of vocabulary to listening narrative comprehension is scant and provides mixed results [17]. Stæhr found a strong relationship between vocabulary knowledge and listening narrative comprehension with advanced second language learners of English [54]. The results showed that vocabulary knowledge was highly correlated with listening comprehension and predicted half of the variance in the listening scores. Roch and Hrzica [31] investigated listening narrative comprehension in Croatian–Italian bilinguals aged 5–7, in both languages, aiming to find out to what extent receptive vocabulary and sentence comprehension predict narrative comprehension skills and possible interdependence between languages. The results show better performance in a narrative comprehension task in L1

than L2 including higher accuracy at answering questions about characters' goals than questions about mental states; they also found a similar contribution of vocabulary and sentence comprehension on narrative comprehension in L1; on the other hand, a smaller contribution of vocabulary and sentence comprehension on narrative comprehension in L2 emerged. These results have been confirmed by a recent study by Valentini and Serratrice [17] according to which vocabulary and morphological knowledge were the most significant predictors of English listening comprehension (L2) in bilingual children in the first two years of formal schooling. These skills specifically determined the children's listening comprehension but not their growth in listening narrative comprehension abilities over time, as is also the case for monolinguals [13].

### *1.2. Crosslinguistic Correlations*

Another important issue that concerns linguistic development in children speaking more than one language pertains to the interrelations between L1 and L2 levels and crosslinguistic transfer. Crosslinguistic transfer has been observed in simultaneous bilingual children in phonology [55], vocabulary [32] and syntax [56,57]. A recent meta-analysis of crosslinguistic transfer of oral language shows a small meta-correlation between L1 and L2 oral language skills and a moderate to large correlation between L1 and L2 phonological awareness and decoding [58].

As for narrative competence, previous research has shown moderate crosslinguistic associations regarding macrostructure. Several studies found a relation between measures of macrostructure in L1 and L2 in the first grade but not in kindergarten [49,50,52]. Explanations for the results pointed out that there could be a shared conceptual knowledge of macrostructure of stories in L1 and L2, and this might facilitate crosslinguistic transfer while more experience (e.g., through schooling) is acquired in both languages [52]. The linguistic interdependence hypothesis [59], according to which every language presents its own superficial manifestations while underlying cognitive processes are common across languages, has been applied as a theoretical framework in studying narrative competence and in the analysis of the relationship between the macrostructure and the microstructure. For bilingual speakers, researchers hypothesize that the macrostructure is invariant across the two languages due to its dependency on cognitive skills. On the other hand, narrative microstructure, being more language specific, is less likely to transfer from one language to another and thus suffers from the effect of exposure to a specific language.

Crosslinguistic correlations of listening narrative comprehension in bilinguals need more attention since all the results of crosslinguistic transfer that have been reported in preschool children concern narrative production rather than comprehension [60] and in simultaneous rather than sequential bilingual preschoolers. Recently, Roch & Hrzica [31] analyzed the possible crosslinguistic transfer in narrative comprehension of sequential bilingual speakers of Croatian and Italian. The results suggest a degree of interdependence between L1 and L2: each language comprehension measure (vocabulary, sentence and narrative) in one language correlated with the same measure in the other language; each comprehension level in one language, however, correlated weakly with the other levels in the other language. These results, in line with previous works [50,61], highlight the complexity of the relations between L1 and L2 in bilinguals. Understanding crosslinguistic influence might help us both theoretically and practically, providing information on how the development of narrative comprehension in children learning two (or more) languages differs from that of children learning only one and then aiding the design of successful educational interventions that might help bilingual children, especially in preschool children, as a way to promote school readiness.

### *1.3. The Current Study*

Findings reported above highlight the need for the further advancement of our understanding of the relationship between L1 and L2 vocabulary and listening narrative

comprehension of children speaking more than one language, particularly in the transition between preschool and primary school.

The main aim of the current work concerns whether the generally reported weak vocabulary of bilingual preschoolers and first graders may constrain broad, higher-level language processing, namely listening narrative comprehension. The rationale is that any weakness in core oral language skills (i.e., vocabulary) may act as a bottleneck and constrain the ability to engage in higher-level comprehension processes, impeding a successful listening narrative comprehension. The following research questions guided the current study and, although there is a lack of consensus about the relationship between receptive vocabulary and narrative comprehension in bilingual speakers, we advance some predictions:

(1) To what extent do children show a different performance in L1 and L2 receptive vocabulary and listening narrative comprehension and to what extent does their performance change between 5 and 7 years? In line with the previous literature [38], although they were consistently exposed to L2 at school for at least three years, it is predicted that: (a) they will show some disadvantage of L2 over L1 in listening narrative comprehension; (b) there is a greater L2–L1 gap in receptive vocabulary between 5 and 7 years than in listening narrative comprehension.
(2) To what extent are the two linguistic systems related? We expect to find high correlations between vocabulary and listening narrative comprehension in each of the two languages, as predicted from the previous literature [10] and weak correlations between the two languages, as shown in previous works using different tasks and involving different language combinations [31,61]. Based on previous studies [52], it is possible that the pattern of these relationships changes with age.
(3) To what extent does L1 and L2 vocabulary contribute to listening narrative comprehension in both languages? Previous studies reported mixed results concerning this point and therefore we did not put forward specific predictions.

The findings of this study will provide evidence in both L1 and L2 listening narrative comprehension in relation to vocabulary comprehension in preschoolers and first graders and of the contribution of receptive vocabulary in their listening narrative comprehension. Because of the distributed nature of exposure to their languages, bilingual children also offer a unique opportunity for investigating the role of a relative amount of input in the process of listening narrative comprehension in the corresponding language [17]. Usually, the heterogeneity of this population, regarding the degree and quality of exposure to more languages in the home context, limits the generalizability of the findings [62]. In this study, we tried to control these variables (the amount and quality of linguistic input) by involving children born and raised in Italy, by Italian parents, but enrolled in an international school in English. In this way, the quantity of input and its quality are to be considered the same for all study participants, providing information that allows for the greater generalizability of the results. In addition, to analyze any variations between preschoolers and first graders, and the effect of the amount of language exposure, we involved two groups of different ages that are characterized by different amounts of language input in both languages and particularly in L2. Finally, the use of a standardized test to assess the comprehension of an oral narrative that does not involve expressive skills (described below) could provide insights into the contribution of receptive vocabulary to the listening narrative comprehension in both languages, Italian and English.

## 2. Materials and Methods

### *2.1. Participants*

Study participants were 62 Italian children attending an English international school in northeastern Italy. Thirty children attended the last year of preschool and were thus not yet conventional readers (Mage 5;5, SD = 3 months, range 5–6 years) and 32 children attended the first year of primary school (Mage 6;6, SD = 4 months, range 6–7; 2 years) and were mostly exposed to prereading exercises. The children's mean age of first exposure to

English was 3 years, 3 months (SD = 1 month) for the younger group and 3 years, 6 months for the older group (SD = 2 months). The children's parents completed a short questionnaire investigating the amount of linguistic input in L1 and L2. In Supplementary Materials, we report the short questionnaire developed during the COST Action IS1804. To avoid confounding effects due to socioeconomic background, we only selected children from middle to high SES families. Additionally, to avoid confounding effects due to the quantity and quality of linguistic input in English, we selected only children enrolled full-time. Both groups, preschoolers and first graders, were exposed to English daily, in different activities appropriate for their age, for approximately 8 h every day. Moreover, unlike previous studies, all the children involved in this study had the same L1 (Italian) allowing us also to control for possible effects related to the language of origin. The study was approved by the ethics committee of the University of Padua (protocol n. 1521) and performed in accordance with the principles expressed in the Declaration of Helsinki. Only children with signed parental consent participated in the study.

### *2.2. Materials and Procedure*

Two trained psychology graduate students tested each child individually in a quiet room during the school day with standardized tests (described below). Tasks were administered in a fixed order, preferable for investigating individual differences [63]. Each child was tested in two sessions lasting approximately 30 min each, on two different days (one for language), at the end of which the testers thanked the child for their participation and rewarded him/her with free playtime.

Receptive vocabulary: The children's English vocabulary was assessed through the Peabody Picture Vocabulary Test—PPVT-R [64], whereas the adapted and standardized version for Italian was used to assess their Italian vocabulary [65]. Adapted versions of PPVT keep the same procedure as the original version but introduce changes in the lexical material (order of words, exclusion/inclusion of words) to obtain a similar level of difficulty. It consists of a list of words presented to participants who are asked to point out which, out of four pictures, best represents the target word. The items are presented in order of increasing difficulty. Testing is then continued until the participant obtains six incorrect answers in eight consecutive items. Raw scores correspond to the number of correct answers (range 0–175); age-specific standard scores (with a mean of 100 and standard deviation of 15) are provided in the PPVT-R manual. The reliability for the PPVT-R, evaluated with split-half procedure, is 0.88.

Narrative comprehension: The test TOR 3-8 is a standardized test for Italian children aged between 3 and 8 years of age that measures listening narrative comprehension without involving expressive skills [66]. Assessing children's comprehension through listening narrative comprehension tasks such as these allows for the minimization of the constraints of oral language skills involved in narrative retelling tasks, as well as difficulties in answering verbal comprehension questions. To assess narrative comprehension in English, all the material of this standardized test was translated into English using the back translation method. The test consists of two short stories of equal difficulty and length. One story was presented in each of the two languages. The story is read to the child and his/her comprehension is evaluated by asking 10 questions, followed by a multiple-choice task with four possible answers, which were represented by pictures. The tester pauses the reading at two pre-established points and asks the questions in order to avoid overloading memory resources and to guarantee the child maintains attention. All the questions concerned information that is necessary for an adequate understanding of the story. Half of the questions are based on explicit information while the others concerned information that could be inferred from the text through the generation of text-based or knowledge-based inferences. The score consists of the sum of correct answers, 10 for each story, with a maximum score of 20. Raw scores can be converted into scaled scores (M = 10, SD = 2). Cronbach's alpha over items ranges from 0.52 to 0.72.

## 3. Results

In order to answer the first research question, the average performance of participants in the two tasks used in both languages was observed. Table 1 reports mean scores and standard deviations, in brackets, obtained in the two linguistic tasks, namely receptive vocabulary (PPVT) and listening narrative comprehension (TOR) for both L1 (Italian) and L2 (English) as a function of the age group (preschoolers and first graders). Both raw (first row) and standardized (second row) scores are reported in Table 1.

**Table 1.** Descriptive statistics.

| | | L1 (Italian) | | L2 (English) | |
|---|---|---|---|---|---|
| | | **PPVT** | **TOR** | **PPVT** | **TOR** |
| 5 years ($n$ = 30) | Raw score | 79 (14) | 6.5 (1.7) | 56 (9) | 5.2 (1.4) |
| | Standard score | 91.4 (11.6) | 10.5 (1.5) | 83.1 (9.5) | 9.6 (1.1) |
| 6 years ($n$ = 32) | Raw score | 90 (20) | 5.7 (2.3) | 64 (9) | 4.9 (1.7) |
| | Standard score | 94.1 (14.9) | 9.7 (2.1) | 87.9 (11.3) | 9.1 (1.7) |

PPVT standard score: mean 100; SD 10; TOR standard score: mean 10; SD 2.

Descriptive statistics show that L2 constituted a weaker language. However, while for listening narrative comprehension children show age-appropriate performance in both languages, in receptive vocabulary, children show a delay in L2, with -1SD performance compared to monolingual scores in the normative sample.

### *3.1. Levels of Narrative Comprehension and Receptive Vocabulary: The Role of Age and Language*

To analyze whether the advantage of L1 over L2 decreases between 5 and 7 years, we performed a mixed ANOVA 2 Ages (preschoolers and first graders) × 2 Languages (L1 and L2) on each of the two dependent variables: Receptive Vocabulary and Listening Narrative Comprehension. Age was a between subjects' factor and Language was a within subjects' factor. The assumptions of normality and of homogeneity of variance were verified. According to the Shapiro–Wilk test, two of our observed variables, namely English vocabulary and Italian text comprehension are normally distributed (0.987 and 0.966, respectively), while the other two variables, namely Italian vocabulary and English text comprehension are not normally distributed ($p < 0.05$). Additionally, we verified the homogeneity of the variance through the Levene's test and we found that only the Italian vocabulary reported a difference in the variances among the two groups ($p < 0.05$); in fact, most of the children have low vocabulary.

In the case of the receptive vocabulary, both main factors yielded significance: the effect of Age was significant [F (1,60) = 11,40, $p < 0.001$, $\eta^2 = 0.116$] indicating higher receptive vocabulary for older children; the effect of Language was significant [F (1,60) = 145.79, $p < 0.001$, $\eta^2 = 0.708$] indicating a richer receptive vocabulary in L1 (Italian) than in L2 (English). The interaction Age × Language was not significant. Different results emerged from the analysis with listening narrative comprehension as the dependent variable: only the Language factor yielded significance [F (1,60) = 11.61, $p < 0.001$, $\eta^2 = 0.162$], while Age and the Interaction between the two factors, in both cases, were not significant.

### *3.2. Relationship between Narrative Comprehension and Receptive Vocabulary*

To investigate to what extent the two linguistic systems are related, we performed correlational analyses between the two measures within each language and across the two languages. Table 2 shows the results of preschoolers and of the first graders.

Correlation analyses suggest a different pattern of relations for preschoolers and first graders. Regarding preschoolers, the two language domains correlate only in the weaker language (L2), whereas for first graders, receptive vocabulary and listening narrative comprehension correlate in both languages. As far as the crosslinguistic relationships are concerned, it emerged that the preschoolers' receptive vocabulary correlated significantly

between the two languages whereas listening narrative comprehension in the two languages were not correlated. On the other hand, no significant crosslinguistic correlations emerged between the vocabulary in L1 and L2 for first graders, whereas a significant correlation emerged between the listening narrative comprehension in L1 and the receptive vocabulary in L2. To summarize, receptive vocabulary and listening narrative comprehension correlate in each of the two languages for first graders and only in L2 for preschoolers.

**Table 2.** Relationships between the two tasks and the two languages for preschoolers and first graders.

| | | **Preschoolers** | | | |
|---|---|---|---|---|---|
| | | **L1 (Italian)** | | **L2 (English)** | |
| | | **PPVT-R** | **TOR 3-8** | **PPVT-R** | **TOR 3-8** |
| L1 (Italian) | PPVT-R | - | 0.142 | 0.624 ** | 0.278 |
| | TOR 3-8 | | - | 0.207 | 0.326 |
| L2 (English) | PPVT-R | | | - | 0.403 * |
| | TOR 3-8 | | | | - |
| | | **First graders** | | | |
| | | L1 (Italian) | | L2 (English) | |
| | | PPVT-R | TOR 3-8 | PPVT-R | TOR 3-8 |
| L1 (Italian) | PPVT-R | - | 0.534 ** | 0.273 | 0.042 |
| | TOR 3-8 | | - | 0.426 * | 0.092 |
| L2 (English) | PPVT-R | | | - | 0.434 * |
| | TOR 3-8 | | | | - |

* $p < 0.05$; ** $p < 0.001$.

### *3.3. The Contribution of Receptive Vocabulary to Listening Narrative Comprehension*

Finally, to analyze the contribution of receptive vocabulary in accounting for individual differences in listening narrative comprehension, we performed two multivariate linear regressions, one on listening narrative comprehension in L1 and the other on narrative comprehension in L2. In both regressions, we used the same potential predictors: in the first step the score obtained in the listening narrative comprehension task in the other language was inserted; in the second step, we inserted age in months to control for developmental changes; and finally, in the third step, receptive vocabulary scores, in both languages, were included. Table 3 reports the results of the regression on the L1 listening narrative comprehension.

**Table 3.** Summary of multivariate linear regressions analysis for variables predicting listening narrative comprehension in L1 ($n = 62$): $R^2 = 0.283$ [$F(4,61) = 5.6$, $p < 0.001$].

| | | **$R^2$ Change** | | **B** | **SE B** | **B** |
|---|---|---|---|---|---|---|
| Step 1 | Narrative comprehension L2 | 0.036 ° | | | | |
| | | | Narrative comprehension L2 | 0.251 | 0.178 | 0.189 |
| Step 2 | Age | 0.059 °° | | | | |
| | | | Narrative comprehension L2 | 0.224 | 0.165 | 0.169 |
| | | | Age | 0.074 | 0.038 | 0.245 |
| Step 3 | Receptive vocabulary L1<br>Receptive vocabulary L2 | 0.188 °°° | | | | |
| | | | Narrative comprehension L2 | 0.108 | 0.164 | 0.082 |
| | | | Age | 0.133 | 0.038 | 0.439 * |
| | | | Receptive vocabulary L1 | 0.045 | 0.015 | 0.387 ** |
| | | | Receptive vocabulary L2 | 0.032 | 0.031 | 0.153 |

° F change (1, 60) = 2.21, $p = 0.142$; °° F change (1, 59) = 3.89, $p = 0.053$; °°° F change (2, 57) = 7.47, $p < 0.01$; * $p < 0.05$; ** $p < 0.01$.

The multivariate linear regression model predicting listening narrative comprehension in L1 (Italian) explained 28% of variance. Listening narrative comprehension in L2, inserted in the first step, explained around 4% of variance, which did not yield significance. Age added a small and marginally significant portion of variance, namely 6%. The third step accounted for 18.8% of unique variance in L1 listening narrative comprehension. A closer inspection of the results reveals that only receptive vocabulary in L1 was significantly related to listening narrative comprehension in the same language ($\beta = 0.387$, $p < 0.01$) but not the receptive vocabulary in the other language ($\beta = 0.153$, $p = 0.281$). Table 4 reports results of the regression performed on listening narrative comprehension in L2 (English).

**Table 4.** Summary of multivariate linear regressions analysis for variables predicting listening narrative comprehension in L2 ($n = 62$): $R^2 = 0.186$ [F (4,61) = 3.26, $p < 0.01$].

| | | $R^2$ Change | | B | SE B | B |
|---|---|---|---|---|---|---|
| Step 1 | Narrative comprehension L1 | 0.036 ° | | | | |
| | | | Narrative comprehension L2 | 0.142 | 0.096 | 0.189 |
| Step 2 | Age | 0.059 °° | | | | |
| | | | Narrative comprehension L2 | 0.135 | 0.100 | 0.180 |
| | | | Age | 0.008 | 0.030 | 0.035 |
| Step 3 | Receptive vocabulary L1<br>Receptive vocabulary L2 | 0.188 °°° | | | | |
| | | | Narrative comprehension L2 | 0.070 | 0.106 | 0.093 |
| | | | Age | 0.038 | 0.033 | 0.166 |
| | | | Receptive vocabulary L1 | −0.013 | 0.013 | −0.151 |
| | | | Receptive vocabulary L2 | 0.074 | 0.023 | 0.462 * |

° F change (1, 60) = 2.21, $p = 0.142$; °° F change (1, 59) = 0.07, $p = 0.794$; °°° F change (2, 57) = 5.22, $p < 0.01$; * $p < 0.01$.

The multivariate linear regression model predicting listening narrative comprehension in L2 (English) explained 18.6% of total variance. Listening narrative comprehension in L1, inserted in the first step, explained 3.6% of variance, which did not yield significance. Age inserted in the second step did not add further variance. The third step accounted for a 14.9% of significant unique variance in narrative comprehension. A closer inspection of the results reveals that only the receptive vocabulary in L2 was significantly related to listening narrative comprehension in the same language ($\beta = 0.462$, $p < 0.01$).

## 4. Discussion

The main aim of our study was to investigate the role and the contribution of receptive vocabulary in listening narrative comprehension of bilingual children aged between 5 to 7 years, during the transition from preschool to school. The findings extend previous research, which has scarcely investigated the effects of multilingual exposure listening narrative comprehension and vocabulary skills among preschool and school-aged children raised in Italian-speaking families in Italy. The investigation of these effects was enriched by considering age and controlling for SES and L1 differences. In contrast to many studies on bilingual migrants or heritage language speakers, the participants in our study are Italian children born and raised in Italian families but attending an English international school and thus exposed to English daily for approximately 8 h.

Three main issues were addressed, and the following results were obtained.

Regarding the first research question, we found that even after different years of continuative exposure to two languages, L2 remains a weaker language. Children performed better in L1 than in L2 both in vocabulary and in listening narrative comprehension. However, a comparison of the participants' performance to the monolingual norms revealed that listening narrative comprehension falls within the age-appropriate performance in both languages, whereas vocabulary lags behind the typical performance showing a 1 SD delay, again in both languages spoken. Vocabulary growth was evident within the age range considered whereas listening narrative comprehension performance remained stable.

We failed to find any interaction between the effect of language and age, indicating that the pattern of results is similar for both languages.

Regarding the second research question, we investigated the relationship between receptive vocabulary and listening narrative comprehension both within each language and across the two languages (Italian vs. English). For school-aged children, we found that the receptive vocabulary and listening narrative comprehension correlate within each language, whereas for preschoolers only in L2. Moreover, significant correlations emerged between the receptive vocabularies in both the languages in preschool children and between vocabulary in L1 and listening narrative comprehension in L2 in first graders.

Finally, concerning the role of receptive vocabulary in listening narrative comprehension (the third research question), the two regression analyses indicated that receptive vocabulary accounted for a relevant amount of the total variance in listening narrative comprehension, namely 18% in the native language (L1) and 14% in the second language (L2). Neither in L1 nor in L2 listening narrative comprehension did vocabulary in the other language (namely, L2 skills for L1 narrative comprehension and vice versa) provide a significant contribution to the model.

The results of the current study are discussed for their theoretical relevance as well as for the practical implications for education within two different sections: on the one hand, concerning the level of skills reached in each linguistic domain and, on the other hand, the role of vocabulary in listening narrative comprehension in each language and across the two languages.

### *4.1. Receptive Vocabulary and Listening Narrative Comprehension in Bilingual Children*

The greatest advantage of measuring the language comprehension of bilinguals in both languages is that this allows for the comparison of the level attained in a native and second language. The current findings appear to be generally in line with what is reported in the literature on sequential bilinguals [31,47]: even after several years of exposure, vocabulary lags behind monolingual performance (i.e., −1 SD) in each language, and the vocabulary in L2 is significantly weaker than in L1. These results are in line with the results of Vettori and colleagues [67] who, working with bilingual language-minority children, found a statistically significant difference in lexical competence compared with monolingual children.

This result is not surprising for sequential bilinguals given that for the first years of life they were exposed only to one language, and from the introduction of the L2, they have been exposed less than monolinguals to both languages. In other words, the L1–L2 advantage in linguistic skills in our sample reflects their status as sequential bilinguals. In parallel, better outcomes for listening narrative comprehension for L1 than for L2 were also found, whereas we failed to find an age effect. The absence of an age effect for bilingual children could suggest that the amount of input, rather than age, is a better predictor in bilingual children because the amount of language experience is not just a function of age in this group [17]. Language input is one of the strongest predictors of the rate of language development in monolingual and bilingual children [68]. For our sample, although children belong to two different school grades (preschool vs. primary school), the amount of cumulative input in L2 is very similar (3.3 years vs. 3.6), while the daily amount of input in L2 is very similar (8 h). Additionally, qualitatively, no differences can be hypothesized among the children regarding input in L2 since they all attended the same school.

This result, although in line with the results of Roch and Hrzica [31], in part contradicts previous studies that analyzed narrative comprehension in bilingual speakers. Usually, no significant differences are found between children's L1 and L2 narrative comprehension [45,69]. However, concrete comparisons between these studies are difficult because the researchers used different methodologies, narrative comprehension measures and stimuli. For instance, narrative comprehension is usually measured through comprehension questions that are asked after a picture story has been told or retold. To the

best of our knowledge, this is the first study that examined narrative comprehension of a listened story (without pictorial stimuli) followed by comprehension questions among such young bilinguals in both languages. In addition, we used a standardized test (e.g., TOR 3-8), which allowed us to measure the extent to which listening narrative comprehension develops compared with monolinguals. This measure is very similar to that adopted for older children when measuring their reading comprehension, with the difference that in the current study we measured listening narrative comprehension. These preliminary data need to be further confirmed with future studies involving children of different age groups to understand developmental trajectories of listening narrative comprehension in both languages by bilingual speakers—simultaneous, sequential and second language learners—and how this ability can promote good outcomes in emergent literacy and transfer to reading comprehension.

### *4.2. The Relationship between Vocabulary and Narrative Comprehension in Each Language and between the Two Languages*

Vocabulary represents a relevant predictor for listening narrative comprehension for bilinguals, which is similar for monolingual children [17]. More interestingly, the current findings suggest that low vocabulary scores obtained in both languages did not prevent children of our sample from comprehending adequately an oral narrative text in each language. Receptive vocabulary emerged as an equally important predictor of narrative comprehension in both L1 and L2 and explained the 18% and 14%, respectively, of significant variation in listening narrative comprehension after narrative comprehension and age have been controlled for. In both languages, a monolingual-like pattern of relations was found [9,67]. In line with Roch and Hrzica [31], we analyzed and quantified the contribution of vocabulary in listening narrative comprehension in children's narrative comprehension, and we argue that the contribution is relevantly high. Although the materials and stimuli are different in these two studies, we found similar results about the contribution of vocabulary in children's comprehension. At the same time, it is also evident that there is a conspicuous variation in narrative comprehension that cannot be attributed solely to vocabulary. This puts forward a hypothesis that other contributing factors may clarify how other skills, presumably cognitive in nature, may promote narrative comprehension processes in children acquiring more than one language [19]. Multicomponent approaches of text comprehension emphasize that the construction of a coherent mental representation of the narrative is based not only on linguistic components but also on higher-level integrative processes, such as inferential abilities, knowledge of story structure and comprehension monitoring [4]. These higher-level cognitive components might be even more important for narrative comprehension in bilingual speakers—who cannot rely completely on (poor) linguistic skills—and their role should be investigated in future studies.

This sheds light on the fact that there is a need for further studies that investigate broader linguistic comprehension in bilingual speakers in early stages of development and before they start formal education. This might facilitate the early identification of possible risk factors for reading comprehension failure and might prevent future learning difficulties.

## 5. Limitations

A limitation of our research is that, contrary to some models in previous research [16], we did not consider the contribution and the possible mediation effect of lower and higher cognitive abilities on the relationship between vocabulary skills and listening narrative comprehension. It is possible that lower-level cognitive skills, such as memory and attention, and higher-level cognitive skills, such as inferential ability and comprehension monitoring, might have a subtler relationship with listening comprehension, mediated via a relationship between these skills and vocabulary skills. It could be, for instance, that children with better working memory might be better word learners, with better vocabulary skills, and that better vocabulary skills will positively affect their listening

narrative comprehension [17]. We believe our model is of value in highlighting the specific importance of receptive vocabulary in both languages in predicting listening narrative comprehension in L1 and L2; however, we cannot rule out the possible (mediated) effects of other linguistic and cognitive skills involved in the comprehension process.

Another limitation of our research is the lack of longitudinal measurement of the predictors of listening narrative comprehension. We acknowledge that including longitudinal measures of the predictors might have accounted for more variability in listening narrative comprehension, especially in relation to longitudinal changes in English input and its growth over time. Another limitation is related to the sample size since our sample is too small and thus has low statistical power to detect the relationships among the variables in our study limiting the generalizability of our results. Finally, another limitation of our research is related to the use of Italian norms for the English translated version of TOR 3-8, the task used to assess children's listening narrative comprehension. The lack of normative data for English does not allow us to properly assess performance in the English-language task such that future studies should use standardized tests in both languages.

## 6. Conclusions and Implications for Education

The results of this study, albeit considering its many limitations, represent a novel contribution to a better understanding of the determinants of listening narrative comprehension in bilingual children and have relevant pedagogical implications. The results are relevant since worldwide the number of children exposed to more than one language is increasing exponentially. To the best of our knowledge, ours is one of the few studies that analyzes the contribution of vocabulary skills in listening narrative comprehension simultaneously in both languages. The majority of the studies focus on narrative production and analyze the contribution of vocabulary in the same language [67]. Additionally, in our study, we controlled for the amount and quality of linguistic input by involving Italian children enrolled in an international school in English. Since language input is one of the strongest predictors of the rate of language development in monolingual and bilingual children [68], controlling for this variable allows for greater generalizability of the results.

The specific focus on listening narrative and vocabulary skills was derived from a number of different studies that recognized the implications of these skills for emergent literacy skills development and later reading and writing skills. Narratives are a text type in which bilinguals may become competent and in which they are able to overcome their vocabulary limitations if they are given adequate qualitative and quantitative input in both languages. Bilingual children benefit from the fact that narratives are a universal text on which they have developed knowledge in L1 and that they transfer to L2. Understanding the specific contribution of vocabulary in listening narrative comprehension is crucial for educational settings and schools to design specific pedagogical actions and interventions to ensure high-quality teaching and strategies to foster children's language development.

For teachers and school practitioners, it is important to know that time and quality restrictions may negatively influence children's performances and that for bilinguals it seems more useful to learn vocabulary through activities and tasks in which they feel competent in order to support their motivation and enjoyment through learning. Increasing children's high-quality lexical representations, particularly by providing them with more information about the meaning and use of words, is likely to have a positive cascading effect on their understanding of spoken language. Interventions aimed at improving school learning skills with bilinguals through oral narratives could have the secondary benefit of also improving positive self-image, relationships and wellbeing [70].

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13080780/s1. Background questions. This questionnaire was developed during the COST Action IS1804 Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment.

**Author Contributions:** Conceptualization, M.R. and M.C.L.; methodology, M.R.; formal analysis, R.D.; investigation, M.R.; resources, M.C.L.; data curation, M.R.; writing—original draft preparation, M.R. and R.D.; writing—review and editing, M.R. and R.D.; visualization, M.R.; supervision, M.R.; project administration, M.C.L.; funding acquisition, M.C.L. 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 Ethics Committee of the University of Padua (protocol code 1521, March 2021).

**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.

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

## References

1. Wise, J.C.; Sevcik, R.A.; Morris, R.D.; Lovett, M.V.; Wolf, M. The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *JSLHR* **2007**, *50*, 1093–1109. [CrossRef]
2. Hoover, W.A.; Gough, P.B. The simple view of reading. *Read. Writ.* **1990**, *2*, 127–160. [CrossRef]
3. Kintsch, W. *Comprehension: A Paradigm for Cognition*; Cambridge University Press: Cambridge, UK, 1998.
4. Oakhill, J.; Cain, K. (Eds.) Introduction to comprehension development. In *Children's Comprehension Problems in Oral and Written Language: A Cognitive Perspective*; The Guilford Press: New York, NY, USA, 2007; pp. 3–40.
5. Kim, Y.S. Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child. Dev.* **2015**, *86*, 128–144. [CrossRef]
6. Kendeou, P.; van den Broek, P.; White, M.J.; Lynch, J.S. Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *J. Educ. Psychol.* **2009**, *101*, 765–778. [CrossRef]
7. Oakhill, J.V.; Cain, K.; Bryant, P.E. The dissociation of word reading and text comprehension: Evidence from component skills. *Lang. Cogn. Process.* **2003**, *18*, 443–468. [CrossRef]
8. Dicataldo, R.; Roch, M. Direct and indirect pathways of variation in length of exposure to the majority language, cognitive and language skills in preschoolers' listening narrative comprehension. *Children* **2021**, *8*, 636. [CrossRef]
9. Florit, E.; Roch, M.; Altoè, G.; Levorato, M.C. Listening comprehension in preschoolers: The role of memory. *Br. J. Dev. Psychol.* **2009**, *27*, 935–951. [CrossRef]
10. Florit, E.; Roch, M.; Levorato, M.C. Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Process.* **2011**, *48*, 119–138. [CrossRef]
11. Florit, E.; Roch, M.; Levorato, M.C. The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Appl. Psycholinguist.* **2013**, *34*, 395–415.
12. Strasser, K.; del Río, F.D. The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Read. Res. Q.* **2014**, *49*, 169–187. [CrossRef]
13. Lepola, J.; Lynch, J.; Laakkonen, E.; Silvén, M.; Niemi, P. The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Read. Res. Q.* **2012**, *47*, 259–282. [CrossRef]
14. Alonzo, C.N.; Yeomans-Maldonado, G.; Murphy, K.A.; Bevens, B. Language and reading research consortium (larrc), predicting second grade listening comprehension using prekindergarten measures. *Top. Lang. Disord.* **2016**, *36*, 312–333. [CrossRef]
15. Currie, N.K.; Cain, K. Children's inference generation: The role of vocabulary and working memory. *J. Exp. Child Psychol.* **2015**, *137*, 57–75. [CrossRef] [PubMed]
16. Kim, Y.S. Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *J. Exp. Child Psychol.* **2016**, *141*, 101–120. [CrossRef] [PubMed]
17. Valentini, A.; Serratrice, L. Longitudinal predictors of listening comprehension in bilingual primary school-aged children. *Lang. Learn.* **2023**, *73*, 5–46.
18. Strickland, D.S.; Riley-Ayers, S. Early literacy: Policy and practice in the preschool years. *Presch. Policy Brief* **2006**, *10*, 259–282.
19. Babayiğit, S. The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *J. Res. Read.* **2014**, *37*, S22–S47.
20. Dickinson, D.K.; Neuman, S.B. *Handbook of Early Literacy Research*; The Guilford Press: New York, NY, USA, 2007.
21. de Jong, P.F.; van der Leij, A. Effects of phonological abilities and linguistic comprehension on the development of reading. *Sci. Stud. Read.* **2002**, *6*, 51–77. [CrossRef]

22. Oakhill, J.V.; Cain, K. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Sci. Stud. Read.* **2012**, *16*, 91–121. [CrossRef]
23. Seigneuric, A.; Ehrlich, M.F. Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Read. Writ.* **2005**, *18*, 617–656. [CrossRef]
24. Verhoeven, L.; van Leeuwe, J. Prediction of the development of reading comprehension: A longitudinal study. *Appl. Cogn. Psychol.* **2008**, *22*, 407–423. [CrossRef]
25. Meara, P. The vocabulary knowledge framework. *Vocab. Acquis. Res. Group Virtual Libr.* **1996**, *5*, 1–11.
26. Bonifacci, P.; Barbieri, M.; Tomassini, M.; Roch, M. In few words: Linguistic gap but adequate narrative structure in preschool bilingual children. *J. Child Lang.* **2018**, *45*, 120–147. [CrossRef]
27. Fiani, R.; Henry, G.; Prévost, P.; Bohnacker, U.; Gagarina, N. Narrative comprehension in Lebanese Arabic-French bilingual children. In *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*; Bohnacker, U., Gagarina, N., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2020; pp. 31–60.
28. De Lamo White, C.; Jin, L. Evaluation of speech and language assessment approaches with bilingual children. *Int. J. Lang. Commun. Disord.* **2011**, *46*, 613–627.
29. Bialystok, E.; Luk, G.; Peets, K.F.; Sujin, Y. Receptive vocabulary differences in monolingual and bilingual children. *Biling. Lang. Cogn.* **2010**, *13*, 525–531.
30. Morales, J.; Calvo, A.; Bialystok, E. Working memory development in monolingual and bilingual children. *J. Exp. Child Psychol.* **2013**, *114*, 187–202. [PubMed]
31. Roch, M.; Hržica, G. Narrative comprehension by Croatian-Italian bilingual children 5–7 years old. *Dev. Narrat. Compr. Multiling. Assess. Instrum. Narrat.* **2020**, *61*, 171–192.
32. Nicoladis, E. Cross-linguistic transfer in deverbal compounds of preschool bilingual children. *Biling. Lang. Cogn.* **2003**, *6*, 17–31.
33. Oller, D.K. The distributed characteristic in bilingual learning. In *ISB4: Proceedings of the 4th International Symposium on Bilingualism*; Cohen, J., McAlister, K.T., Rolstad, K., MacSwan, J., Eds.; Cascadilla Press: Somerville, MA, USA, 2005; pp. 1744–1749.
34. Oller, D.K.; Eilers, R.E. *Language and Literacy in Bilingual Children*; Multilingual Matters: Clevedon, UK, 2002; Volume 2.
35. Geva, E.; Farnia, F. Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Read. Writ.* **2012**, *25*, 1819–1845.
36. Scheffner Hammer, C.; Lawrence, F.R.; Miccio, A.W. Exposure to English before and after entry into Head Start: Bilingual children's receptive language growth in Spanish and English. *Int. J. Biling. Educ. Biling.* **2008**, *11*, 30–56.
37. Dicataldo, R.; Roch, M. Are the effects of variation in quantity of daily bilingual exposure and socioeconomic status on language and cognitive abilities independent in preschool children? *Int. J. Environ. Res. Public. Health* **2020**, *17*, 4570–4592.
38. Thordardottir, E. The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* **2011**, *15*, 426–445.
39. McWilliam, N. *What's in a Word? Vocabulary Development in Multilingual Classrooms*; Stylus Publishing, LLC.: Sterling, VA, USA, 1998.
40. Roth, F.P.; Speece, D.L.; Cooper, D.H. A longitudinal analysis of the connection between oral language and early reading. *J. Educ. Res.* **2002**, *95*, 259–272.
41. Cain, K.; Oakhill, J.V.; Barnes, M.A.; Bryant, P.E. Comprehension skill, inference-making ability, and their relation to knowledge. *Mem. Cognit.* **2001**, *29*, 850–859. [PubMed]
42. Cleave, P.L.; Girolametto, L.E.; Chen, X.; Johnson, C.J. Narrative abilities in monolingual and dual language learning children with specific language impairment. *J. Commun. Disord.* **2010**, *43*, 511–522. [PubMed]
43. Iluz-Cohen, P.; Walters, J. Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Biling. Lang. Cogn.* **2012**, *15*, 58–74.
44. Pesco, D.; Bird, E.K.R. Perspectives on bilingual children's narratives elicited with the Multilingual Assessment Instrument for Narratives. *Appl. Psycholinguist.* **2016**, *37*, 1–9.
45. Bohnacker, U. Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Appl. Psycholinguist.* **2016**, *37*, 19–48.
46. Boerma, T.; Leseman, P.; Timmermeister, M.; Wijnen, F.; Blom, E. Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *Int. J. Lang. Commun. Disord.* **2016**, *51*, 626–638.
47. Hoff, E.; Core, C.; Place, S.; Rumiche, R.; Señor, M.; Parra, M. Dual language exposure and early bilingual development. *J. Child Lang.* **2012**, *39*, 1–27.
48. Gutiérrez-Clellen, V.F. Narratives in two languages: Assessing performance of bilingual children. *Linguist. Educ.* **2002**, *13*, 175–197.
49. Squires, K.E.; Lugo-Neris, M.J.; Peña, E.D.; Bedore, L.M.; Bohman, T.M.; Gillam, R.B. Story retelling by bilingual children with language impairments and typically developing controls. *Int. J. Lang. Commun. Disord.* **2014**, *49*, 60–74.
50. Uccelli, P.; Páez, M.M. Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Lang Speech Hear Serv Sch.* **2007**, *38*, 225–236.
51. Rodina, Y. Narrative abilities of preschool bilingual Norwegian-Russian children. *Int. J. Biling.* **2017**, *21*, 617–635.
52. Roch, M.; Florit, E.; Levorato, C. Narrative competence of Italian–English bilingual children between 5 and 7 years. *Appl. Psycholinguist.* **2016**, *37*, 49–67.

53. Pearson, B.Z. Narrative competence among monolingual and bilingual school children in Miami. In *Language and Literacy in Bilingual Children*; Oller, D.K., Eilers, R.E., Eds.; Multilingual Matters: Clevedon, UK, 2002; pp. 135–174.
54. Stæhr, L.S. Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Stud. Second. Lang. Acquis.* **2009**, *31*, 577–607.
55. Brulard, I.; Carr, P. French-English bilingual acquisition of phonology: One production system or two? *Int. J. Biling.* **2003**, *7*, 177–202.
56. Yip, V.; Matthews, S. Syntactic transfer in a Cantonese–English bilingual child. *Biling. Lang. Cogn.* **2000**, *3*, 193–208.
57. Paradis, J.; Navarro, S. Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *J. Child Lang.* **2003**, *30*, 371–393.
58. Melby-Lervåg, M.; Lervåg, A. Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *J. Res. Read.* **2011**, *34*, 114–135.
59. Cummins, J. Linguistic interdependence and the educational development of bilingual children. *Rev. Educ. Res.* **1979**, *49*, 222–251.
60. Nicoladis, E. Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Biling. Lang. Cogn.* **2006**, *9*, 15–32.
61. Marchman, V.A.; Martínez-Sussmann, C.; Dale, P.S. The language-specific nature of grammatical development: Evidence from bilingual language learners. *Dev. Sci.* **2004**, *7*, 212–224. [PubMed]
62. Kim, A.Y.; Park, A.; Lust, B. Simultaneous vs. successive bilingualism among preschool-aged children: A study of four-year-old Korean–English bilinguals in the USA. *Int. J. Biling. Educ. Biling.* **2018**, *21*, 164–178.
63. Carlson, S.M.; Moses, L.J. Individual differences in inhibitory control and children's theory of mind. *Child Dev.* **2001**, *72*, 1032–1053.
64. Dunn, L.M.; Dunn, L.M. *Peabody Picture Vocabulary Test-III*; American Guidance Service: Circle Pines, MN, USA, 1997.
65. Stella, G.; Pizzoli, C.; Tressoldi, P.E. *Peabody test di Vocabolario Recettivo Omega Edizioni*; Omega: Torino, IT, USA, 2000.
66. Levorato, M.C.; Roch, M. *Valutare la Comprensione del Testo Orale: Il TOR 3-8 [Evaluation of Listening Comprehension: The TOR 3-8]*; Organizzazioni Speciali: Florence, IT, USA, 2007.
67. Vettori, G.; Bigozzi, L.; Incognito, O.; Pinto, G. Relationship between oral narrative and vocabulary skills among bilingual language-minority children and their monolingual peers in primary school. *Int. J. Biling.* **2022**, 13670069221089064.
68. Pearson, B.Z. Social factors in childhood bilingualism in the United States. *Appl. Psycholinguist.* **2007**, *28*, 399–410. [CrossRef]
69. Kapalková, S.; Polišenská, K.; Marková, L.; Fenton, J. Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Appl. Psycholinguist.* **2016**, *37*, 145–164.
70. Lao, C. Parents' attitudes toward Chinese–English bilingual education and Chinese-language use. *Biling. Res. J.* **2004**, *28*, 99–121. [CrossRef]

education sciences

MDPI

*Article*

# Where SLA and Language Education Meet: The Transfer from Task-Based Needs Analysis to Task Design

**Roger Gilabert**

Department of Modern Languages and Literatures and English Studies, University of Barcelona, 08007 Barcelona, Spain; rogergilabert@ub.edu

**Abstract:** This theoretical article sets a goal to investigate how task-based needs analysis (TBNA) as part of instructed second language acquisition (ISLA) may mediate between constructs and concepts in second language acquisition (SLA) and task-based design. It is claimed that as an instantiation of instructed second language acquisition (ISLA), TBNA in task-based language teaching (TBLT) may bring together decision-making during task design and what is known about SLA products and processes. The article then explores some of the key SLA concepts in ISLA around the constructs of input, intake, knowledge, and output and relates them to task-based research. Thirdly, TBNA is defined and analyzed in terms of how it may inform task and syllabus design. Finally, SLA constructs are directly associated with pedagogic task design that springs from TBNA. After a close inspection of all dimensions of TBNA, the article concludes that task design based on detailed TBNA has indeed the potential to bring reality into instruction through thorough task descriptions and methods, as well as to trigger well-known and specific processes that bring SLA and language education closer.

**Keywords:** task-based language teaching; task-based needs analysis; second language acquisition theories

**Citation:** Gilabert, R. Where SLA and Language Education Meet: The Transfer from Task-Based Needs Analysis to Task Design. *Educ. Sci.* **2023**, *13*, 1015. https://doi.org/10.3390/educsci13101015

Academic Editor: Martin Howard

Received: 6 August 2023
Revised: 23 September 2023
Accepted: 25 September 2023
Published: 8 October 2023



## 1. Introduction

SLA and language education have often been at odds with each other. Back at the beginning of this century, Long [1] already suggested that: "Most SLA theories and most SLA theorists are not primarily interested in language teaching, and in some cases not at all interested" (p. 17). In the same fashion, Gregg [2] reminded us that "... the connections between SLA theory and L2 instruction are indirect, complex, and tenuous at best when they are not non-existent." (p. 153). This may be the case because the goal of SLA theorists is to identify what is *necessary* and *sufficient* in order to acquire a second or foreign language, while the goal of language teachers and theorists is to identify the most *efficient* practices, procedures, and conditions that will quickly and effortlessly lead to language learning [1,3].

In between those two fields with apparently divergent goals, instructed SLA (ISLA) tries to shed light on how SLA products and processes, as well as any practices tapping into them, may be associated with second language instruction. As defined by Loewen in the Encyclopedia of Applied Linguistics [4]: "Instructed second language acquisition (ISLA) is a subfield of second language acquisition (SLA) that investigates any type of second language (L2) learning or acquisition that occurs as a result of the manipulation of the L2 learning context or processes." (Second Language Acquisition section). A well-consolidated line of research within ISLA originates from the task-based language teaching (TBLT) approach, which we take as an instantiation of language education in this article. Since its conception, TBLT has been a research-based teaching approach (a 'researched pedagogy' in [5]) with strong underlying principles that have primarily, although not exclusively, fed on cognitive-interactionist SLA theories and constructs to explain L2 performance and development. The TBLT research agenda has evolved from the early studies in interaction [6] in the early 90s and performance studies in the mid-90s [7,8], and in the 2000s, with a main interest in complexity, fluency, and accuracy to the design of tasks for the acquisition of different

dimensions of language. In theoretical terms, the field has moved from almost exclusively interaction and performance theories as reflected in the work by Long, Gass, Robinson, Skehan, and Ellis, to investigating the interface between TBLT and other neighboring areas of interest, such as computer-assisted language learning constructs and concepts [9] as well as theory coming from pragmatics [10] and phonetics and phonology theories [11], writing studies [12,13], and multimedia learning [14], among others, all of which have enriched the TBLT field in the last two decades.

*Task-Based Needs Analysis*

Out of the many aspects that TBLT is concerned with, here, we focus on needs analysis as a first and crucial step into program development, since the decisions about what to teach determine every other aspect of syllabus construction such as our pedagogical choices, methodological approach, or testing. In this article, we claim that the connections between SLA and TBLT may be specified, mediated, and informed by task-based needs analysis (TBNA). TBNA is a professional, in-depth inquiry into what learners need to learn, and it will identify the specific tasks, the skills, and the language associated with them, which a particular community of learners needs to be able to perform in their second or foreign language(s). Briefly, needs analysis is about 'what' learners need to learn. Task design informed by methodological principles and pedagogical procedures (and so the 'how') will follow, and it will be adjusted to try to alter the mental processes involved in second language acquisition with the hope of facilitating and accelerating the progress of learners' interlanguage systems. Task-based needs analysis (TBNA), often the first step in task-based syllabus design, is at the very core of TBLT, and it is responsible for bringing reality into instruction.

We will claim that by carefully identifying and describing the tasks that need to be taught, task-based design may be supported and facilitated in combination with task-based methodological principles and pedagogical procedures, thus bridging the gap between what is the reality of second language use and second language instruction. Even if task-based needs analysis does not say anything directly about the mental processes that engaging in pedagogic task performance may entail, and how these may eventually lead to acquisition, needs analysis can be a powerful and reliable source of information for decision-making for pedagogic task design and instruction that will tap into what we know about second language acquisition.

In order to reflect on the mediating role of task-based needs analysis between SLA and language instruction, we will first define some key concepts emerging from SLA that are particularly associated with language education in general and with TBLT in particular. Such concepts will revolve around the products and processes associated with input, intake, knowledge, and output. We will then zoom in on how, as part of TBLT, task-based needs analysis can inform pedagogic task design. Then, we will try to bring together SLA processes and task design, and we will conclude with some final reflection on what is left to be resolved.

## 2. SLA, Instructed SLA, and Task-Based Language Teaching (TBLT)

Numerous handbooks, books, and articles have defined the goals and scope of SLA as a field [15–19], among others. It will suffice to say here that SLA is a consolidated, diverse, and growing research area exploring the learning and loss of second languages by people of all ages and with distinct individual differences (e.g., motivation, cognitive abilities, or purposes). In both formal and informal contexts, including second and foreign language contexts as well as lingua franca settings, SLA research is interested in both individual and whole community learning. The issues and topics SLA researchers are interested in are many and varied and may depend on whether theoretical accounts are linguistic, cognitive, or social in nature [20].

While drawing on SLA theories in general, the field of ISLA has been particularly interested in concepts and constructs that are more closely related to language instruction,

and so in SLA products and processes involved in input, output, and interaction, the role of instruction and feedback, the intentional/incidental and the implicit/explicit debates in relation to L2 learning, and learners' individual differences, among many others. Less directly related but highly relevant to teachers are issues of transfer, the Critical Period, developmental sequences, rule-based and memory-based features of the linguistic system, or ultimate attainment. Yet, as Long [21] pointed out: "SLA theories may provide insight into putatively universal methodological principles, in other words, while saying little or nothing about the inevitable particularity of appropriate classroom pedagogical principles, in which the local practitioner, not the SLA theorist, should always be the expert" (p. 19).

As opposed to focus-on-form approaches that organize syllabi around linguistic units or focus-on-meaning approaches that do so around conceptual or subject-matter units, the central argument in focus-on-form approaches in general and task-based language teaching (TBLT), in particular, is that through the performance of real-life communicative tasks, a second language (L2) can be best acquired [3,22–24]. A pedagogic task is defined here as a differentiated process that connects with real-life activity, with a specific goal and a series of steps, which draws on learners' cognitive and communicative resources for its completion. One key reason for that is that communicative task performance may lead learners to notice and retain certain aspects of the language they are using, and this may cause changes in and development of their interlanguage systems. TBLT theory and research have been concerned with the conditions under which learners learn new forms, and when and how they acquire them. In terms of pedagogy, research into tasks can help in the identification of those task characteristics that may influence language processing for comprehension, production, and learning, hence providing empirical evidence for the pedagogic task and materials design. It has been shown that teachers and syllabus designers may design and generate conditions of performance that will gear learners' attention toward specific aspects of the language and, in this way, promote opportunities for learning and development of their interlanguage system and their overall communicative competence.

What are key SLA constructs and processes related to input, intake, knowledge, and output? How does TBLT integrate those processes? In the following sections, we will review some of the key concepts and processes (in italics) coming out of second language acquisition theories and their association with TBLT as an option in ISLA.

### *2.1. Constructs and Processes Related to Input, Intake, Knowledge, and Output: From SLA Theories to TBLT*

#### 2.1.1. Input

There is a clear consensus that *input* is essential to learning [25]. Exposure to input has been shown to trigger both micro-processes and macro-processes [26]. Examples of the former include *noticing* [27], and *cognitive comparison* [26], which are conceptualized as occurring with awareness, by means of which certain elements of the input receive selective attention and are briefly and momentarily compared to whatever other knowledge the learner may possess. Just mere exposure to input in any of its forms (e.g., text, video, audio, or a combination of them, or conversation) will enable learners to select and begin to process certain elements in the input. Input selection will depend on multiple factors such as the learner's developmental readiness, their internal syllabus, their communicative needs, perceived task demands, L2 proficiency, L1 features and L1-L2 distance, processing capabilities, motivation, task construal, and agency, among many other factors or combinations of factors (for a review, see [28,29]).

Language learners may be exposed to second language "potentially processable language data" or input outside the classroom by chance or in the classroom by design [30]. In the latter case, the amount, quality, and access to input can vary considerably across teaching approaches, practices, and contexts [23,30]. Nonetheless, many EFL learners in regular classrooms around the world receive minimal amounts of input with only a few hours per week, typically in a decontextualized manner and without enough opportunities for actual communicative practice, and with access being limited to mostly the classroom

setting. By its nature, input is external to the learner, and what learners notice and further process will depend on two main sets of factors: one, the characteristics of linguistic elements in the input (e.g., frequency, saliency, communicative value of the forms–what Nick Ellis calls the 'usual suspects') and, two, what the learner brings to the task (e.g., what they are ready to learn according to their internal syllabus and proficiency in the L2, the interplay between their communicative needs and perceived task demands, their interest and motivation and agency among many other factors or combinations of factors). Exposure to comprehensible input (e.g., input that is at learners' reach), preferably modified by means of elaborated input that facilitates understanding of new or difficult items (as opposed to simplified input that eliminates them), has been claimed as crucial to learning [3]. More recently, SLA research has brought to the forefront the importance of multimodal input, which is clearly a part of our everyday experience through audiovisual products such as TV series or video games mostly outside but more and more often inside the classroom. Theories of multimodal input [31] and dual-coding [32] are often cited as support to the idea that multimodal input is a positive contributor to second language development in the areas of grammar, vocabulary, formulaic sequences, pragmatics, and phonetics. Such theories have advanced the idea that input processed through one channel (e.g., audio) can be reinforced, without interference, by input processed through a different channel (e.g., text in captions). In addition to positive evidence, negative evidence in the form of corrective feedback (e.g., ranging from telling learners directly that they are wrong to indirectly and implicitly prompting self-corrections) has been claimed to also contribute to second language acquisition [33]. Evidence of the effectiveness of corrective feedback has been shown in both the oral [34] and written [35] modes. The debate is ongoing regarding the degree of explicitness that is required with direct, explicit explanations on one end, and indirect and more implicit recasts on the other.

In sum, not everything learners are exposed to gets selected to be processed, and even when they notice certain elements from the input, there is no guarantee what they have extracted from the input will be processed any further. As we learned from [36], for noticed features in the input to be processed in more depth, input needs to become *intake*.

#### 2.1.2. From Input to Intake

While micro-processes happen with some degree of awareness, underlying macro-processes are also activated during exposure to input. For instance, input internalization or *intake* and *intake processing* is a process by which a subset of data in the linguistic input is held in working memory (WM) for further processing against prior knowledge [36]. According to [37], for the conversion of input into intake, *form-function mapping* is necessary, and it constitutes the continual and cumulative process of linking concepts to phonological forms and grammatical functions. Establishing form–meaning connections can be affected by different factors such as the nature of the input, learner characteristics, and the learning context and instruction [38]. As opposed to the effects of input or output on the acquisition, both from a general ISLA and a task-based perspective, intake as a product and as a process has been considerably under-researched. Other processes include *item learning* and *chunking*, where some elements may be picked from the input and taken into long-term memory as unanalyzed material for later processing [21]; *analysis,* the continual and cumulative process by which linguistic and conceptual representations become susceptible to inspection [39]; *hypothesis formation* and *testing*, which entails the generation of learner's internal theories about how the language works [40,41].

As suggested by [13], in the context of TBLT, the input to a pre-task or task often stands as positive evidence of the target language or is presented as corrective feedback, and it is typically part of a dynamic, goal-oriented, input-output-feedback cycle whether in oral or written modes, or a combination of them. In task-based design, content and language input is typically present in text, audio, video, and their combinations (e.g., audiovisual material with L2 captions) in the pre-task, and it makes exposure possible with the hope that either ideas or language or both will be recycled in the performance

of the main task. Ellis et al. [22] have listed some examples of pre-tasks that are carriers of such input. In them, input is subject to intervention by (1) increasing the presence of target language features through *input flooding*; (2) by highlighting certain elements in the input through *input enhancement*; (3) by facilitating comprehension and retention of new or difficult forms through *input elaboration*; (4) or by pushing learners towards the processing of specific forms that that become *task essential language* without which the task cannot be completed. Input is not only provided as positive evidence but also part of the *corrective feedback* that teachers or more competent users of the language provide reactively during learners' performance of the task typically in the form of recasts, or post-actively once they have completed it. These interventions on the input (the external product) are meant to generate an effect on input processing in learners that will hopefully engage second language acquisition processes leading to development [42,43]. The link between input modification during task design and second language acquisition has been particularly proven for input enhancement (see [43] for an example of the effects of input enhancement on grammar learning), with less literature behind input flooding, input elaboration, or task essential language. By enhancing certain elements in the input (e.g., typically verb endings, individual words, sounds, or collocations) teachers and designers seek processes such as *noticing* and *cognitive comparison* to be more likely engaged during exposure, and this will lead to other processes such as *intake* (chunking or unanalyzed item learning, analysis, hypothesis testing, form-function mapping) and *knowledge processing* (internalization, restructuring, and consolidation). With some caveats, the facilitating effects of both oral and written corrective feedback on second language development has been extensively documented over the years (see for example [44]).

#### 2.1.3. Knowledge and Knowledge Processing

*Knowledge* and *knowledge processing* is about internalizing, modifying, and consolidating L2 knowledge [45], and it is one of the areas that has received little attention in TBLT. *Restructuring*, the abrupt process by means of which some aspects of interlanguage become more efficiently represented in the learner's mind, may lead to *grammatization* and *syntactization* in the L2 [46], *automatization* [47], and *consolidation* [48] of memories. Certainly, TBLT research is in great need of studies in the area of knowledge processing. An explanation for this lack of studies could be that knowledge processing is not open to direct inspection, despite the considerable advances in our understanding and measurement of implicit and explicit knowledge in [49].

Regarding *output* and *output processing*, since the mid-1990s, research has made considerable efforts to measure the effects of manipulating task design features on both L2 performance (operationalized as complexity, accuracy, and fluency, or CAF [50,51] and L2 acquisition. From an acquisitional perspective, the output hypothesis [52] has posited that "the act of producing language (speaking or writing) constitutes, under certain circumstances, part of the process of second language learning" (p. 471), pushing learners from semantic processing in comprehension to more syntactic processing in production. Syntactic processing demands higher attention to linguistic forms and deeper language analysis, with potentially consequential effects on language development. The production of output is postulated to trigger the whole range of beneficial processes, such as noticing and focusing on form, hypothesis testing, metalinguistic reflection, and automatization [53–55]. What the TBLT paradigm precisely offers is the optimal context for the sustained and context-embedded type of output practices that theorists consider vital for L2 development. Noticing linguistic problems can occur in both oral and written tasks, although researchers have claimed that the written mode poses advantages for such processes to take place [13].

As opposed to more traditional approaches that have typically assumed the idea that what is taught, typically explicitly, is what gets learned, TBLT, since its origins, has aligned with SLA theories that provide evidence that second language acquisition is a slow and complex phenomenon that requires numerous and meaningful input-output-feedback cycles over an extended period of time in order for it to come to fruition. TBLT

advocates have been aware that whether all the processes we have just revised are engaged or not during task performance will depend on multiple factors. While unable to fully predict the kind of language products and processes that will be engaged during task performance; however, task-based researchers, task-based designers, and instructors have made an attempt to make the process of second language acquisition predictable and at least partially manageable through task design and instruction. In the following sections, we will claim that NA may be instrumental to such decision-making by teachers/designers and that NA may actually be useful in predicting the kinds of SLA processes that may be involved during task performance.

## 3. Bridging the Gap between NA and Pedagogic Task Design

Syllabus design is the instantiation of our theories about how languages are learned [23]. The way we conceptualize what language is and how it works will most likely determine the type of units that we choose to organize our syllabus around. In turn, our choice of units for our syllabus design will largely determine every other aspect of the syllabus, such as how the units will be graded and sequenced, how they will need to be designed pedagogically, the methods that we will need to use in order to teach them, as well as the assessment methods required to assess those units and evaluate the program in which they appear. If for instance, we choose content or conceptual units as in English as a medium of instruction (EMI), our sequencing will most likely be conducted with the logic of the subject matter. In math, for example, from addition to equations and algorithms, learners are presented with units increasing in the complexity and intricacy of mental operations they require. Typically, units will be pedagogically designed to make such content available to and manageable for learners, and learners will be tested on their mastery of such content. Finally, the program will be judged on the basis of whether it achieves the goals of having learners put content and language to good use outside the EMI classroom. If instead of content/conceptual units, we select linguistic units as the organizing principle of our syllabus structure, units will quite likely be organized according to some notion of 'difficulty' or 'usefulness', and pedagogic design will be tailored around the deductive or inductive teaching of those units, the mastery of those linguistic units will be assessed with language-based tests, and the effectiveness of the structural/lexical syllabus will be tested against what learners end up knowing about the L2 system. In TBLT, syllabus design has often taken a cognitive-interactive approach to syllabus construction that revolves around pedagogic tasks. Additionally, tasks are sequential and susceptible to pedagogic and/or research intervention. Tasks are dynamic processes, which are susceptible to modifications and adaptations in ever-changing social, academic, and professional environments. Although not always, the tasks that constitute task-based programs are versions of the real target tasks detected by means of systematic NA. From the detailed descriptions of such target tasks, pedagogic tasks are created that will prepare learners for the typically highly complex tasks that people need to perform in their everyday personal or professional environments. While still under debate, sequencing is often decided upon on the basis of cognitive task complexity, and tasks are taught mostly inductively in pre-task-task–post-task cycles where language is embedded in all phases of the task. Ideally, the performance of pedagogic tasks is assessed in terms of task completion and their approximation to real target task performance. Finally, program evaluation checks whether pedagogic tasks have actually helped learners prepare for real target task performance outside the classroom. While task-based NA is a professional inquiry into 'what' a specific community of learners needs to be able to do in terms of tasks, and so it is the first step into syllabus design, in [56,57], we claimed that NA may actually inform all aspects of program development.

In those two chapters [56,57], the issue of transfer from NA to task design was thoroughly investigated. It was seen that the information retrieved from careful and detailed needs analysis can inform all other aspects of syllabus design, that is, pedagogic task selection, sequencing, pedagogic design, methodological implementation, assessment, and

program evaluation. By means of multiple sources and methods, the task dimensions NA may investigate are divided into seven broad categories: (1) 'general aspects' the goals associated with the tasks, the frequency with which the task is performed, its outcome(s), task-related topics, sub-/target tasks, and how it fits into the general picture of the domain; (2) 'participants and interaction' is a dimension linked to information exchange and communication between participants involved in a task, the rules of interaction, psycholinguistic aspects, intercultural communicative aspects, and non-verbal aspects [6,58,59]; (3) the 'physical space' where tasks take place enquires into factors that have to do with the spatial and psychosocial setting of tasks [59]; (4) the 'cognitive demands' dimension tries to tap into tasks' attentional and memory demands, mental processes, and perceived difficulty of tasks, as well as the recruiting of higher and lower order skills [60,61]; (5) tasks' 'linguistic demands' include the linguistic resources necessary to complete a task [62,63]; (6) communication and technology seeks to retrieve information on the communication channels and technological tools and platforms associated with performing a task [9]; and (7) the 'other dimensions' category grouped together assessment, task support, and tasks' non-verbal aspects, attitudinal values, concepts, and norms, as well as sequence of procedures.

In terms of task selection, if the needs analysis is properly conducted [64,65], a list of target tasks and associated sub-tasks should contain information about the frequency, difficulty, and need for training (based on their importance or priority) of each of the tasks. Gilabert & Malicka [56,57] suggested 'the need for training' as a reliable choice whenever possible, defined as the time and effort that a person needs to invest in order to master task performance. An additional criterion may be the degree of perceived difficulty and complexity by domain experts, that is, people who did not know how to perform a task in the L2 and later mastered it. This type of information is collected during needs analysis, and it can greatly facilitate the decision-making process about which tasks should be selected for the syllabus. Some tasks may be perceived as difficult or higher stakes by experts and hence require more mental effort. Those target tasks may be better targets for selection than simple tasks or sub-tasks that may be more common but may not require so much training.

As for pedagogic task design, TBNA can clearly help to identify task goals, that is, the ultimate objective of the real-life task, such as 'solving a problem' or 'reaching an agreement'. The information gathered in a TBNA can inform decisions about aspects such as the number of participants and the type of performance (monologic vs. dialogic), or the information flow between them (one-way, two-way, multiple-way). Furthermore, participant observation in TBNA can help to identify the status of parties involved in professional tasks, which will help with the creation of roles based on actual psychological profiles and positions of power. Additionally, important for task design is what TBNA can say about the linguistic demands and skills called upon by tasks. The analysis may also include information about the language associated with the task in terms of specific terminology, discourse features and grammatical features, or speech acts required by the task. In focused pedagogic tasks, a specific item or a number of items may be targeted when they are known to emerge from and be required by the task, and they are typically detected and selected after several iterations of the task. During pedagogic task design of such focused tasks, target language aspects may be addressed preemptively (e.g., by means of input flooding, input enhancement, or task essential language as focus-on-form techniques) and distributed throughout the pre-task, task, and post-task phases. In unfocused tasks that do not target any specific items, the focus-on form may take the shape of recasts or other forms of reactive feedback as learners run into language problems and miscommunication [26]. From a cognitive point of view, needs analysis provides information on the attention and memory demands real-life tasks place on those performing them. TBNA should help us find out specific attributes of tasks such as the mental operations recruited to perform them, how many pieces of information need to be stored in working memory at the same time, or whether tasks are conducted under time pressure or there is time available to plan. These attributes of real-life tasks can then be translated into pedagogical variables, which can be manipulated in task design. Very importantly for task design, TBNA should also be able to

inform us of how these cognitive factors are perceived in terms of their relative difficulty by those who perform them since this will help with task sequencing. While still largely unresolved, task sequencing may be realized in terms of the cognitive complexity of tasks (as perceived by domain experts during needs analysis), by considering their linguistic difficulty [62,63].

As for task methodology, TBLT has traditionally adhered to some of the "language teaching universals" [22] that have oriented teachers when implementing tasks. The use of 'authentic input', which TBNA can help with during sample collection and associated discourse analysis of the language included in those samples. Additionally, TBNA can help with decisions as to what kind of focus-on form (i.e., techniques such as input elaboration, input flooding, input enhancement, and task essential language) may be associated with each task or each phase of the task (e.g., input flooding in the input included in the pre-task and recasts for the task phase [26]). It may also help advance and predict the difficulties learners may encounter with language and that will require corrective feedback [66] and hence prepare for them. Needs analysis may also help with the pedagogical options to be chosen during the pre-task (e.g., strategic planning or modelling) as well as the task (e.g., by providing information about the number of participants involved in a task, their roles and status, as well as how information may flow among participants and the divergence or convergence of their goals).

TBNA can also contribute to task-based assessment. Semi-structured interviews and task performance observations may provide useful information about what the performance standards of tasks are. As Malicka et al. [67] suggest, assessment tasks that build on insights obtained from TBNA have the potential to mirror authentic situations and are therefore valid indices of candidate preparedness to deal with requirements of tasks encountered in real-life situations.

But what can help us bridge the gap between SLA and decision-making in language education?

## 4. Bridging the Gap between SLA and Pedagogic Task Design

In Section 1, we looked at the connection between SLA theories, ISLA, and TBLT. In Section 2, we explored the connections between SLA, ISLA, and TBLT. In Section 3, we saw how TBNA may inform pedagogic task design, and in this section, we explore how TBNA may establish the link between SLA constructs and pedagogic design, and how it may help out with decision-making decisions during task design. It is important to stress that the point here is not to make a claim that needs analysis will be able to inform exactly about what SLA processes will be activated as a consequence of design but, rather, to speculate from a theoretical standpoint about which processes design will most likely trigger and to what variable extent. This will be achieved by carefully considering what we know about SLA products and processes as well as what we have learned about TBNA and task-based design over the year as discussed in Sections 2 and 3.

What aspects of the *input* do we decide will be targeted during task design and, hence, instruction? What SLA processes can be associated with each of our design decisions as mediated by needs analysis? As we saw in Section 3, TBNA analysis can help extract information about content, skills, and language that may be relevant to the performance of the task. Whether the focus falls more clearly on the content or the form will depend on the perspective and context we are designing tasks for. If designing tasks for a CLIL program, the emphasis on the task may lean towards the mastery of content even if language is also targeted as part of the design. If instead, tasks are being designed for a program conceived from a strong version of TBLT that includes a TBNA, the focus-on form will be more in balance with the focus on content. TBNA can extract very specific information about what language is associated with each task, and it can do so in at least four different ways. Firstly, semi-structured interviews where domain experts are asked to describe the kind of language that each task requires and that they typically describe in 'non-linguistic' jargon. Researchers must interpret such descriptions and classify them into standardized

categories (e.g., vocabulary, pragmatic moves, or formulaic sequences) that they can use as a reference for pedagogic task design. Secondly, non-participating observations may also help researchers describe tasks and their associated language with high precision. Usually, observations are assigned to some of the tasks identified through semi-structured interviews. Lastly, we can also use discourse analysis where samples are collected and analyzed to minute detail (see [68] for an outstanding example of discourse analysis by L2 Korean learners). Such an analysis can provide information about contextual factors, typical choices, and specific language associated with the performance of a task. In the fourth place, we can use recordings or annotations of multiple iterations of the task in order to determine what language L2 learners make attempts at using, or the language they report they do not have but would like to use, during task performance. All of these methods that can potentially be used during TBNA will help with the selection of target linguistic features that are relevant to each task. In this way, resources and efforts will be most efficiently directed to the language that matters for a specific task completion without teaching too much irrelevant language or too little important and task-oriented language. Without losing track of the learner's volition and agency at picking features from the input, this is a design and instructional attempt at initiating the noticing of what matters for the successful completion of the task.

Once linguistic features have been accurately and precisely selected, this can be coupled with the input transformation techniques we saw in Section 2.1. This applies to focused tasks since unfocused tasks would not make any predictions or include any preemptive attention to any particular language during design and would deal with it reactively [22]. Techniques include input flooding, input enhancement, input elaboration, and task essential language in order to bring attention to items that will be necessary to process and partially internalize during the pre-task in order to perform the main task. From a theoretical standpoint, by flooding the text with more examples of the target features we guarantee that a given feature does not appear just once (or a small number of times), is maybe *noticed*, and then goes (as is often the case when noticing happens right in the middle of a conversation), but rather that we have more than one chance of moving it from mere registration to some degree of cognitive comparison (in [26]) and initial *form-function mapping* [37]. This also applies to input enhancement, which will gear attention to the target features over others. As we saw, also in Section 2.1., there are plenty of factors that may explain what learners will end up noticing, but the use of input modification techniques, by themselves or in combination, may help secure at least partial noticing. Although typically applied in the pre-task phase of the task, such techniques and the SLA processes they potentially trigger may be distributed throughout the different phases of the task. In sum, while in unfocused tasks, the focus-on form is left to happen exclusively incidentally, TBNA can inform design in focused tasks in such a way that certain items are targeted. Detected task-related linguistic items during needs analysis can be potentially matched against the same or different linguistic items that emerge from several iterations of the task. Finally, by applying focus-on-form techniques, they will be hopefully processed as they incidentally arise during task implementation by engaging input and intake processing processes, such as noticing mechanisms, cognitive comparison, and form-function mapping.

However, for *input* to become *intake* and so for more in-depth processes to be engaged, conditions that allow for further processing beyond simple registration or noticing will need to be created. Still, at a low level of processing, item learning and non-systematized chunking may be allowed if WM is liberated and enough attention and memory resources are made available. The TBLT literature has provided us with several ways to reduce task demands and reduce cognitive load. Here, we include three ways in which demands on WM may be reduced. Firstly, the inclusion of pre-task planning time in task design has been shown to liberate resources by providing learners with enough time to process the input [69] before task performance in order to predict what they will be saying and doing, even practice and train for it, and engage with the input at ease before task performance. Creating pedagogic tasks with conditions that approximate real task performance is, of

course, of utmost importance since pre-task planning time may not always be available. TBNA is conducive to obtaining information about performance conditions associated with each task. A second option in task design is that of exact task repetition [70], by which going from usually input in the pre-task to output in the task and then going through the same cycle a second time increases familiarity and liberates resources that will allow underlying processes to be engaged and to be stretched or pushed to greater depths. Revisiting the input after the first attempt at performing the task has an impact on the second attempt at performance. A third option is that of multimodal input, in which WM load is reduced by reinforcing audio input (often hard to process at certain levels of proficiency) with written input, which may help with auditory word recognition, segmentation, and mapping of phonemes to orthographical form, all of which can be grouped into the SLA process form-function mapping [71].

Finally, for learners to recruit enough resources from intake and knowledge processing, *output* practice needs to be part of the design of tasks. As we learned from Swain's [72] Output Hypothesis, engaging in output production will serve the function of noticing/triggering by having learners realize the gaps in their knowledge and the discrepancies between what they want to say. [73] delved into the conditions that contribute to such noticing. Output will also lead learners into generating and testing their hypothesis about how the language works, and so the conversation in the L2 will be a testing ground for learners that will help them keep good uses of the language and reject ungrammatical or pragmatically inadequate moves, among others. Hypothesis generation and testing entail a much more in-depth and sophisticated processing of language. Through oral and written interaction, learners may generate output on which they also receive feedback, often in the form of correct input samples or models against which to contrast their own incorrect or immature productions. It is predictable that ongoing input-output-feedback will also trigger analysis (in Bialystok's terms) and some basic rule formation and eventually lead to internalization, restructuring, and modification [45].

## 5. Conclusions and Areas for Future Research Bringing SLA and TBLT Together

In this article, we have advanced the claim that TBNA can mediate what we know about task design and the SLA processes it may activate and generate. We first tackled the relationship between SLA, ISLA, and TBLT, and we said that TBLT is one of the options ISLA that draws heavily on SLA constructs and concepts that are particularly relevant to all aspects of task-based syllabus design. We then zoomed in on constructs and processes related to input, intake, knowledge, and output that are relevant to TBLT. We saw that input is the basic product that, when exposed to it under appropriate conditions, will initiate a series of input processing mechanisms such as noticing and cognitive comparison. Under certain conditions, input can become intake and therefore engage more in-depth processing. We then revised some of the functions of output that may push learners to move beyond noticing in order to generate and test hypotheses. In Section 3, we explored how TBNA may, directly and indirectly, inform pedagogic task design and all aspects of syllabus design. In Section 4, the issue of how SLA processes and task design and instruction may be put together was brought into focus, and suggestions were made as to how TBNA affecting pedagogic design may tap into SLA processes.

As a teaching and research approach drawing on SLA knowledge and hence as an instantiation of ISLA, TBLT tries to build points of connection between what is known about SLA products and processes and what we know about task design and instruction. As part of syllabus design, TBNA can be instrumental to all other aspects of syllabus design. It is obvious that TBNA does not solve all aspects of tasks and syllabus design. The unresolved issue of linguistic difficulty and morphologically complex languages are two examples of that. Such languages add extra processing to the understanding and use of certain forms and so they may shift the balance between content and form. In turn, this may have consequences for task design and how learners may need to engage in SLA processes [63]. In addition, unresolved is the issue of task sequencing still haunting the field. Despite some

initial attempts at tackling the complex issue of sequencing, no model exists that will help us [74]. The exact combination of information about the internal complexity of tasks, their perceived cognitive difficulty, as well as their actual and perceived linguistic demands has not been achieved for appropriate and efficient sequencing. As it is, we cannot currently make any robust predictions about SLA processes in relation to sequencing, which should be sufficiently interesting material for subsequent research.

Over the last two decades, TBNA has certainly proven itself worthy of informing task and syllabus design in meaningful and sophisticated ways. While realistically it takes some initial time and effort at the start of program development, we would like to claim that investing such time and effort has an enormous payoff for design and development. By conceptualizing TBNA in the way that we have suggested in this article, by linking ISLA concepts and constructs to pedagogical task design, we hope to be taking a decisive step in integrating reality into instruction, and hence bringing SLA and language education closer.

**Funding:** Grant PID2020-119009RB-I00 funded by Spanish Ministry of Science and Innovation (MCIN) and National Research Agency (AEI) 10.13039/501100011033.

**Acknowledgments:** This article has been made possible by grant PID2020-119009RB-I00 funded by Spanish Ministry of Science and Innovation (MCIN) and National Research Agency (AEI) 10.13039/501100011033. Also the Generalitat de Catalunya 2021SGR0303. Special thanks to the members of the taskGen research group. Group support and resources from the GRAL (Language acquisition research group) at the University of Barcelona must also be acknowledged.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Long, M.H. Acquisition and teaching. In *Routledge Encyclopedia of Language Teaching and Learning*; Byram, M., Ed.; Routledge: New York, NY, USA, 2000; pp. 5–7.
2. Gregg, K.R. Learnability and SLA theory. In *Cognition and Second Language Instruction*; Robinson, P., Ed.; Cambridge University Press: Cambridge, UK, 2001.
3. Long, M.H. *Second Language Acquisition and Task-Based Language Teaching*; Wiley: Oxford, UK, 2008.
4. Loewen, S. Instructed Second Language Acquisition. In *The Encyclopedia of Applied Linguistics*; Chapelle, C.A., Ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2012. [CrossRef]
5. Samuda, V.; Van den Branden, K.; Bygate, M. *TBLT as a Researched Pedagogy*; John Benjamins Publishing Company: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2018.
6. Pica, T.; Kanagy, R.; Falodun, J. Choosing and using communication tasks for second language instruction. In *Tasks and Language Learning. Integrating Theory and Practice*; Crookes, G., Gass, S., Eds.; Multilingual Matters: Clevedon, UK, 1993.
7. Robinson, P. Task complexity and second language narrative discourse. *Lang. Learn.* **1995**, *45*, 99–140.
8. Skehan, P.; Foster, P. Task type and task processing conditions as influences on foreign language performance. *Lang. Teach. Res.* **1997**, *1*, 185–211. [CrossRef]
9. González-Lloret, M.; Ortega, L. Towards technology-mediated TBLT: An introduction. In *Technology-Mediated TBLT: Researching Technology and Tasks*; González-Lloret, M., Ortega, L., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2014; pp. 1–22.
10. Taguchi, N. Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics* **2006**, *16*, 513–535. [CrossRef]
11. Gurzynski-Weiss, L.; Long, A.Y.; Solon, M. TBLT and l2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Stud. Second. Lang. Acquis.* **2017**, *39*, 213–224. Available online: https://www.jstor.org/stable/26330757 (accessed on 24 September 2023). [CrossRef]
12. Byrnes, H.; Manchón, R. (Eds.) *Task-Based Language Learning: Insights from and for L2 Writing*; John Benjamins: Philadelphia, PA, USA, 2014; p. 312.
13. Gilabert, R.; Manchón, R.M.; Vasylets, O. Mode in Theoretical and Empirical TBLT Research: Advancing Research Agendas. *Annu. Rev. Appl. Linguist.* **2016**, *36*, 117–135. [CrossRef]
14. Peters, E.; Muñoz, C. Introduction to the special issue: Language learning from multimodal input. *Stud. Second. Lang. Acquis.* **2020**, *42*, 489–497. [CrossRef]
15. Doughty, C.; Long, K. (Eds.) *The Handbook of Second Language Acquisition*; Wiley-Blackwell: Hoboken, NJ, USA, 2003. [CrossRef]
16. Gass, S.M.; Mackey, A. (Eds.) *The Routledge Handbook of Second Language Acquisition*, 1st ed.; Routledge: New York, NY, USA, 2012. [CrossRef]
17. Herschensohn, J.; Young-Scholten, M. (Eds.) *The Cambridge Handbook of Second Language Acquisition*; Cambridge Handbooks in Language and Linguistics; Cambridge University Press: Cambridge, UK, 2013. [CrossRef]
18. Larsen-Freeman, D.; Long, M. *An Introduction to Second Language Acquisition Research*; Longman: London, UK, 1991.

19. Noels, K.A.; Ritchie, W.C.; Bhatia, T.K. (Eds.) *Handbook of Second Language Acquisition*; Academic Press: Cambridge, MA, USA, 1998.
20. Mitchell, R.; Myles, F.; Marsden, E. *Second Language Learning Theories*; Routledge: London, UK, 2019.
21. Long, M.H. *Problems in SLA*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2007.
22. Ellis, R. *Task-Based Language Learning and Teaching*; Oxford University Press: Oxford, UK, 2003.
23. Ellis, R.; Skehan, P.; Li, S.; Shintani, N.; Lambert, C. *Task-Based Language Teaching: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2019.
24. Skehan, P. *A Cognitive Approach to Language Learning*; Oxford University Press: Oxford, UK, 1998.
25. Gass, S. *Input, Interaction and the Development of Second Languages*; Erlbaum: Mahwah, NJ, USA, 1997; p. xi + 189.
26. Doughty, C.; Williams, J. *Focus on Form in Classroom Second Language Acquisition*; Cambridge University Press: Cambridge, UK, 1998.
27. Schmidt, R.A. Attention, cognition and second language instruction. In *Cognition and Second Language Instruction*; Robinson, P., Ed.; Cambridge University Press: New York, NY, USA, 2001; pp. 3–32.
28. Han, Z.; Park, E.S.; Combs, C.H. Textual Enhancement of Input: Issues and Possibilities. *Appl. Linguist.* **2008**, *29*, 597–618. [CrossRef]
29. Sharwood Smith, M.A. Input Enhancement in Instructed SLA: Theoretical Bases. *Stud. Second. Lang. Acquis.* **1993**, *15*, 165–179. [CrossRef]
30. Long, M.H.; Robinson, P. Focus on form: Theory, research, and practice. In *Focus on Form in Classroom Second Language Acquisition*; Doughty, C., Williams, J., Eds.; Cambridge University Press: Cambridge, UK, 1998; pp. 15–41.
31. Mayer, R. *Multimedia Learning*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009. [CrossRef]
32. Paivio, A. *Mental Representations: A Dual-Coding Approach*; Oxford University Press: New York, NY, USA, 1986.
33. Gass, S.; Alison, M. *Input, Interaction and Output in L2 Acquisition*; Routledge: New York, NY, USA, 2020.
34. Lyster, R.; Saito, K.; Sato, M. Oral corrective feedback in second language classrooms. *Lang. Teach.* **2013**, *46*, 1–40. [CrossRef]
35. Wang, Y.P.; Han, X.B.; Yang, J. Revisiting the Blended Learning Literature: Using a Complex Adaptive Systems Framework. *Educ. Technol. Soc.* **2015**, *18*, 380–393.
36. Reinders, H. Towards a definition of intake in second language acquisition. *Appl. Res. Engl. Lang.* **2012**, *1*, 15–36. [CrossRef]
37. Van Patten, B.; Uludag, O. Transfer of training and processing instruction: From input to output. *Fuel Energy Abstr.* **2011**, *39*, 44–53. [CrossRef]
38. VanPatten, B.; Williams, J.; Rott, S. Form-meaning connections in second language aquisition. In *Form-Meaning Connections in Second Language Acquisition*; VanPatten, B., Williams, J., Rott, S., Overstreet, M., Eds.; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2004; pp. 1–26.
39. Bialystok, E. Factors in the growth of linguistic awareness. *Child Dev.* **1986**, *57*, 498–510. [CrossRef]
40. Bley-Vroman, R. Hypothesis Testing in Second-language Acquisition Theory. *Lang. Learn.* **2006**, *36*, 356–376. [CrossRef]
41. Chaudron, C. Intake: On Models and Methods for Discovering Learners' Processing of Input. *Stud. Second. Lang. Acquis.* **1985**, *7*, 1–14. [CrossRef]
42. Gass, S.M.; Mackey, A. Input, interaction, and output in second language acquisition. In *Theories in Second Language Acquisition*; Vanpatten, B., Williams, J., Eds.; LEA: London, UK, 2007; pp. 175–200.
43. Lee, S.; Huang, H. Visual input enhancement and grammar learning: A Meta-Analytic Review. *Stud. Second. Lang. Acquis.* **2008**, *30*, 307–331. [CrossRef]
44. Li, S. The effectiveness of corrective feedback in SLA: A meta-analysis. *Lang. Learn.* **2010**, *60*, 309–365.
45. Leow, R.P. *Explicit Learning in the L2 Classroom: A Student-Centered Approach*, 1st ed.; Routledge: New York, NY, USA, 2015. [CrossRef]
46. Cadierno, T.; Robinson, P. Language typology, task complexity and the development of L2 lexicalization patterns for describing motion events. *Annu. Rev. Cogn. Linguist.* **2009**, *7*, 245–276. [CrossRef]
47. Dekeyser, R. Skill acquisition theory. In *Theories in Second Language Acquisition*; Van Patten, B., Williams, J., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2007; pp. 97–113.
48. McLaughlin, B. "Conscious" versus "Unconscious" Learning. *TESOL Q.* **1990**, *24*, 617–634. [CrossRef]
49. Andringa, S.; Rebuschat, P. New directions in the study of implicit and explicit learning: An introduction. *Stud. Second. Lang. Acquis.* **2015**, *37*, 185–196. [CrossRef]
50. Robinson, P. *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*; John Benjamins: Hoboken, NJ, USA, 2011.
51. Skehan, P. Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Appl. Linguist.* **2009**, *30*, 510–532. [CrossRef]
52. Swain, M. Three functions of output in second language learning. In *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*; Cook, G., Seidlhofer, B., Eds.; Oxford University Press: Oxford, UK, 1995; pp. 125–144.
53. De Bot, K. The psycholinguistics of the output hypothesis. *Lang. Learn.* **1996**, *46*, 529–555. [CrossRef]
54. Erlam, R.; Loewen, S.; Philp, J. The roles of output-based and input-based instruction in the acquisition of L2 implicit and explicit knowledge. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*; Multilingual Matters: Bristol, UK, 2009; pp. 241–261. [CrossRef]

55. Uggen, M. Reinvestigating the Noticing Function of Output. *Lang. Learn.* **2012**, *62*, 506–540.
56. Gilabert, R.; Malicka, A. From needs analysis to task selection, sesign, and sequencing. In *The Cambridge Handbook of Task-Based Language Teaching*; Ahmadian, M., Long, M., Eds.; Cambridge Handbooks in Language and Linguistics; Cambridge University Press: Cambridge, UK, 2021; pp. 226–249.
57. Gilabert, R.; Malicka, A. From needs analysis to task-based design: Methodology, assessment, and program evaluation. In *Handbook of Task-Based Language Teaching and Assessment*; Sudharshan, N.P., Mukhopadhyay, L., Eds.; Springer Nature: New York, NY, USA, 2022; pp. 93–118. [CrossRef]
58. Bosswood, T.; Marriott, A. Ethnography for specific purposes: Teaching and training in parallel. *Engl. Specif. Purp.* **1994**, *21*, 59–79. [CrossRef]
59. East, M. Addressing the intercultural via task-based language teaching: Possibility or problem? *Lang. Intercult. Commun.* **2012**, *12*, 56–73. [CrossRef]
60. Robinson, P. (Ed.) *Cognition and Second Language Instruction*; Cambridge Applied Linguistics; Cambridge University Press: Cambridge, UK, 2001. [CrossRef]
61. Bloom, B.; Englehart MFurst, E.; Hill, W.; Krathwohl, D. Taxonomy of educational objectives: The classification of educational goals. In *Handbook I: Cognitive Domain*; Longmans, Green: New York, NY, USA; Toronto, ON, Canada,, 1956.
62. Gilabert, R.; Castellví, J. Task and Syllabus Design for Morphologically Complex Languages. In *The Cambridge Handbook of Language Learning*; Schwieter, J., Benati, A., Eds.; Cambridge Handbooks in Language and Linguistics; Cambridge University Press: Cambridge, UK, 2019; pp. 527–549.
63. Pallotti, G. An Approach to Assessing the Linguistic Difficulty of Tasks. *J. Eur. Second. Lang. Assoc.* **2019**, *3*, 58–70. [CrossRef]
64. Long, M.H. *Second Language Needs Analysis*; CUP: Cambridge, UK, 2005. [CrossRef]
65. Serafini, E.J.; Lake, J.B.; Long, M.H. Needs analysis for specialized learner populations: Essential methodological improvements. *Engl. Specif. Purp.* **2015**, *40*, 11–26. [CrossRef]
66. Kartchava, E.; Gatbonton, E.; Ammar, A.; Trofimovich, P. Oral corrective feedback: Pre-service English as a second language teachers' beliefs and practices. *Lang. Teach. Res.* **2018**, *24*, 220–249. [CrossRef]
67. Malicka, A.; Gilabert, R.; Norris, J. From needs analysis to task design: Insights from an English for specific purposes context. *Lang. Teach. Res.* **2019**, *23*, 78–106. [CrossRef]
68. Chaudron, C.; Doughty, C.; Kim, Y.; Kong, D.; Lee, J.; Lee, Y.; Urano, K. A task-based needs analysis of a tertiary Korean as a foreign language program. In *Second Language Needs Analysis*; Long, M., Ed.; Cambridge Applied Linguistics; Cambridge University Press: Cambridge, UK, 2005; pp. 225–262. [CrossRef]
69. Ellis, R.; Yuan, F. The Effects of Planning on Fluency, Complexity, and Accuracy in Second Language Narrative Writing. *Stud. Second. Lang. Acquis.* **2004**, *26*, 59–84. [CrossRef]
70. Bygate, M. (Ed.) *Learning Language through Task Repetition*; John Benjamins: Amsterdam, The Netherlands, 2018; Volume 11. [CrossRef]
71. Galimberti, V.; Mora, J.C.; Gilabert, R. Teaching EFL pronunciation with audiosynchronised textual enhancement and audiovisual activities: Examining questionnaire data. In Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices (EPIP7), Grenoble-Alpes, France, 18–20 May 2023; pp. 70–82. [CrossRef]
72. Swain, M. *The Output Hypothesis: Theory and Research*; Routledge: New York, NY, USA, 2005; pp. 495–508. [CrossRef]
73. Izumi, S. Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Stud. Second. Lang. Acquis.* **2002**, *24*, 541–577. [CrossRef]
74. Baralt, M.; Gilabert, R.; Robinson, P. (Eds.) *Task Sequencing and Instructed Language Learning*; Bloomsbury Academic: London, UK, 2014; p. 248.

education sciences

MDPI

*Article*

# The Development of Receptive Language Skills from Captioned Video Viewing in Primary School EFL Learners

**Daniela Avello * and Carmen Muñoz**

Department of Modern Languages and Literatures and English Studies, University of Barcelona, 08007 Barcelona, Spain; munoz@ub.edu
* Correspondence: diag08041@gmail.com

**Abstract:** The evidence obtained to date supports the use of captioned videos for L2 learning purposes, such as vocabulary acquisition and the development of L2 listening skills. However, little research has been conducted with primary school learners, and even less so on the extent to which L2-captioned videos foster the development of L2 reading skills. Thus, the present investigation aimed to determine the extent to which five groups of primary school EFL learners from Chile ($n$ = 96, 9–11 years old, years 4 and 5) benefited from their viewing experience (11 captioned videos) as regards the development of L2 listening skills and L2 reading efficacy (measured at pretest, posttest, and delayed-posttest). In addition, we assessed the influence of L1- and L2-related factors on learners' performance over time (L1 and L2 reading efficacy, L2 vocabulary knowledge, and L2 listening skills). Overall, the results revealed that the treatment led to significant gains in English listening skills and reading efficacy in fourth and fifth graders. However, learners' performance was also found to be predicted by language-related factors, especially L2 vocabulary knowledge. On the whole, the findings of this investigation support the use of age-appropriate captioned videos at primary school to increase children's exposure to the target language and enhance the development of receptive language skills.

**Keywords:** audiovisual input; young learners; foreign language learning; reading skills; listening skills; individual differences

**Citation:** Avello, D.; Muñoz, C. The Development of Receptive Language Skills from Captioned Video Viewing in Primary School EFL Learners. *Educ. Sci.* **2023**, *13*, 479. https://doi.org/10.3390/educsci13050479

Academic Editor: Martin Howard

Received: 10 March 2023
Revised: 4 May 2023
Accepted: 4 May 2023
Published: 9 May 2023



## 1. Introduction

The literature has increasingly provided evidence of the positive effects of audiovisual input on L2 learning [1,2]. Nonetheless, most of the investigations have been conducted with teenagers and adult L2 learners [3]. While various studies on out-of-school learning have demonstrated that primary school students do benefit from their exposure to audiovisual input, it is important to bear in mind that the extent to which learners watch TV in a foreign language from an early age depends on contextual factors and national policies [4–6]. In the case of input-limited contexts, where dubbing is usually the norm, learners' exposure to TV series and movies in a foreign language is very limited [4,7] and restricted to their access to video-sharing websites, pay-per-view TV channels, and streaming platforms. However, viewers may still control the settings and choose to watch the videos in their L1. In fact, the evidence has indicated that children and families are not necessarily aware of the potential benefits of audiovisual input and text support for L2 learning [4]. As a result, the implementation of a principled approach to viewing in the classroom, which develops families' awareness of the actual benefits and the strategies that may be used to improve the experience, might be key to encouraging L2 learners to do this activity at home and increase their exposure to the L2 [8].

There is little evidence on the effects of extensive viewing on the development of receptive L2 skills (e.g., [5]). Overall, the few existing experimental studies that have examined the effects of captioned video viewing on the development of listening skills

by means of fill-in-the-blanks and shadowing tasks have demonstrated that the use of onscreen text enhances speech perception, namely bottom-up processing, after a short intervention [9–11]. Still, it is uncertain whether these results could be replicated with young learners by means of short interventions because, in foreign language contexts, primary school students—whose cognitive skills are still developing—have been found to be less efficient learners than teenagers and adults [12]. As for the development of reading skills as a result of captioned video viewing, most of the studies that have explicitly focused on this issue have been conducted in L1 contexts [13–15]. Therefore, further research is required to determine whether the findings emerging from L1 settings may also be obtained in foreign language contexts. Given these gaps in the literature, the present investigation attempted to determine the extent to which extensive captioned video viewing fostered the development of receptive L2 skills in primary school learners and to explore the role of language-related factors (L1 and L2) on the outcomes. On the whole, the results of this investigation may strengthen the implementation of language learning programs with primary school students in input-limited foreign language learning contexts to enhance the development of receptive language skills [16–18]. These young learners, who are still developing their L1 reading skills, need to increase their exposure to the target language by means of activities that are suited to their characteristics.

### *1.1. Captions and L2 Listening Skills Development*

The literature suggests that listening comprehension tasks may be quite challenging for L2 learners, particularly at lower proficiency levels, because specific information and ideas must be extracted at a speed that listeners cannot control [19,20]. Existing investigations on captioned video viewing have shown evidence of the positive effects of onscreen text on viewing comprehension, which may be associated with speech segmentation and learners' capacity to identify word boundaries in the stream of speech [10]. The use of text support has been found to be key for aural word recognition [9,21], especially when it comes to the learning of a language with opaque orthography [22], and when learners are mostly exposed to written input.

The few experimental studies that have explored the development of L2 listening skills through captioned videos have indicated that the provision of text support improves learner-viewers' speech perception after relatively short interventions (1–2 episodes; 25–60 min). Still, it is uncertain whether the results from studies with university students may also be replicated with primary school students. In fact, the investigation by Tragant et al. [23] with fifth graders from Spain revealed that learners' exposure to 21 graded readers (i.e., texts adapted for foreign language learners) with audio support failed to enhance the development of L2 listening and reading skills, a finding that was attributed to the insufficient length of the treatment.

### *1.2. Captions and L2 Reading Skills Development*

L2 reading is a highly complex task that integrates lower- and higher-level reading processes (e.g., word recognition and global comprehension, respectively). The complexity of L2 reading might potentially explain why this activity has not proved very popular in foreign language contexts (e.g., [24,25]). This is a great limitation considering that practice in reading is crucial to show significant improvement in the development of this receptive language skill [16,26]. In L1 contexts, a handful of investigations with primary school learners have demonstrated that the use of L1 captions enhances the development of reading skills (e.g., [13–15]). Nonetheless, these findings from L1 contexts may not necessarily be translated to foreign language contexts. L1 and L2 readers differ in terms of language proficiency and their exposure to written texts, which are crucial factors for becoming familiar with L2 orthographic patterns and for automatizing lower-level reading skills [27]. In sum, the extent to which the use of onscreen text supports the development of L2 reading skills in foreign language settings is still unknown.

Up to now, the studies conducted with primary school learners in L1 contexts have demonstrated that the use of bimodal verbal input (i.e., audio and text) enhances the development of reading skills [13]. However, the processing patterns and the specific aspects that may benefit from this activity seem to depend on the extent to which lower-level reading skills are automatized. On the whole, the evidence has indicated that learners' exposure to captions improves word decoding skills [13–15]. However, at earlier stages of reading skills development (7–9 years old, 2nd grade), children seem to focus their attention on lower-level reading skills while processing captions, which is a factor that might hinder their capacity to focus on details or less relevant elements from the input [14]. As explained by Sadoski and Paivio [28], when the decoding process is effortful, learners devote greater attention to lower linguistic levels, leaving fewer cognitive resources available to process other elements, such as images and gestures, and to make referential connections. As Linebarger et al. [15] hypothesize, there seems to be a stage where the reading of captions is neither too challenging nor too easy to follow, so the use of text support may successfully aid comprehension and foster reading skills development.

Furthermore, the reading aspects that may benefit from the use of onscreen text may also depend on the length of the treatment. The results obtained by Linebarger et al. [15] with second and third graders in the US (native and second-language learners) indicated that the students who were exposed to captions improved in terms of non-word reading (English patterns) but not in terms of oral reading fluency. In this regard, six episodes may not have been enough to lead to significant improvement. In other words, primary school learners might need intensive exposure to captions to significantly benefit from their viewing experience.

While it is true that the ultimate goal of reading instruction is to achieve high levels of comprehension, the literature suggests that the instruction and development of lower-level reading skills are crucial to attaining this objective [28]. Learners' development of both higher- and lower-level reading processes is key to building coherent mental representations [16,29,30]. With this in mind, it could also be stated that although captions support the development of lower-level reading skills at the expense of students' immediate viewing comprehension at early reading stages [14], the use of on-screen text might still be seen as a contribution to the learning process. That is to say, despite learners' greater efforts to cope with the speed of captions at early learning stages, the beneficial effects of captions on the automatization of orthographic and phonological processing may eventually result in higher levels of comprehension and motivation to read [22]. That being the case, captions may have the potential to break the vicious circle of low-achievers' reluctance to read [26] and, to a certain extent, counteract learners' lack of exposure to L2 print due to the complexity of L2 reading.

### *1.3. The Influence of Language-Related Factors*

#### 1.3.1. L1- vs. L2-Related Factors

The extent to which L2 learners process audiovisual input with ease, and benefit from their viewing experience, seems to be strongly predicted by L2 proficiency level [31]. At lower proficiency levels, the processing of onscreen text seems to be more effortful, increasing the amount of time viewers spend on captions/subtitles [32,33]. This finding is to be expected considering the complexity of L2 reading and the fact that learners' ability to read texts with ease and a high level of comprehension appears to be mainly explained by L2-related factors [34,35]. Although the literature has consistently shown evidence of the influence of L1 reading skills on L2 reading [23,26,36,37], it has also been suggested that L2-related factors might be stronger predictors of L2 reading (e.g., [34,35,38,39]). The strong relationship between L2 reading and L2-related factors has been explained in terms of the Simple View of Reading model [40,41] (see Figure 1), which postulates that reading comprehension is mainly explained by word decoding and general oral language comprehension [35].

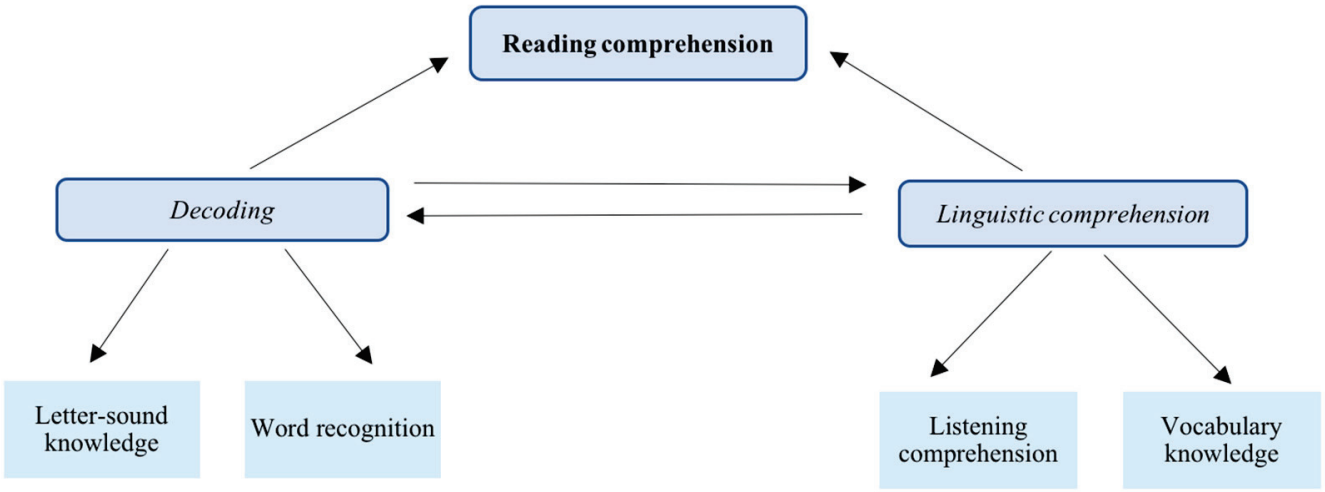



**Figure 1.** The modified Simple View of Reading.

#### 1.3.2. L1 and L2 Reading Skills

To date, few researchers have focused their attention on the direct relationship between L2 receptive language skills (reading and listening) and L2 learning from captioned videos [42,43], and no one, to the best of our knowledge, has studied the specific influence of L1 reading skills in this regard. L1 and L2 reading skills may be particularly relevant in the case of young learners as their reading skills are still developing [17], also due to the effort required to cope with the speed of captions [32], as well as the need to integrate verbal and pictorial information to attain appropriate levels of comprehension [28,44]. Concerning the influence of L1 reading skills, the literature suggests that learners' L1 orthography may support and facilitate L2 reading to compensate for L2 knowledge gaps and lack of practice, as long as there is an overlap between the two systems [45]. Therefore, learners are thought to progressively assimilate and accommodate their linguistic infrastructure to the characteristics of the L2 [26,37,46], which is a process that relies on their L2 proficiency and familiarity with the characteristics of the target language [46]. Hence, one may expect that at least in the case of young L2 learners, both L1 and L2 reading skills might play a role in the processing of captions and in their ability to learn from them.

#### 1.3.3. L2 Listening Skills

As for the role of L2 listening skills, the majority of studies on audiovisual input have focused on viewing comprehension or the development of listening skills rather than on the role of L2 listening in language learning from viewing, with few exceptions (e.g., [43,47]). Overall, it is widely accepted that L2 viewing without text support may be quite challenging for lower proficiency learners [48,49]. Therefore, one may assume that L2 proficiency, including L2 listening skills, may play a significant role in L2 learning from viewing without captions. Nonetheless, the scant evidence available indicates that listening skills may also predict the outcomes under the presence of captions, which is a factor that may not only be attributed to the input received through the aural channel but also to the general comprehension processes involved while viewing.

#### 1.3.4. L2 Vocabulary Knowledge

The empirical evidence has demonstrated that a minimum level of vocabulary knowledge is required in order to show appropriate levels of comprehension in different modalities, namely aural and/or written (e.g., [50–52]). Once this threshold is surpassed, L2 learning is likely to occur due to the lower effort required in the processes of decoding and comprehension, which means that sufficient resources are available to notice unknown target language constructions [53,54]. While the lexical coverage that ensures appropriate levels of comprehension in viewing has been found to be less demanding than in reading-

only and listening-only conditions (80%, [50]) due to the presence of imagery, the majority of studies on audiovisual input have identified L2 vocabulary knowledge as a significant predictor of L2 learning from viewing in different age groups (e.g., [55,56]).

## 2. Materials and Methods

This investigation attempted to fill some of the gaps in the literature by determining whether the use of age-appropriate captioned videos was conducive to the development of L2 receptive skills in a group of primary school learners of English as a foreign language. In addition, this study investigated the role played by a series of language-related factors on the outcomes (see Table 1).

**Table 1.** Research design.

| 1. Pretest: | | 2. Treatment | 3. Immediate Posttest: | 4. Delayed Posttest: |
|---|---|---|---|---|
| **Target Factors** | **Language-Related Factors** | | **Target Factors** | **Target Factors** |
| L2 reading efficacy | (L2 vocabulary knowledge, L1 reading efficacy, L2 listening skills) | 11 episodes of the animated cartoon Charlie and Lola | L2 reading efficacy | L2 reading efficacy |
| L2 listening skills | (L2 vocabulary knowledge, L1 reading efficacy, L2 reading efficacy) | | L2 listening skills | L2 listening skills |

The research questions that guided this study were the following:

(1) To what extent does the use of captioned videos foster the development of L2 listening skills and L2 reading efficacy in fourth and fifth graders?
(2) To what extent do language-related factors influence the development of L2 listening skills and L2 reading efficacy? (i.e., L2 vocabulary knowledge, L2 listening skills, and L1 and L2 reading efficacy).

### 2.1. Participants

This study was conducted with a convenience sample of 96 L1-Spanish primary school learners in year 4 (aged 9–10; $n$ = 47, 3 groups) and year 5 (aged 10–11; $n$ = 49, 2 groups) from a private school in Chile. The outcomes of both year levels were compared due to the fast and marked changes experienced by children during middle childhood as regards their physical, socio-emotional, and cognitive development [57]. The number of female and male participants was comparable (girls = 47; boys = 49). The following requirements were considered for the inclusion of the participants in the statistical analyses: watching 100% of the episodes, doing at least a set of pre and posttests to assess their progress over time, and not being part of the group of students with special educational needs (who in fact received additional support to complete the activities). In addition, parents' consent was requested prior to the intervention, which was complemented by children's oral confirmation of their willingness to participate in the viewing experience.

Not all private and semi-private schools in Chile ensure high levels of proficiency through formal instruction, especially when schools are located outside the three metropolitan centers [58]. The participants from this school had received around six hours a week of formal L2 instruction since preschool. However, the amount of out-of-school contact with the target language reported in a questionnaire by the participants was either limited or non-existent, a factor that may explain their low level of proficiency (approximately pre-A1 according to the CEFR).

Two groups of students from the same school participated as control groups, which only completed a specific set of tests and attended their regular English lessons. Specifically, control group 1 (CG1; $n$ = 16 students, nine male and seven female, year 5) took the L2 listening tests (pretest and posttest; 5-week interval) a year before the intervention with the experimental groups (2020), while control group 2 (CG2; $n$ = 17 students, nine male and

eight female, year 4) took the L2 reading efficacy tests (pretest and posttest; 5-week interval) a year after the actual experiment (2022).

### 2.2. Treatment

The experimental groups watched 11 episodes of the animated cartoon Charlie and Lola [59], which was considered to be age- and content-appropriate. An important characteristic of this animated cartoon is that each episode lasts 10 min. Previous research suggested that after 10 min, young L2 learners may not be able to cope with the cognitive demands of the viewing task, in particular those who are less skilled readers [7]. Moreover, the characteristics of this animated cartoon seem to ensure appropriate levels of comprehension. Specifically, the dialogues contain sufficient visual support, and there is a clear connection between verbal input and the actions performed by the characters [60]. The linguistic analyses of the scripts (VocabProfiler on Lextutor; [61]) revealed that, on average, the episodes reached 91.3% vocabulary coverage at the K1 level. Based on the threshold proposed by previous research on vocabulary coverage with audiovisual input (80%; [50,51]), the analyses suggested that Charlie and Lola ensured appropriate levels of comprehension in low-proficiency learners. This was confirmed by a pilot group of fourth graders who reported that the episodes were suitable and motivating. Likewise, the results obtained in our pilot study are consistent with the outcomes obtained by Tragant and Pellicer-Sánchez with eye-tracking methodology [33]. By exploring fifth graders' eye movements while watching an episode of Charlie and Lola, the researchers concluded that despite the participants' higher level of attention to written input, they were fully capable of processing both text and images.

### 2.3. Instruments

Due to the pandemic, the instruments were carefully designed to be administered in pen-and-paper and online format (Google Forms). Although this investigation was conducted onsite with the experimental groups, the students who were put in quarantine had to complete some activities online through Microsoft Teams, which was the official platform used at the school. These online activities were organized either individually or in small groups (four students maximum). The participants were asked to keep their cameras on during the whole session.

#### 2.3.1. Reading Efficacy in English and Spanish

The concept of reading efficacy (see [36]) integrates learners' reading speed (words read per minute = WPM) and comprehension, because both lower- and higher-level reading processes are thought to be equally important for discriminating high- and low-achievers [27]. Silent reading was selected over reading aloud because the former may better resemble the processing of captions and may be a more reliable instrument to test comprehension [29].

Several procedures were followed to ensure that the three texts selected to test reading efficacy in Spanish and English were comparable. As for Spanish reading efficacy (SR efficacy hereafter), the fiction texts (A, B, and C) were adapted from a set of sample materials [62] developed to train Chilean fourth graders for the national standardized test on reading skills. The ATOS readability formula, which measures text complexity as a function of average sentence length, average word length, and word difficulty level, indicated that the texts were appropriate for the sample groups. In addition, the six multiple-choice comprehension questions focused either on textually explicit/literal (four questions) or textually implicit information (two questions). Each test item consisted of a correct answer, three distractors, and the 'I don't know' option to prevent learners from guessing. Due to the participants' low L2 proficiency level, the non-fiction texts (A, B, and C) selected for the English reading efficacy tests (ER efficacy hereafter) were adapted from Pre-A1 starters' sample papers [63]. The assessment of text readability (i.e., Flesch–Kincaid reading ease and grade level) indicated that all the texts were easy to read. More specifically, we confirmed that the texts used at each testing time were comparable and suitable for the target groups.

The five multiple-choice comprehension questions followed the same structure as the SR efficacy tests. Four questions focused on textually explicit information while only one tested textually implicit information.

The measures obtained from these instruments were reading speed (WPM = [nº of words in the text/number of seconds used to read the whole passage] × 60) and comprehension. Each comprehension question was assigned one point. Then, the raw comprehension score was used to calculate the percentage of comprehension (number of correct answers × 100/Nº of questions). Finally, the formula used to calculate reading efficacy was ([WPM × % comprehension]/100). In view of the fact that reading speed (WPM) varied among the participants, there was no maximum score for this test.

Prior to their administration, these tests were pilot tested with 14 primary school learners from Chile and the conflictive items were improved before the intervention. With the experimental and control groups, the reading efficacy tests were administered in small sub-groups (four students maximum) to more accurately track reading speed (by using a stopwatch per child) and to ensure that the instructions were appropriately followed. First, the students were asked to read the texts at their own pace for comprehension purposes. This instruction was repeated several times to fulfill the main aim of this instrument. In addition, the instructions highlighted that the text had to be read only once. Having listened to the instructions, the students were explicitly told to start reading. For practical reasons, the texts contained red circles, which signaled the beginning and the end of the reading process. Specifically, the learners had to raise their hands when they reached the end. While answering the questions, the students did not have access to the text again.

#### 2.3.2. L2 Listening Skills

Two sample Movers tests (paper A and paper B) [63] were implemented to measure learners' listening skills at the three testing times (i.e., listening for specific information). Paper A was previously pilot tested onsite with two groups of fifth graders from the same school (1.5 years before the actual experiment) to ensure its suitability. As a result, it was decided to distribute the test in two sessions.

The maximum score was 20 points. Correct answers were given one point, so each section was worth a total of five points. The test was administered onsite in pen-and-paper format. An online version was also developed in order to assess the control group, as well as the students that were in quarantine. Paper A was administered at pretest and delayed posttest, while paper B was given at posttest. The Cronbach alpha coefficient obtained for test A was 0.722 with the pilot group. As for the experimental groups, the Cronbach alpha values obtained for paper A were 0.655 at pretest, and 0.794 at delayed posttest, while the coefficient obtained for paper B at posttest was 0.682. Given that reliability analyses performed with a small number of items tend to lead to low values [64], and this test generated four main scores, the results suggest that these instruments' internal consistency is acceptable [65].

#### 2.3.3. L2 Vocabulary Knowledge

The EFL picture vocabulary test assessed general vocabulary knowledge at the level of meaning recognition. This test was adapted from Puimège and Peters' [65] version of the Picture Vocabulary Size Test (PVST) created by Anthony and Nation [66] to assess young learners' vocabulary knowledge (L1 and L2 English speakers). The original instrument was pilot tested, and the results suggested that this test was not suitable for our context. Then, we designed an instrument with 50 items by keeping the same format. The target words, selected from the A2 key for schools' vocabulary list developed by Cambridge English Assessment [67], contained an equal number of items from the K1 and K2 frequency bands based on the analysis performed on Lextutor [61].

As for the testing procedures, the target words were uttered in isolation and then in a non-defining sentence [65]. These stimuli were simultaneously presented in written and oral form. The audio had been previously recorded by an English native speaker. Out

of four pictures, the students had to select the one that represented the meaning of the target word (A, B, C, or D). Additionally, the students had the possibility of selecting the 'I don't know' option to prevent guessing. The students had only 10 s to select the correct alternative for each testing item. The questions were presented through a video to ensure that the testing procedures were the same in all the experimental groups. In the online format, students could hear the audio and see the pictures on their form, while in the pen-and-paper format, a projector and speakers were used to show the video, while the students had to record their responses on an answer sheet.

The instrument was first pilot tested on six groups of EFL learners from Chile ($N$ = 188; Third-sixth graders). The Cronbach alpha coefficients obtained with the pilot groups were satisfactory (0.908 for K1 words, 0.898 for K2 words, and 0.898 for the whole test). In the experimental groups, the Cronbach alpha coefficients obtained were 0.866 for K1 words, 0.814 for K2 words, and 0.913 for the whole test. In addition, at the end of the pilot testing sessions, the EFL teachers answered a questionnaire (see [68]), which confirmed that the instrument was appropriate for the context. The pictures that were considered confusing were replaced by clearer options.

### 2.4. Analyses

Data analyses were performed in SPSS v.25. First, we ran a series of ANOVAs and T-tests to ensure that the groups in each year level were comparable in terms of language-related factors. Then, Pearson's correlations were performed to study the relationships between variables (receptive L2 skills and language-related factors). The factors that were not normally distributed (Kolmogorov–Smirnov/Shapiro–Wilk = $p < 0.05$) were square root (SQRT) transformed to run parametric tests. To measure learners' progress over time and the influence of language-related factors, a series of GLMMs were performed (generalized linear mixed models; L2 listening skills: binary logistic regressions, ER efficacy: linear models). These analyses included Satterthwaite approximation and robust covariances, which are suggested for small sample groups and unbalanced data. Additionally, multiple linear regressions were run to assess the influence of language-related factors and to calculate the specific contribution of each variable on ER efficacy. Prior to the calculation of GLMMs and multiple linear regressions, we assessed collinearity between variables (tolerance > 0.3; VIF < 3.33). As regards GLMMs, model fit was estimated through AIC (Akaike information criterion). The GLMMs consisted of a compound-symmetry structure with student identification as subjects and time as a repeated measure.

## 3. Results

### 3.1. Preliminary Analyses

Independent-sample T-tests were calculated to compare fourth and fifth graders in language-related factors. The results demonstrated that at pretest, fifth graders scored significantly higher than fourth graders in vocabulary knowledge ($t(86) = 3.195$, $p = 0.006$, r = 0.32), L2 listening skills ($t(94) = 17.921$, $p < 0.001$, r = 0.87), and ER efficacy ($t(107) = 4.320$, $p < 0.001$, r = 0.38). However, the difference between the two year levels in SR efficacy only approached significance ($t(106) = 1.923$, $p = 0.057$, r = 0.18) (see Tables 2 and 3).

**Table 2.** L2 vocabulary knowledge and SR efficacy: descriptive statistics.

| | **Year Level** | | | |
|---|---|---|---|---|
| | **Year 4** | | **Year 5** | |
| | **M** | **(SD)** | **M** | **(SD)** |
| Vocabulary knowledge | 14.47 | (6.28) | 20.5 | (10.81) |
| SR efficacy | 79.15 | (40.17) | 94.85 | (46.33) |

**Table 3.** L2 listening skills and ER efficacy: descriptive statistics.

| | | Listening Pretest | | Listening Posttest | | Listening Delayed | |
|---|---|---|---|---|---|---|---|
| | | **Mean** | **(SD)** | **Mean** | **(SD)** | **Mean** | **(SD)** |
| Group | Year 4 | 8.46 | (3.35) | 11.19 | (3.46) | 11.15 | (3.78) |
| | Year 5 | 11.31 | (4.51) | 13.63 | (3.81) | 14.16 | (3.85) |
| | CG1-fifth | 11.56 | (3.12) | 10.06 | (3.09) | | |
| | | **ER Efficacy Pretest** | | **ER Efficacy Posttest** | | **ER Efficacy Delayed** | |
| | | **Mean** | **(SD)** | **Mean** | **(SD)** | **Mean** | **(SD)** |
| Group | Year 4 | 45.00 | (30.36) | 68.98 | (46.17) | 80.81 | (33.03) |
| | Year 5 | 72.72 | (34.82) | 109.59 | (52.17) | 124.29 | (48.46) |
| | CG2-fourth | 50.06 | (31.24) | 42.53 | (21.18) | | |

### *3.2. Development of L2 Listening Skills from Captioned-Video Viewing and the Influence of Language-Related Factors*

We ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects and time as a repeated measure to assess each year level's progress over time. The model was built with learners' scores at the three testing times by setting 20 (maximum score) as the denominator. The fixed effects included in the analyses were time, year level, and their interaction. The results yielded significant main effects for time ($F(2,190) = 63.966$, $p < 0.001$) and year level ($F(1,89) = 15.844$, $p < 0.001$), but a non-significant interaction between these two factors. As shown in Table 4, the experimental groups showed significant progress from pretest to posttest, and from pretest to delayed posttest, regardless of their year level. The results also showed that learners' scores did not significantly decrease from posttest to delayed posttest. The significant effects of year level were associated with the fact that the higher performance of fifth graders was kept over time. On the whole, the results indicated that the treatment was similarly beneficial for both year levels (see Figure 2).

**Table 4.** Listening skills: Time pairwise contrasts.

| Time Pairwise Contrasts | Contrast Estimate | Std. Error | t | df | Adj. Sig. | 95% CI | |
|---|---|---|---|---|---|---|---|
| | | | | | | **Lower** | **Upper** |
| Pretest–posttest | −0.129 | 0.014 | −9.224 | 241 | 0.000 | −0.162 | −0.095 |
| Pretest–delayed | −0.142 | 0.015 | −9.785 | 197 | 0.000 | −0.175 | −0.109 |
| Posttest–delayed | −0.014 | 0.015 | −0.894 | 148 | 0.373 | −0.044 | 0.016 |

The sequential Bonferroni-adjusted significance level is 0.05. Confidence interval bounds are approximate.

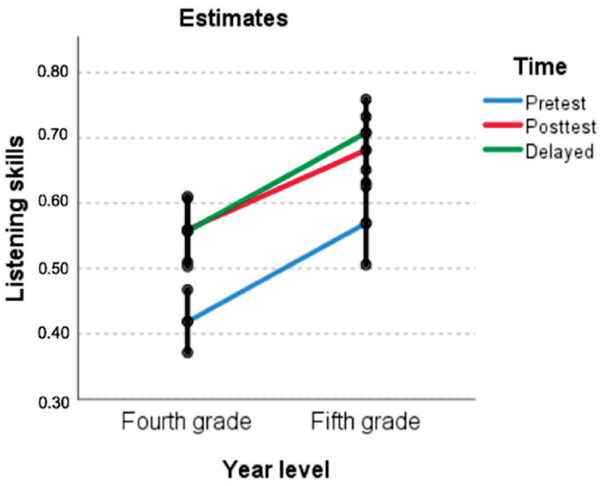



**Figure 2.** Listening skills: the trajectory of each year level over time.

To compare the experimental groups' performance with the control group (CG1-fifth) at pretest and posttest, we ran a compound symmetry structure GLMM (binary logistic regression). The results revealed significant effects for group ($F(2,135) = 6.768$, $p = 0.002$), time ($F(1,97) = 16.887$, $p < 0.001$), and their interaction ($F(2,102) = 14.495$, $p < 0.001$). Specifically, the Bonferroni-adjusted results indicated that the groups significantly improved from pretest to posttest ($p < 0.001$), except for the control group, which in fact scored lower at posttest (see Table 5).

**Table 5.** Listening skills: Time pairwise contrasts by class (pretest and posttest).

| Groups | Time Pairwise Contrasts | Contrast Estimate | Std. Error | t | df | Adj. Sig. | 95% CI | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Fourth grade | Pretest–posttest | −0.141 | 0.021 | −6.853 | 95 | 0.000 | −0.182 | −0.100 |
| Fifth grade | Pretest–posttest | −0.115 | 0.019 | −6.155 | 143 | 0.000 | −0.152 | −0.078 |
| CG1-fifth | Pretest–posttest | 0.075 | 0.036 | 2.105 | 89 | 0.038 | 0.004 | 0.146 |

The sequential Bonferroni-adjusted significance level is 0.05. Confidence interval bounds are approximate.

As for the influence of language-related factors on the development of L2 listening skills, a series of generalized linear mixed models (binary logistic regression) with repeated measures (time) compound-symmetry structure were calculated. To this end, the following factors were entered into the model: L2 vocabulary knowledge, SR efficacy, ER efficacy, year level, time, and all possible two-way interactions. The best-fitted model was determined by a backward elimination procedure. The results yielded significant effects for vocabulary knowledge ($F(1,82) = 56.549$, $p < 0.001$), SR efficacy ($F(1,119) = 6.754$, $p = 0.011$), and time ($F(2,173) = 59.388$, $p < 0.001$). The exponential coefficients in Table 6 indicate that the odds of obtaining a correct response in the listening test increased by 69% per each additional point on the EFL vocabulary test, and by 6.8% per each additional point in SR efficacy.

**Table 6.** Listening skills: best fitted model obtained to assess the influence of language-related factors on learners' scores.

| Model Term | Coef | Std. Error | t | Sig. | 95% CI | | Exp (Coef) | 95% CI for Exp(Coef) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Lower | Upper | | Lower | Upper |
| Intercept | −2.166 | 0.2748 | −7.882 | 0.000 | −2.711 | −1.621 | 0.115 | 0.066 | 0.198 |
| Vocabulary knowledge | 0.528 | 0.0702 | 7.520 | 0.000 | 0.388 | 0.667 | 1.695 | 1.474 | 1.949 |
| SR efficacy | 0.066 | 0.0252 | 2.599 | 0.011 | 0.016 | 0.115 | 1.068 | 1.016 | 1.122 |
| Pretest | −0.604 | 0.0669 | −9.029 | 0.000 | −0.736 | −0.472 | 0.547 | 0.479 | 0.624 |
| Posttest | −0.018 | 0.0730 | −0.252 | 0.802 | −0.163 | 0.126 | 0.982 | 0.850 | 1.134 |
| Delayed posttest | 0 [b] | | | | | | | | |

Probability distribution: binomial. Link function: logit. [b] This coefficient is set to zero because it is redundant.

### *3.3. Development of ER Efficacy from Captioned-Video Viewing and the Influence of Language-Related Factors*

To compare the performance of fourth and fifth graders over time, we calculated a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as a repeated measure. The model was fitted with learners' scores at the three testing times and the following fixed factors: time, year level, and their interaction. The results yielded significant main effects for year level ($F(1,94) = 27.711$, $p < 0.001$) and time ($F(2,166) = 72.697$, $p < 0.001$), but not for their interaction. However, the interaction was kept in the model to further observe each year level's outcomes (see Figure 3 and Table 7).

On the whole, the experimental groups showed significant gains from pretest to posttest, and from pretest to delayed posttest, regardless of their year level. Moreover, the results indicated that learners' scores significantly increased from posttest to delayed posttest. The significant effects of year level confirmed that fifth graders consistently outperformed fourth graders over time. As shown in Table 7, fifth graders seemed to obtain slightly higher gains from the treatment.

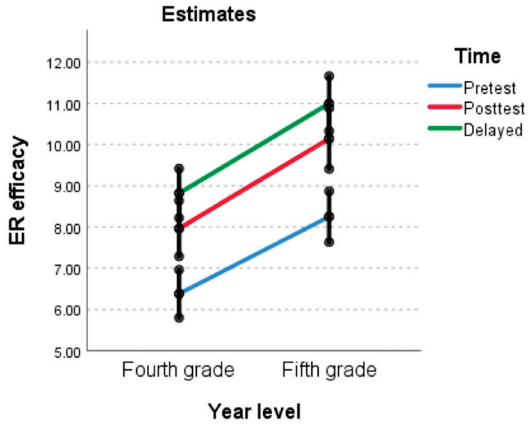



**Figure 3.** ER efficacy: the trajectory of each year level over time.

**Table 7.** ER efficacy: time pairwise contrasts by year level.

| Year Level | Time Pairwise Contrasts | Contrast Estimate | Std. Error | t | df | Adj. Sig. | 95% CI | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Fourth grade | Pretest–posttest | −1.584 | 0.256 | −6.199 | 261 | 0.000 | −2.160 | −1.008 |
| | Pretest–delayed | −2.440 | 0.300 | −8.145 | 185 | 0.000 | −3.164 | −1.716 |
| | Posttest–delayed | −0.856 | 0.302 | −2.831 | 178 | 0.005 | −1.453 | −0.259 |
| Fifth grade | Pretest–posttest | −1.893 | 0.275 | −6.884 | 201 | 0.000 | −2.514 | −1.272 |
| | Pretest–delayed | −2.744 | 0.349 | −7.855 | 96 | 0.000 | −3.596 | −1.893 |
| | Posttest–delayed | −0.851 | 0.304 | −2.804 | 146 | 0.006 | −1.451 | −0.251 |

The sequential Bonferroni-adjusted significance level is 0.05. Confidence interval bounds are approximate.

To compare the performance of the two year levels and the control group, a new compound symmetry structure GLMM (linear model) was calculated with student identification as subjects, and time as a repeated measure. To this end, we fitted learners' listening scores as the target variable, and entered group, time, and their interaction into the model as fixed factors. The results yielded significant effects for time ($F(1,120) = 29.693$, $p < 0.001$), group ($F(2,130) = 16.853$, $p < 0.001$), and a significant interaction between time and group ($F(2,116) = 11.889$, $p < 0.001$). Specifically, the Bonferroni-adjusted results revealed that both year levels significantly improved from pretest to posttest ($p < 0.001$), while the control group obtained similar scores at the two testing times (see Table 8).

**Table 8.** ER efficacy: time pairwise contrasts by group (pretest and posttest).

| Groups | Time Pairwise Contrasts | Contrast Estimate | Std. Error | t | df | Adj. Sig. | 95% CI | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Fourth grade | Pretest–posttest | −1.584 | 0.256 | −6.199 | 122 | 0.000 | −2.090 | −1.078 |
| Fifth grade | Pretest–posttest | −1.892 | 0.275 | −6.877 | 95 | 0.000 | −2.438 | −1.346 |
| CG2-fourth | Pretest–posttest | 0.443 | 0.411 | 1.079 | 134 | 0.282 | −0.369 | 1.256 |

The sequential Bonferroni-adjusted significance level is 0.05. Confidence interval bounds are approximate.

Additionally, a new model was built to assess each year level's progress in silent reading speed (number of words read per minute). Specifically, we ran a series of repeated measures (time) compound-symmetry structure GLMMs (linear model) with learners' scores in silent reading speed as the target factor, and the following variables as fixed factors: year level, time, and their interaction. The results revealed significant effects for year level ($F(1,89) = 21.346$, $p < 0.001$) and time ($F(2,175) = 20.950$, $p < 0.001$), while their interaction did not reach significance ($F(2,175) = 0.914$, $p = 0.403$). Nonetheless, this interaction was kept in the model to further explore the progress of each year level. As shown in Table 9, both year levels significantly improved from pretest to posttest, and from pretest to delayed posttest ($p < 0.05$). However, the results suggest that fifth graders

showed greater progress in reading speed over time, especially from pretest to delayed posttest, and from posttest to delayed posttest.

**Table 9.** ER efficacy: pairwise contrasts of the outcomes obtained by each group (WPM) over time.

| Year Level | Pairwise Contrasts | Contrast Estimate | Std. Error | t | df | Adj. Sig. | 95% CI | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Fourth grade | Pretest–posttest | −0.561 | 0.217 | −2.588 | 138 | 0.021 | −1.053 | −0.070 |
| | Pretest–delayed | −0.718 | 0.211 | −3.398 | 191 | 0.002 | −1.229 | −0.208 |
| | Posttest–delayed | −0.157 | 0.223 | −0.702 | 149 | 0.484 | −0.598 | 0.284 |
| Fifth grade | Pretest–posttest | −0.743 | 0.192 | −3.872 | 237 | 0.000 | −1.175 | −0.310 |
| | Pretest–delayed | −1.119 | 0.208 | −5.376 | 188 | 0.000 | −1.622 | −0.616 |
| | Posttest–delayed | −0.377 | 0.209 | −1.804 | 187 | 0.073 | −0.788 | 0.035 |

The sequential Bonferroni-adjusted significance level is 0.05. Confidence interval bounds are approximate.

As regards the influence of language-related factors, we calculated a series of repeated-measures (time) compound-symmetry structure GLMMs (linear models). To this aim, we entered ER efficacy (students' scores at the three testing times) as the target factor, and the following variables as fixed effects: year level, time, L2 listening skills, SR efficacy, L2-vocabulary knowledge, and all possible two-way interactions. By following a step-back procedure, the non-significant interactions and main effects were removed from the model one by one until the best-fitted model was obtained (see Table 10). The results revealed significant main effects for year level ($F(1,53) = 9.186$, $p = 0.004$), listening skills ($F(1,66) = 4.525$, $p = 0.037$), L2 vocabulary knowledge ($F(1,80) = 29.818$, $p < 0.001$), SR efficacy ($F(1,104) = 11.395$, $p = 0.001$), and time ($F(2,121) = 4.302$, $p = 0.016$). In addition, the results yielded a significant interaction between L2 listening skills and time ($F(2,154) = 4.351$, $p = 0.015$).

**Table 10.** ER efficacy: The influence of L2-related factors on learners' scores.

| Model Term | Coef | Std. Error | t | Sig. | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 3.439 | 0.8089 | 4.251 | 0.000 | 1.828 | 5.049 |
| Year level | −0.934 | 0.3082 | −3.031 | 0.004 | −1.552 | −0.316 |
| Listening skills | 0.116 | 0.0561 | 2.068 | 0.041 | 0.005 | 0.227 |
| Vocabulary knowledge | 0.970 | 0.1777 | 5.461 | 0.000 | 0.617 | 1.324 |
| SR efficacy | 0.212 | 0.0628 | 3.376 | 0.001 | 0.087 | 0.337 |
| Pretest | −1.694 | 0.6257 | −2.707 | 0.008 | −2.936 | −0.452 |
| Posttest | −1.464 | 0.5762 | −2.541 | 0.012 | −2.605 | −0.323 |
| Delayed | 0 [b] | | | | | |
| Listening skills *[pretest] | −0.093 | 0.0538 | −1.728 | 0.086 | −0.199 | 0.013 |
| Listening skills *[posttest] | 0.055 | 0.0487 | 1.125 | 0.262 | −0.041 | 0.151 |
| Listening skills *[delayed] | 0 [b] | | | | | |

Probability distribution: normal. Link function: identity. [b] This coefficient is set to zero because it is redundant. * Interaction.

To calculate the contribution of each factor on students' ER efficacy scores at pretest, posttest and delayed posttest, we ran multiple linear regressions for each testing time. The predictor variables included were as follows: listening skills, vocabulary knowledge, and SR efficacy. At pretest, L2 listening skills were not found to contribute significantly ($p > 0.05$), so this factor was removed from the analysis. The results indicated that vocabulary knowledge and SR efficacy predicted 44% of the variance ($F(2,85) = 36.404$, $p < 0.001$, $R^2 = 0.449$). Specifically, the standard coefficients indicated that vocabulary knowledge was the strongest predictor ($\beta = 47\%$, $p < 0.001$), followed by SR efficacy ($\beta = 31\%$, $p = 0.001$). At posttest, the results indicated that listening skills, vocabulary knowledge, and SR efficacy predicted 57% of the variance ($F(3,83) = 39.437$, $p < 0.001$, $R^2 = 0.573$). The standard coefficients indicated that vocabulary knowledge was the strongest predictor ($\beta = 44\%$, $p < 0.001$),

followed by listening skills ($\beta$ = 27%, $p$ = 0.007) and SR efficacy ($\beta$ = 17%, $p$ = 0.042). Finally, at delayed posttest, the results revealed that SR efficacy was no longer significant ($p$ = 0.100), so this factor was removed from the analysis. The results showed that vocabulary knowledge and listening skills predicted 48% of the variance ($F(2,76) = 38.210$, $p < 0.001$, $R^2 = 0.488$). Again, the standard coefficients indicated that vocabulary knowledge was a stronger predictor ($\beta$ = 45%, $p < 0.001$) than L2 listening skills ($\beta$ = 32%, $p$ = 0.004).

## 4. Discussion

This study aimed to determine the extent to which primary school learners may benefit from captioned video viewing with regard to the development of L2 listening skills and ER efficacy. In addition, we studied the influence of language-related factors on the outcomes. With respect to the first research question, the results indicated that in comparison with the control groups, the experimental groups obtained significant gains from the treatment in terms of L2 listening skills and ER efficacy. As for L2 listening skills, the results seemed to confirm that the use of captions enhanced bottom-up processing after a relatively short intervention (11 episodes), as previous studies have also found [9–11]. Likewise, the results concerning the development of ER efficacy suggested that the findings obtained by studies conducted in L1 settings (e.g., [13–15]) may be extrapolated to foreign language contexts. Additionally, the comparisons between year levels regarding their progress in silent reading speed indicated that both year levels significantly improved over time, although fifth graders were found to obtain greater gains in this regard. This outcome seems to indicate that learners' progress in ER efficacy may not only be associated with their improvement in silent reading speed but also with their capacity to devote fewer attentional resources to text decoding to improve their levels of comprehension [28,30]. As the literature suggests, the automatization of lower-level reading skills requires plenty of exposure to print [16]. Thus, fourth graders might need a higher number of episodes to show greater gains in silent reading speed [15].

The fact that the use of captioned videos enhanced the development of both receptive language skills may not be surprising, given that reading and listening have been found to have a bidirectional relationship [26,35,38,41], which may have been enhanced by learners' simultaneous exposure to both modalities [38]. As for the second research question, the results indicated that learners' outcomes were influenced by both L1 and L2-related factors. Specifically, their progress in L2 listening skills was found to be significantly predicted by L2 vocabulary knowledge and SR efficacy, while their performance in ER efficacy over time was significantly explained by vocabulary knowledge, L2 listening skills, and SR efficacy. However, when assessing the exact contribution of each variable, the findings suggested that learners' performance in L2 listening and reading skills over time was mainly explained by L2-related factors, particularly L2 vocabulary knowledge [34,35]. The finding that indicated that reading efficacy in Spanish but not in English influenced the development of L2 listening skills may be explained by the target participants' limited practice in L2 reading and viewing. Thus, in line with the literature, the outcomes suggest that the participants relied on their L1 linguistic infrastructure to process onscreen text [26,46]. In other words, young and low-proficiency learners may rely on L1 reading skills to compensate, to a certain extent, for their L2 knowledge gaps and lack of practice [39,45].

In general, the treatment appeared to be similarly beneficial for both year levels. Although it is true that fifth graders consistently outperformed fourth graders at the three testing times in L2 listening skills, the results may be attributed to their significantly higher proficiency level, and possibly to their higher cognitive development [12,17]. A similar picture was observed when assessing learners' performance in ER efficacy over time. However, as mentioned above, fifth graders appeared to show marginally higher improvement in ER efficacy, especially as regards silent reading speed. This outcome may be associated with their stronger L1 literacy skills [17], which may have allowed them to cope with the speed of captions and encouraged them to stay on the reading task

while viewing. Fifth graders may have been better equipped to rely on their L1 linguistic infrastructure to cope with the L2 input demands [26,37].

As for learners' trajectory from posttest to delayed posttest, the results indicated that their scores in L2 listening skills did not show significant variability. Yet, fifth graders showed a slight improvement between these two testing times. Similarly, fifth graders' scores in ER efficacy presented a marginally significant increase. While these outcomes may well be attributed to test effects, it may also be the case that the intervention encouraged fifth graders to watch videos at home. Prior research with adolescents has demonstrated that this type of intervention may have a positive impact on learners' viewing habits (e.g., [69]). In fact, the data collected in the present investigation by means of an interview with sample participants suggested that fifth graders were more open to experimenting with viewing and the use of onscreen text as a result of the intervention (see [68]). Nonetheless, more evidence needs to be gathered to confirm these outcomes. Future studies should explore further the extent to which young learners' viewing experiences at school may have an effect on their viewing habits at home [8].

The findings of the present study do not coincide with those of the study conducted by Tragant et al. [23] with fifth graders. In that study, learners' exposure to 21 graded readers (with and without audio support) was not conducive to significant gains in listening or reading skills in comparison with a control group. As Tragant et al. [23] explained, the length of the intervention may have been insufficient to observe significant gains from learners' exposure to graded readers, because the development of receptive language skills requires a great deal of practice [16], particularly in the case of young L2 learners, who have been found to be less efficient [18]. A key difference between the two studies was that our participants' viewing experience was limited or practically non-existent, while the students in Tragant et al.'s [23] investigation were already familiar with graded readers in L1 and L2. Thus, the gains in our study might be the result of the sudden increase in the participants' exposure to captioned videos (11 viewing sessions). Furthermore, the conflicting results may also be associated with the presence of moving images and the fact that the participants in the present study were not able to control the viewing process as had been the case in Tragant et al.'s [23] investigation, where each child could manipulate the audio (e.g., pause) and read the books at their own pace. Thus, learners' gains in L2 listening skills and ER efficacy may have been enhanced by the supporting role of imagery in terms of comprehension and the greater effort involved in the processing of captions during screen exposure. Further work is required to confirm whether the implementation of relatively short interventions that include learners' simultaneous exposure to audio and text may foster the development of receptive L2 skills. In summary, the comparison between reading-while-listening and captioned videos might also be a fruitful area for further research. By the same token, the use of eye-tracking methodologies would be of great help to explore young learners' processing patterns while reading static and dynamic texts with audio support, as well as to compare different age or proficiency groups [33].

Perhaps the main limitation of our investigation was that our data were collected during the COVID-19 pandemic, because our decisions were restricted by the contingency measures taken by the Chilean government. In 2021, few schools implemented in-person classes, and most of them were reluctant to welcome researchers into their classrooms. Therefore, we were unable to implement this research design at a different school to increase our sample size and/or compare learners' outcomes as a function of the characteristics of their language program. In addition, the present study measured gains after a relatively short intervention, because we worked under the threat of going online at any moment. Hence, we prioritized the completion of the experiment over the measurement of learners' outcomes after a higher number of sessions.

## 5. Conclusions and Pedagogical Implications

Taken together, the evidence obtained in this investigation supports the use of captioned videos with primary-school learners to foster the development of receptive L2 skills.

The results seem to demonstrate that fourth graders' lower proficiency level (in L1 and L2) and lack of exposure to the target language did not prevent them from benefiting from the treatment. Thus, the results lead to the conclusion that the use of captioned videos in the L2 classroom may help learners cope with the challenges entailed in the implementation of listening-only and reading-only activities [19,20] and may break down the potential vicious circle of low achievers' reluctance to complete reading and listening activities [26]. As stated above, young learners' exposure to the target language is a fundamental aspect of their language learning process [17,18]; therefore, the systematic use of captioned videos in the L2 classroom might enhance the outcomes of foreign language learning programs implemented in input-limited contexts.

**Author Contributions:** Conceptualization, D.A. and C.M.; methodology, D.A. and C.M.; analyses, D.A.; investigation, D.A. and C.M.; resources, D.A. and C.M.; data curation, D.A.; writing—original draft preparation, D.A.; writing—review and editing, D.A. and C.M.; supervision, C.M.; project administration, C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the APIF scholarship granted by the University of Barcelona, and the grant PID2019-110594GB-100 from the Spanish Ministry of Science, Innovation and Universities.

**Institutional Review Board Statement:** The study was conducted in accordance with the ethical principles proposed by the Research Ethics Committee at the University of Barcelona (2020-22). The study protocol guaranteed good practices in data collection, anonymization, processing, and storage. The study respected the privacy and confidentiality of all the participants involved. Apart from ensuring that parents' informed consent was provided, we requested students' verbal confirmation and respected their actual willingness to complete the activities implemented at school.

**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 available from the corresponding author, upon reasonable request.

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

## References

1. Montero Perez, M. Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Lang. Teach.* **2022**, *55*, 163–192. [CrossRef]
2. Muñoz, C. Audiovisual input in L2 learning. *Lang. Interact. Acquis.* **2022**, *13*, 125–143. [CrossRef]
3. Montero Perez, M.; Rodgers, M. Video and language learning. *Lang. Learn. J.* **2019**, *47*, 403–406. [CrossRef]
4. Black, S. Subtitles as a tool to boost language learning? Children's views and experiences of watching films and television programmes in other languages with interlingual subtitles. *J. Audiov. Transl.* **2022**, *5*, 73–93.
5. Lindgren, E.; Muñoz, C. The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *Int. J. Multiling.* **2013**, *10*, 105–129. [CrossRef]
6. Muñoz, C.; Cadierno, T.; Casas, I. Different starting points for English language learning: A Comparative Study of Danish and Spanish Young Learners. *Lang. Learn.* **2018**, *68*, 1076–1109. [CrossRef]
7. Marzá, A.; Torralba, G. Incidental language learning through subtitled cartoons: Is it possible in a dubbing country? In *Subtitles and Language Learning. Principles, Strategies and Practical Experiences*; Gambier, Y., Caimi, A., Mariotti, C., Eds.; Peter Lang: Bern, Switzerland, 2015; pp. 199–219.
8. Webb, S. Extensive viewing: Language learning through watching television. In *Language Learning beyond the Classroom*; Nunan, D., Richards, J., Eds.; Routledge: New York, NY, USA, 2015; pp. 159–168.
9. Birulés-Muntané, J.; Soto-Faraco, S. Watching subtitled films can help learning foreign languages. *PLoS ONE* **2016**, *11*, e0158409. [CrossRef]
10. Charles, T.; Trenkic, D. Speech segmentation in a second language: The role of bi-modal input. In *Subtitles and Language Learning. Principles, Strategies and Practical Experiences*; Gambier, Y., Caimi, A., Mariotti, C., Eds.; Peter Lang: Bern, Switzerland, 2015; pp. 173–197.
11. Mitterer, H.; McQueen, J.M. Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* **2009**, *4*, e7785. [CrossRef]
12. Muñoz, C. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Appl. Linguist.* **2008**, *24*, 578–596. [CrossRef]
13. Kothari, B.; Takeda, J.; Joshi, A.; Pandey, A. Same language subtitling: A butterfly for literacy? *Int. J. Lifelong Educ.* **2002**, *21*, 55–66. [CrossRef]

14. Linebarger, D. Learning to read from television: The effects of using captions and narration. *J. Educ. Psychol.* **2001**, *93*, 288–298. [CrossRef]
15. Linebarger, D.; Taylor Piotrowski, J.; Greenwood, C. On-screen print: The role of captions as a supplemental literacy tool. *J. Res. Read.* **2010**, *33*, 148–167. [CrossRef]
16. Grabe, W.; Stoller, F. *Teaching and Researching Reading*, 3rd ed.; Routledge: New York, NY, USA, 2020.
17. Holmes, B.; Myles, F. White Paper: Primary Languages Policy in England. The Way Forward. RiPL. 2019. Available online: www.ripl.uk/policy/ (accessed on 15 May 2022).
18. Muñoz, C. *Age and the Rate of Foreign Language Learning*; Multilingual Matters: Clevedon, UK, 2006.
19. Kormos, J.; Kosak, M.; Pizorn, K. The role of low-level first language skills in second language reading, reading-while-listening and listening performance: A study of young dyslexic and non-dyslexic language learners. *Appl. Linguist.* **2019**, *40*, 834–858. [CrossRef]
20. Newton, J.; Nation, I.S.P. Extensive listening. In *Teaching ESL/EFL Listening and Speaking*, 2nd ed.; Newton, J., Nation, I.S.P., Eds.; Routledge: New York, NY, USA, 2021.
21. Bird, S.A.; Williams, J.N. The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Appl. Psycholinguist.* **2002**, *23*, 509–533. [CrossRef]
22. Toscano-Fuentes, C.M.; Julián de Vega, C. Videos musicales en el aula de inglés de primaria para la mejora de la fluidez lectora. *Tejuelo* **2018**, *28*, 43–66.
23. Tragant, E.; Llanes, À.; Pinyana, À. Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English. *Read. Writ.* **2019**, *32*, 819–838. [CrossRef]
24. Muñoz, C. Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *Rev. Española De Lingüística Apl.* **2020**, *33*, 171–201.
25. Peters, E. The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *Int. J. Appl. Linguist.* **2018**, *169*, 142–168.
26. Birch, B.; Fulop, S. *English L2 Reading: Getting to the Bottom*, 4th ed.; Routledge: New York, NY, USA, 2021.
27. Grabe, W. *Reading in a Second Language. Moving from Theory to Practice*; Cambridge University Press: Cambridge, UK, 2009.
28. Sadoski, M.; Paivio, A. A dual theoretical model of reading. In *Theoretical Models and Processes of Reading*; Alvermann, D.E., Unrau, N.J., Ruddell, R.B., Eds.; International Reading Association: Newark, NJ, USA, 2013; pp. 886–922.
29. Alderson, J.C.; Haapakangas, E.-L.; Huhta, A.; Nieminen, L.; Ullakonoja, R. *The Diagnosis of Reading in a Second or Foreign Language*; Routledge: New York, NY, USA, 2015.
30. Nassaji, H. The role and importance of lower-level processes in second language reading. *Lang. Teach.* **2015**, *47*, 1–37. [CrossRef]
31. Gesa, F.; Miralpeix, I. Effects of watching subtitled TV series on foreign language learning. Does proficiency level matter? In *Foreign Language Learning in the Digital Age. Theory and Pedagogy for Developing Literacies*; Lütge, C., Ed.; Routledge: Oxford, UK, 2022; pp. 159–173.
32. Muñoz, C. The role of age and proficiency in subtitle reading: An eye tracking study. *System* **2017**, *67*, 77–86. [CrossRef]
33. Tragant, E.; Pellicer-Sánchez, A. Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System* **2019**, *80*, 212–223. [CrossRef]
34. Alderson, J.C.; Huhta, A.; Nieminen, L. Characteristics of weak and strong readers in a foreign language. *Mod. Lang. J.* **2016**, *100*, 853–879. [CrossRef]
35. Sparks, R.L. Identification and characteristics of strong, average, and weak foreign language readers: The simple view of reading model. *Mod. Lang. J.* **2021**, *105*, 507–525. [CrossRef]
36. Llanes, A. Reading in English as a foreign language: Examining differences in reading speed, comprehension, efficacy and L1 cross-linguistic influence across grades. *Investig. Sobre Lect.* **2018**, *9*, 1–49.
37. Perfetti, C.A.; Liu, Y.; Fiez, J.; Nelson, J.; Bolger, D.J.; Tan, L. Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Biling. Lang. Cogn.* **2007**, *10*, 131–146. [CrossRef]
38. Verhoeven, L.; van Leeuwe, J. The simple view of second language reading throughout the primary grades. *Read. Writ.* **2012**, *25*, 1805–1818. [CrossRef]
39. Yamashita, J. Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *J. Res. Read.* **2002**, *25*, 81–95. [CrossRef]
40. Gough, P.; Tunmer, W. Decoding, reading, and reading disability. *Remedial Spec. Educ.* **1986**, *7*, 6–10. [CrossRef]
41. Tunmer, W.E.; Chapman, J.W. The simple view of reading redux. *J. Learn. Disabil.* **2012**, *45*, 453–466. [CrossRef]
42. Muñoz, C.; Pattemore, A.; Avello, D. Exploring repeated captioning viewing as a way to promote vocabulary learning: Time lag between repetitions and learner factors. *Comput. Assist. Lang. Learn.* **2022**. [CrossRef]
43. Suárez, M.; Gesa, F. Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *Lang. Learn. J.* **2019**, *47*, 479–496. [CrossRef]
44. Pellicer-Sánchez, A.; Tragant, E.; Conklin, K.; Rodgers, M.; Serrano, R.; Llanes, A. Young learners' processing of multimodal input and its impact on reading comprehension. An eye tracking study. *Stud. Second Lang. Acquis.* **2020**, *42*, 577–598. [CrossRef]
45. Birch, B. *English L2 Reading: Getting to the Bottom*, 3rd ed.; Routledge: New York, NY, USA, 2015.
46. Jiang, J.; Ouyang, J.; Liu, H. Interlanguage: A perspective of quantitative linguistic typology. *Lang. Sci.* **2019**, *74*, 85–97. [CrossRef]

47. Pattemore, A.; Muñoz, C. Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System* **2020**, *93*, 102303. [CrossRef]
48. Pujadas, G.; Muñoz, C. When to Switch Off Captions? In Proceedings of the EuroSLA 31st Conference, Fribourg, Switzerland, 24–27 August 2022.
49. Vanderplank, R. 'Gist watching can only take you so far': Attitudes, strategies and changes in behaviour in watching films with captions. *Lang. Learn. J.* **2019**, *47*, 407–423. [CrossRef]
50. Durbahn, M.; Macís, M.; Rodgers, M.; Peters, E. Lexical coverage in L2 viewing comprehension: An extension of van Zeeland and Schmitt (2012). Presented at the EuroSLA 31st Conference, Fribourg, Switzerland, 24–27 August 2022.
51. Durbahn, M.; Rodgers, M.; Peters, E. The relationship between vocabulary and viewing comprehension. *System* **2020**, *88*, 102166. [CrossRef]
52. van Zeeland, H.; Schmitt, N. Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Appl. Linguist.* **2012**, *33*, 457–479. [CrossRef]
53. Kim, S.K.; Webb, S. Individual difference factors for L2 vocabulary. In *The Routledge Handbook of Second Language Acquisition and Individual Differences*; Li, S., Hiver, P., Papi, M., Eds.; Routledge: New York, NY, USA, 2022; pp. 282–293.
54. Lin, P.M.S.; Siyanova-Chanturia, A. Internet television for L2 vocabulary learning. In *Language Learning Beyond the Classroom*; Nunan, D., Richards, J., Eds.; Routledge: New York, NY, USA, 2015; pp. 149–158.
55. Alexiou, T. Vocabulary uptake from Peppa Pig: A case study of preschool EFL learners in Greece. In *Current Issues in Second/Foreign Language Teaching and Teacher Development: Research and Practice*; Gitsaki, C., Alexiou, T., Eds.; Cambridge Scholars Publishing: Newcastle, UK, 2015; pp. 285–301.
56. Pujadas, G.; Muñoz, C. Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *Lang. Learn. J.* **2019**, *47*, 479–496. [CrossRef]
57. Delgiudice, M. Middle childhood: An evolutionary-developmental synthesis. In *Handbook of Life Course Health Development*; Halfon, N., Forrest, C., Lerner, R., Faustman, E., Eds.; Springer: New York, NY, USA, 2018; pp. 95–107.
58. Enever, J. *Policy and Politics in Global Primary English*; Oxford University Press: Oxford, UK, 2018.
59. Carrington, M.; Child, L. (Executive Producers); *Charlie and Lola [Animated TV Series]*; Tiger Aspects Productions: London, UK, 2005–2008.
60. Donaghy, K. Using film to teach languages in a world of screens. In *Using Film and Media in the Language Classroom: Reflections on Research-Led Teaching*; Herrero, C., Vanderchelden, I., Eds.; Multilingual Matters: London, UK, 2019; pp. 3–16.
61. Lextutor. Available online: https://www.lextutor.ca/vp/eng/ (accessed on 30 October 2019).
62. Santillana. *Lenguaje y Comunicación Para Cuarto Básico Serie Bicentenario*; Santillana: Santiago, Chile, 2014.
63. Cambridge Assessment English. PreA1 Starters, A1 Movers and A2 Flyers Sample Papers. Available online: https://www.cambridgeenglish.org/exams-and-tests/movers/preparation/ (accessed on 15 October 2019).
64. Pallant, J. *SPSS Survival Manual. A Step by Step Guide to Data Analysis Using IBM SPSS*, 6th ed.; McGraw Hill Education: Berkshire, UK, 2016.
65. Puimège, E.; Peters, E. Learners' English vocabulary knowledge prior to formal instruction: The effect of learner-related and word-related factors. *Lang. Learn.* **2019**, *69*, 943–977. [CrossRef]
66. Anthony, L.; Nation, I.S.P. Picture Vocabulary Size Test (Version 1.2.0). Available online: http://www.laurenceanthony.net/software/pvst (accessed on 2 May 2019).
67. Lanes, A.; Love, R.; Kalman, B.; Brenchley, M.; Pickles, M. *Updating the A2 Key and B1 Preliminary Vocabulary Lists*; Cambridge Research English Research Notes; Cambridge University Press: Cambridge, UK, 2019; pp. 1–9.
68. Avello, D. L2 Learning from Captioned-Video Viewing in Primary School Students. Ph.D. Thesis, Universitat de Barcelona, Barcelona, Spain, February 2023.
69. Pujadas, G. Language Learning through Extensive TV Viewing. A Study with Adolescent EFL Learners. Ph.D. Thesis, Universitat de Barcelona, Barcelona, Spain, November 2019.

education sciences

MDPI

*Article*

# Extensive Reading and Science Vocabulary Learning in L2: Comparing Reading-Only and Reading-While-Listening

**Raquel Serrano**

Department of Modern Languages and Literatures and English Studies, University of Barcelona, 08007 Barcelona, Spain; raquelserrano@ub.edu

**Abstract:** This paper presents a study analyzing second language vocabulary gains after an extensive reading program that included non-fiction graded readers of scientific content in English. The study was conducted in a Spanish primary school ($N = 96$) and implemented in two different modalities: reading-only and reading-while-listening, which included audiobooks. The study lasted one school year and involved 39 science graded readers, making it unique in its duration and scope. The findings indicate that the practice of extensive reading resulted in notable improvements in vocabulary acquisition during the first half of the school year; however, the advantages were less evident in the second half. Different factors intrinsic to the program but also related to students' motivation will be discussed in order to explain the findings.

**Keywords:** extensive reading; graded readers; academic vocabulary learning

**Citation:** Serrano, R. Extensive Reading and Science Vocabulary Learning in L2: Comparing Reading-Only and Reading-While-Listening. *Educ. Sci.* **2023**, *13*, 493. https://doi.org/10.3390/educsci13050493

Academic Editors: Graham Crookes and Martin Howard

Received: 22 March 2023
Revised: 12 April 2023
Accepted: 10 May 2023
Published: 13 May 2023



## 1. Introduction

The current study presents an analysis of immediate and long-term vocabulary gains after an extensive reading (ER) program that employed non-fiction books of scientific content in English. The target books were graded readers, which were adapted to the learners' proficiency. The ER program was implemented in a Spanish primary school in two different modalities: reading-only and reading-while-listening. As Nation [1,2] suggests, ER should be an important component of second language (L2) learning programs. There is research that supports the benefits of ER for a variety of L2 areas, mostly, vocabulary learning, reading comprehension, and reading fluency [3]. Most studies on ER have focused on older learners, mainly college students, and programs that used fiction graded readers. In spite of this, it has been suggested that ER programs could also be beneficial for young learners [4] and that non-fiction books should also be included in order to cater to the diverse reading preferences of students [5]. Although including a variety of genres is usually recommended, a study focusing solely on non-fiction books related to a particular subject, for example natural sciences, would be a promising avenue for research. On the one hand, such a study could potentially add to the existing body of literature on ER and its impact on L2 acquisition. Additionally, it could provide valuable insights into how science graded readers could contribute to the acquisition of scientific vocabulary and enhance science learning. This type of investigation would be particularly informative for schools that follow the Content and Language Integrated Learning (CLIL) approach and offer science education in English.

There are several variables that have been demonstrated to impact the extent to which vocabulary acquisition can occur through reading. One such variable is the reading mode, and several studies have compared reading with audio support, also known as assisted reading or reading-while-listening (RWL), and reading only (RO). These studies indicate that RWL tends to be more advantageous [6,7], although in some instances, the differences between the two reading modes have not been found to be statistically significant [8,9]. Therefore, additional research is necessary to gain further insight into how the presence of audio affects the degree of vocabulary learning from reading.

The present study contributes to the literature on the effects of ER for L2 development, and offers new insights by comparing RO and RWL, including an under-researched population (i.e., young learners in primary school) reading graded readers of scientific content instead of fiction, as is common in the literature. Another novelty of the present study is that it involves a program that lasted one school year and that included 39 graded readers, in contrast to most studies that have examined shorter periods and fewer books.

In the following subsections, a review of the literature will be presented with the aim of contextualizing the current study and identifying the research gaps that motivated this investigation.

### *1.1. Extensive Reading*

Nation [1] asserts in his influential paper on "The Four Strands" that a well-structured L2 curriculum must maintain an equilibrium among four essential components: meaning-oriented input, meaning-oriented output, language-focused activities, and fluency practice. For the first strand, he advocates ER as an effective means to promote language learning. In a subsequent paper primarily addressed to language teachers, Nation [2] further suggests that "The single most effective change a teacher can make to a language course is to include an extensive reading program" (p. 6). Similarly, after performing a meta-analysis on the benefits of extensive reading, Nakanishi [4] (p. 6) concludes: "In sum, the available research to date suggests that extensive reading improves students' reading proficiency and should be a part of language learning curricula".

At the heart of an ER program is the objective of providing L2 learners with the opportunity to read extensively, typically at a rate of one book per week, silently and independently over an extended period [5]. The main objective is for students to read fluently and focus on understanding the meaning of the texts, which is why it is crucial that the books are at the right level for the students. These principles are what Waring and McLean [10] considered the "core elements" of ER, as it has been operationalized in the literature. However, according to the authors, there are other "variable elements", many of which were included in the Top Ten Principles of ER [5], that may or may not be present, such as reading exclusively for pleasure or allowing students to choose books.

Research suggests that ER is beneficial for different L2 aspects, including reading fluency [11–14], and reading comprehension [15,16]. There is also evidence that reading books in an L2 contributes to vocabulary learning. Nakanishi's meta-analysis on ER research [4] found that the effect of ER for vocabulary learning was larger ($d$ = 1.25) than for reading comprehension ($d$ = 0.72) or reading speed ($d$ = 0.61).

Numerous studies have demonstrated that L2 learners can expand their vocabulary through reading. Empirical support is derived, in part, from controlled investigations in which participants read one or a few books containing a subset of target words that were not readily available beyond the confines of the book(s). Waring and Takaki [17] examined the learning of 25 non-words that replaced real words in one graded reader (5872 words) in the case of 15 Japanese college students. The results of the immediate post-tests showed that on average, participants were able to recognize the form of 15.3 of the target words, the meaning of 10.6 and were able to translate 4.6. These gains decayed over time to 8.4 target words on the form-recognition test, 6.1 on the meaning-recognition test and 0.9 in the translation test.

Pellicer-Sánchez and Schmitt [18] also performed a controlled study to analyze how much vocabulary a group of 20 university EFL students from Spain gained after reading a single non-adapted novel in English (around 67,000 words), which included some African words, 34 of which were selected as target. The results of the study are interesting, not only because they show that vocabulary can be learned through reading but also because they demonstrate that some components of vocabulary knowledge are more easily learned by reading a book than others (see also [19]), with participants being able to recognize the meaning of 43% of the target words in a multiple-choice test, but being able to actively

recall the meaning of only 14%. Spelling recognition and word class recall fell in between, with 34% and 20% gains, respectively.

Although these studies provide insights as to how vocabulary can be learned from reading while controlling for several potentially confounding variables, such as exposure to the target words outside the program, such manipulations in research do not reflect real learning conditions. In ER programs, books are not meant to be the only source of input, as ER is only one part of a larger vocabulary-learning program that also includes other types of meaning-focused input, as well as opportunities for deliberate learning of words in isolation [2]. Additionally, ER programs promote reading of many different books, which is why research is necessary in order to examine how vocabulary knowledge develops thanks to reading multiple books through an extended period of time in authentic classroom-learning situations.

The study by Webb and Chang [20] examines vocabulary learning gains in an ER program that included audiobooks, in a Taiwanese high school. The participants ($N$ = 82) read and listened to the same 10 graded readers over the course of 13 weeks. After RWL, the students were given the opportunity to engage in various post-reading activities, such as writing book reports and maintaining a learning journal. The authors administered a meaning recognition test, including 100 words that appeared in the graded readers, at three testing times: pre-test, post-test, and delayed post-test. Approximately half of those words were already known at the time of the pre-test and the authors calculated the relative gains, considering the words that the students learned out of those that were unknown. The results showed relative gains of 44.06% of the target words on the immediate post-test and 36.66% on the three-month delayed post-test. These results were in sharp contrast with those of a control group that followed their regular L2 classes with no ER, who showed 5.19% gain on the post-test and no gain on the delayed post-test.

According to one of the principles proposed by Day and Bamford [5], in ER, reading is its own reward. However, many classroom-based studies, as Webb and Chang [20] above, include post-reading activities, and some authors have even stated that ER should be integrated within the language curriculum and not included as an isolated self-contained activity [21]. The study conducted by Boutorwick et al. [22] further examined to what extent including reading activities associated with the books contributed to L2 learning by comparing vocabulary-learning gains in an ER program that did not include post-reading activities (ER-only) and another one which included post-reading discussions in small groups (ER-plus). The study took place in a university in New Zealand, where the students read the same five graded readers. The results of the study suggest that, while both approaches led to vocabulary gains in word association knowledge, the students in the ER-plus group made more gains in mid-frequency target words that were focused on during language related episodes in their small-group discussions. These findings would support the implementation of post-reading activities in ER programs. Through these activities learners have the opportunity to continue practicing the vocabulary encountered in their reading materials. These results align with those of previous studies that have examined shorter texts and found that "reading plus" conditions, which incorporate post-reading vocabulary exercises, facilitate greater vocabulary acquisition than "reading only" conditions that solely focus on meaning comprehension [23,24].

### *1.2. Reading Only (RO) versus Reading-While-Listening (RWL)*

It has been demonstrated that vocabulary learning can be affected by the input mode through which the students are exposed to the target words. In this respect, learners' exposure could be exclusively aural (listening), written (RO), or a combination of both written and aural (RWL). Additionally, vocabulary learning may occur in multimodal conditions, in which written and aural input is also supported by the presence of images representing the target words, with those images being either static (books with pictures) or dynamic (videos). It has been shown that learners interact with written text differently in multimodal conditions including audio and pictorial support [25–27], and that multimodal

conditions facilitate vocabulary learning [28]. Although the books used in the current study also contained pictures, the main aim of the present investigation only concerns in which way the presence or absence of aural input affected vocabulary learning.

While some studies have investigated the effectiveness of ER programs that incorporate audiobooks (e.g., [20]), there is a lack of research comparing the effectiveness of RO versus RWL in the case of ER. Brown et al. [8] conducted one of the first studies in this direction. The authors examined vocabulary learning gains among Japanese college students who were learning English as a foreign language (EFL) and read three graded readers in either RO, RWL, or listening only (LO) modes. To ensure that vocabulary gains were solely attributed to reading, the authors substituted 28 words in each book with non-words. The multiple-choice test results revealed that 48% of the target non-words were acquired in the RWL mode, compared to 45% in the RO mode and 29% in the LO mode for meaning recognition. However, in the translation test, the gains were smaller, with only 16%, 15%, and 2% in the RWL, RO, and LO modes, respectively. Notably, the differences between LO and the other two modes were statistically significant, while the differences between RO and RWL were not.

Webb and Chang [6] further compared RO and RWL in the context of repeated reading in a Taiwanese high school, in which EFL learners read 28 short stories of approximately 300 words, and around 2 min 30 s in the RWL mode, over two seven-week periods. The students had to read each story at least twice with the main aim of understanding and enjoying the content, although they could also ask questions or use dictionaries if necessary. The results of the vocabulary tests showed that repeated reading was a successful approach that encouraged vocabulary learning in both modes, but RWL was significantly more beneficial than RO.

More recently, several studies have compared the two reading modes for the acquisition of collocations, instead of single words. Webb and Chang [7] provided further evidence in support of RWL versus RO for learning of English collocations through a graded reader, in the case of college students in Taiwan. These results were replicated in Vu and Peters' [29] study with Vietnamese EFL college students, who read three different fiction graded readers. According to both studies, RWL could be more beneficial because the audio support helps learners segment the input into meaningful chunks, which facilitates comprehension. The incorporation of the prosodic features of the audio is claimed to be especially useful in the noticing and processing of collocations. However, the advantages of RWL over RO are still far from being established.

Dang et al. [30] investigated the acquisition of collocations presented in various modes during an academic lecture. These modes included RWL and RO, as well as LO, viewing, and viewing with captions. The study was conducted with Chinese college students. In contrast to the previously cited studies, RO facilitated the learning of academic collocations more than RWL. According to the authors, the content of the lecture might have been challenging for learners, who might have lacked enough resources to notice and focus on the target collocations while doing LO or RWL, in contrast to the self-paced RO mode.

In a similar vein, Tuzcu [9] did not find the RWL mode to be more advantageous than the RO mode for the acquisition of medical collocations in the case of college students learning English in the US. One of the reasons the author presents for the lack of advantage of the RWL mode is that the learners' proficiency was quite high and they might have read ahead of the audio, which could have slightly disrupted the reading process. In line with claims made by Conklin et al. [31], we need a better understanding of how reading alignment (or misalignment) with the audio affects processing and vocabulary learning in RWL. Another study by Dang et al. [32], focusing on single words, also failed to find differences between RO and RWL to an academic lecture.

The conflicting results from the previously described studies suggest that more research is necessary in order to investigate whether vocabulary learning from reading is encouraged more easily by RO, or by including audio support, as in RWL. Considering previous findings, text genre might also influence whether one input mode is more benefi-

cial than the other; more specifically, it seems that academic vocabulary learning does not benefit as much from the RWL mode.

### 1.3. Research Gaps and Research Questions

There are some research gaps that the current study aims to fill. First, there is a lack of research on vocabulary learning through ER that exclusively consists of science graded readers. Considering previous findings from studies focusing on academic language, the presence of the audio in ER programs that include non-fiction graded readers might be less beneficial than in studies that have used fiction graded readers. Apart from the theoretical interest in investigating the conditions in which audio support is beneficial for vocabulary learning from reading, examining ER programs that include science graded readers would also be relevant for pedagogical reasons. Obtaining some insights into how vocabulary learning can be fostered through science graded readers would be useful in many contexts in which primary school students learn science in English through CLIL.

Second, not many studies on ER have focused on younger learners in primary school, and the studies that have examined the effect of reading mode (RO vs. RWL) have mostly included adults. Nakanishi's [4] meta-analysis on ER research did not include any ER program with children, but the results suggest that the benefits of ER programs increase with participants' age. Despite this finding, in his recommendations for further research, Nakanishi advocates for further research involving younger learners, positing that ER programs could be particularly motivating for this age group (see also [33,34]). Research is needed, therefore, to shed more light on the potential benefits of ER for a younger population. Since the publication of Nakanishi's meta-analysis, some studies have appeared focusing on primary school L2 learners, analyzing L2 gains as well as students' attitudes towards ER [34–36]. Overall, studies with younger learners indicate that ER is as effective as teacher-fronted instruction and has the added benefit of being particularly motivating for students. In light of the previously mentioned gaps, the present study aims to answer the following research questions:

1. To what extent can primary school students learn science vocabulary through an ER program including graded readers of scientific content?
2. Are there differences between RO and RWL?

The data from the present study is part of a larger project that, apart from vocabulary, examined the development of different aspects of L2 learning, including listening and reading comprehension and fluency, and students' L2 learning motivation [35,36]. The present study focuses on vocabulary learning through the whole academic year, and adds to the findings reported by Tragant et al. [36], which measured different L2 areas during the first half of the year.

## 2. Materials and Methods

### 2.1. Participants

For this study, 96 10–11-year-old students in four intact grade 5 classes at the target school were considered, comprising the entire available population for this grade. Within those classes, two were randomly assigned to an extensive RWL program ($N$ = 47; 20 girls), one to a RO program ($N$ = 24; 11 girls) and another one served as a control group ($N$ = 25; 10 girls). As will be seen in the results, not all the students were present at all testing times. The students attended a school in Barcelona, partly funded by the government, where most families could be considered middle-class and highly educated: according to a background questionnaire administered to the students, 70% of the mothers in the student sample had a university degree. Most students had an English proficiency of A1, according to their teacher's estimation. All the students had good knowledge of Catalan and Spanish, which are the official languages in Barcelona.

The school provides more English hours than the typical schools in Spain and has a strong focus on reading (e.g., the students always carry a book in their backpack, which they read when they finish the assigned activities in any class period). All the students

received 7 h of English instruction per week: 3 h were devoted to regular EFL lessons, 2 h to science in English following the CLIL approach, and the remaining 2 h were devoted to extensive RO/RWL or to communicative practice in the case of the control group.

### 2.2. The ER Program

The ER program ran from October through May, roughly aligning with the school year, which goes from September until June. The program aimed to enhance the children's English language skills by exposing them to a large amount of input in this language. It also supported their learning of science and vocabulary in their English science class.

The ER program included 39 science graded readers, 21 of which were read during the first part of the school year (October–February) and 18 during the second part (February–May). For the sake of simplicity, this paper refers to term 1 and term 2, respectively, for these two blocks, even though these terms did not correspond to the actual academic terms. The students devoted one class session (50 min) to each book, except for three sessions in term 1 that included two short books. The books were chosen from the following collections: Macmillan Science Readers, Macmillan Children's Discover, Oxford Read and Discover, and Benchmark Education. Some titles included in the collections were *Volcanoes*, *Life in the Forest*, *Animals at Night*, *The Power of Storms*, etc. The books were between 15–31 pages long, and their difficulty increased as the school year advanced to adapt to learners' proficiency. Considering the importance of choosing reading material that is appropriate for the learners [3], in term 1, the selected graded readers were aimed at grades 3–4 and had on average 908 words; while in term 2, they were 1615 words on average and were aimed at grades 4–5. During the first term, the duration of the audio in the RWL condition was 12 min, whereas in the second term the average was 19 min. The researchers together with the teachers decided which book(s) should be read in each session and all students read the same book(s).

### 2.3. Reading Procedure

In order to adapt to the learners' needs and preferences, the procedure was slightly different in term 1 and 2. In term 1, at the beginning of each session, the teachers distributed the material, which included the books, a dictionary, and a workbook, which contained different post-reading activities related to the books. Additionally, the learners in the RWL condition received headphones, and MP3 players. When the students had all the material, they first browsed through the assigned book and then started reading/reading-while-listening independently. After a first reading, they were asked to write eight words they would like to remember and their corresponding L1 translation, either by using a dictionary, asking the teacher, a classmate, etc. After this, the students were asked to read the book independently again, unless the book was very long (audio 20 min or longer), in which case they only re-read some parts. After re-reading, the students were instructed to write a minimum of three questions, either true/false or multiple-choice, based on the content of the book. At the conclusion of the reading session, all materials were collected.

In order to learn about students' attitudes towards the ER program, at the end of term 1, a series of interviews were conducted [35]. Despite the overall positive attitudes towards the reading sessions, some aspects of the program received less favorable feedback. Most notably, students expressed dissatisfaction with the limited time allotted for writing questions about the content of the books after the second reading. Furthermore, the students displayed limited interest in reading the books twice.

Bearing in mind the students' feedback, and considering the classroom-based nature of our research project, adjustments were deemed necessary to encourage further engagement in ER during the second term. Instead of requiring students to write their own questions about the book's content, we provided a set of wh- and true/false comprehension questions for each book read in term 2. This new approach allowed students to further work on the content of each graded reader without having to read the entire book again, which was a requirement in term 1.

### *2.4. Vocabulary Tests*

Two different vocabulary tests were designed in order to examine students' learning of science vocabulary through extensive RO/RWL throughout the school year. The first test contained target words that appeared in the books read during term 1, while the second test contained target words from the books read during term 2. Both tests were L2-L1 matching vocabulary tests, which measured students' meaning recognition and were similar to the tests used by Webb and Chang [20]. Meaning recognition was chosen as the target vocabulary component, because it develops earlier and it was considered that learners' vocabulary gains would be better captured through this type of test [8,18].

Each test included 50 items in total, all of which were nouns, distributed over 10 blocks, each containing five target words and five L1-matching translations (provided in both Catalan and Spanish when the word was not the same in the two languages), plus one distractor. The distractor was semantically related to one of the target words. A special effort was made to avoid cognate words. Five different versions of each test, including the same 10 blocks but in a different order, were created in order to prevent cheating. See Figure 1 for an example.

| | |
|---|---|
| 1. tar | carretera ____ |
| 2. fence | tanca/valla ____ |
| 3. tide | arrel/raíz ____ |
| 4. root | quitrà/alquitrán ___ |
| 5. road | marea ___ |
| | branca/rama ___ |

**Figure 1.** Excerpt from the vocabulary test.

Both vocabulary tests had the same format and number of items, but each contained different target words. Appendix A provides a comprehensive list of all the words featured in each test, along with their frequency, as determined by Lextutor [37]. Additionally, we calculated the frequency of the target words in the graded readers, their glossaries, and in the science English coursebook. Although many of the words on the tests were likely to be unknown to most learners, the tests were designed so that children would recognize at least some of the words, thereby avoiding frustration throughout the 50-item vocabulary tests [20]. The suitability of the target words was confirmed by the teachers before the tests were administered. According to Cronbach alpha, both Test 1 (0.909) and Test 2 (0.913) achieved good reliability scores.

### *2.5. Data Collection Procedure*

For vocabulary Test 1, which comprised the vocabulary included in the science graded readers read during the first term, data were collected over three testing times, pre-test (end of September), post-test (February), delayed post-test (June). In contrast, vocabulary test 2 assessed the vocabulary in the books read during the second term and was administered only twice with the pre-test conducted in February and the post-test in June. As June marked the end of the school year, further testing could not be carried out.

The tests were administered in written form in class, where researchers carefully went through the instructions and guided the students through the sample item, which included a block of three words familiar to the students (*house*, *garden*, *dog*). A time limit of 20 min was allotted for each test. As explained before, the data included in this study are part of a larger research project [35,36]. Therefore, apart from taking the vocabulary tests, the students took other tests in the same session, which lasted 50 min. The vocabulary tests were always administered first.

### 2.6. Analyses

The scoring system for the vocabulary tests awarded students one point for accurately matching each L2 target word with its corresponding L1 pair, and zero points for incorrect matches. The results of the vocabulary tests were analyzed in terms of relative gains, following previous research [20]. This approach considers students' initial vocabulary knowledge, which varied both within and between groups, as expected in a real classroom setting. The formula that was used for each participant's scores was the following: [(post-test score − pre-test score)/(total number of items − pre-test score)] × 100. This formula produces scores that indicate the percentage of words learned by students at the end of each term, relative to the words they were capable of learning based on their pre-test scores. Relative gains were also computed to examine long-term learning of the words included in term 1, but instead of using the post-test scores in the above formula, the delayed post-test scores were used.

Statistical analysis was conducted using SPSS version 27 [38]. Since all the data were normally distributed, according to the Kolmogorov—Smirnov test, two one-way ANOVAs were performed with the relative vocabulary gains experienced by the learners in the three conditions (RO, RWL, and control). The first ANOVA analyzed the relative gains with respect to the vocabulary that appeared in term 1 books. This analysis included immediate and long-term gains. The second ANOVA focused on the immediate gains related to the vocabulary included in term 2 books.

## 3. Results

### 3.1. Vocabulary Learning in Term 1

Table 1 shows the descriptive statistics obtained from Test 1, which includes the scores of the pre-test, post-test, and delayed post-test, as well as the immediate and delayed gains. The scores on the pre-test were very similar across conditions. As expected, the students knew some of the vocabulary included in the test. On the post-test, all the learners demonstrated an increase in vocabulary knowledge, but the gains between the two testing times were more obvious for the two ER groups. It can also be observed that the learning gains were maintained through time (four months later), as evidenced by the delayed post-test scores.

**Table 1.** Descriptive statistics Test 1: means and *SD* in parentheses.

| Group | Pre-Test/50 | Post-Test/50 | Delayed Post-Test/50 | Immediate Gains (%) | Long-Term Gains (%) |
|---|---|---|---|---|---|
| RO ($n$ = 22) | 27.31 (10.5) | 37.09 (10.8) | 35.94 (10.3) | 46.7% (27.5) | 44.9 (24.7) |
| RWL ($n$ = 47) | 28.97 (9.4) | 39.00 (9.9) | 38.04 (9.5) | 54.1% (27.0) | 50.4% (26.2) |
| Control ($n$ = 25) | 28.96 (9.3) | 33.80 (9.7) | 32.80 (10.5) | 20.9% (36.0) | 12.6% (37.9) |

In the delayed post-test there were some students who were absent due to other school-related projects and the sample was: RO = 10, RWL = 41, control = 10.

The results of the first ANOVA comparing term 1 relative gains across conditions show that there were significant differences between the three groups in immediate and long-term gains: $F(2, 93) = 10.29$, $p < 0.001$, eta2 = 0.184, and $F(2, 69) = 7.41$, $p = 0.001$, eta2 = 0.184, respectively. Post-hoc analyses with Bonferroni adjustments for multiple comparisons applied to the immediate-gain scores suggest that the mean difference (*MD*) between the control and the RWL groups was statistically significant ($MD = -33.21$, $p < 0.001$); this was also the case between the control and the RO groups ($MD = -25.76$; $p = 0.012$). The difference between RO and RWL was not statistically significant ($MD = -7.45$, $p = 1.00$). Similarly, Bonferroni comparisons for long-term gains indicate that the difference between the control and RO groups was statistically significant ($MD = -32.20$, $p = 0.12$), as well as between the control and RWL groups ($MD = -37.75$; $p = 0.001$), with no differences between RO and RWL ($MD = -5.55$; $p = 1.00$).

### 3.2. Vocabulary Learning in Term 2

Test 2 pre-test scores were very similar across conditions and very similar to those of Test 1, which suggests that the level of difficulty was equivalent considering students' initial knowledge. However, the results of the post-test and the vocabulary gains were different from those reported for term 1. The students learned fewer words in term 2, and the group that made the smallest gains was the RO (see Table 2). According to the results of the ANOVA, there were no differences in relative gains among the three conditions: $F(2, 95) = 0.995$, $p = 0.373$, eta2 = 0.021.

**Table 2.** Descriptive statistics Test 2: means and SD in parentheses.

| Group | Pre-Test/50 | Post-Test/50 | Immediate Gains (%) |
|---|---|---|---|
| RO ($n$ = 24) | 28.83 (9.6) | 30.50 (8.71) | 5.90% (26.77) |
| RWL ($n$ = 47) | 28.59 (10.3) | 33.04 (9.6) | 17.92% (35.4) |
| Control ($n$ = 25) | 27.76 (8.7) | 31.48 (11.2) | 14.84% (37.5) |

## 4. Discussion

The present classroom-based study examined the contribution of RO and RWL to English science vocabulary learning through an ER program in a primary school in Spain. The ER program lasted for the whole school year and vocabulary learning was examined at the end of two terms: term 1 which included 21 graded readers that were read from October until February; and term 2, which concerned 18 books that were read from February until the end of May. The results of term 1 and term 2 will be discussed separately.

### 4.1. Vocabulary Learning in Term 1

The results of vocabulary Test 1 revealed that the use of science graded readers in both the RO and RWL conditions promoted gains in vocabulary knowledge among the students. These gains were significantly higher than those observed in the control group, both immediately after the intervention and in the long-term, four months later. These positive findings align with previous studies that have shown the effectiveness of fiction books in enhancing vocabulary acquisition.

The current ER program produced an average immediate and long-term relative gain in vocabulary of 50.43% and 47.64%, respectively, when combining both approaches (RO and RWL). These gains are consistent with those reported in controlled studies where vocabulary learning was incidental, such as Pellicer-Sánchez and Schmitt [18], who reported a 46% increase in meaning recognition on an immediate post-test. The present results are also similar to those of Brown et al. [8] and Waring and Tataki [17], who reported immediate vocabulary gains of over 40%. Furthermore, the present study aligns with research that has examined the implementation of ER in naturalistic classroom settings, such as Webb and Chang's [20] study on Taiwanese high school students, which reported gains of 44.06% and 36.6% on the immediate and delayed post-tests, respectively. It is worth emphasizing that the vocabulary gains achieved in the current ER program were more successfully retained than in previous studies [8,17,20]. The target students were able to recall most of the words they had learned in term 1 four months later at the end of the school year, experiencing only a slight average loss of 2.8%.

In summary, the findings from term 1 of the current study confirm previous findings regarding the positive effects of ER with fiction graded readers, and thus support ER programs including science books for the learning of scientific vocabulary.

Regarding the second research question, the results of Test 1 showed that, even though RWL promoted more immediate vocabulary gains than RO (54.1% vs. 46.7%) and also better long-term gains (50.4% vs. 44.9%), the difference did not reach statistical significance. These findings are in contrast with those reported in studies that have used fiction graded readers for the learning of collocations, such as Vu and Peters [29] or Webb and Chang [7]. According to these studies, RWL is more beneficial than RO because the audio helps learners

to segment the input in a more target-like fashion, which could facilitate the noticing and processing of L2 collocations. Similarly, in the context of repeated reading of short stories, Webb and Chang [6] also found that the RWL mode encouraged higher vocabulary gains.

On the other hand, the results of our investigation are consistent with Brown et al.'s [8] study, which found gains of 48% and 45% in the RWL and RO conditions, respectively, using three fiction graded readers. Similarly, studies focusing on academic vocabulary acquisition [9,30,32] have failed to show differences between RWL and RO. These findings support the argument that the complexity of scientific content may demand more attentional resources from learners, potentially limiting the benefits of audio input.

Additionally, when implementing RWL, it is crucial to consider whether learners' natural reading pace is aligned with the audio. As Tuzcu [9] suggests, misalignment between the two modes of input might account for cases where RWL does not provide a significant advantage. Given the complexity of the content of the books in the current study, it is plausible that the audio was too fast for some participants, which could have undermined the benefits of RWL.

### *4.2. Vocabulary Learning in Term 2*

The analyses of vocabulary gains experienced during the second term were not in line with those reported during term 1, although there is one common finding, which is the lack of significant differences between the RO and RWL modes, even if the scores in the RWL mode were descriptively higher. Considering this result, it is important to conduct more studies comparing the two modes for different types of text genres (e.g., fiction vs. non-fiction), vocabulary targets (e.g., single words vs. collocations), and learner populations (younger vs. older learners) to explain the conflicting findings in the literature regarding the benefit of audio-supported reading for vocabulary learning.

The vocabulary gains observed in term 2 were disappointing when compared to the gains in term 1. Although students had similar prior knowledge of the target words in Test 1 and Test 2, the percentage of vocabulary gains in term 2 was low (RO = 5.90%; RWL = 17.91%). These gains are more in line with studies investigating incidental vocabulary acquisition through reading alone as opposed to "reading plus" conditions [39].

Although there was a certain intentional learning component in the ER program under study, it is worth noting that the vocabulary task required students to choose their own words to focus on, but this approach may not have been the most effective to promote engagement with vocabulary. During the interviews, many students reported facing various difficulties with this task. For some, it was challenging to identify unknown words in the graded readers or to find some words in the dictionary, while others reported choosing words they already knew to avoid the extra effort of looking up the meaning of new words. We can speculate that these challenges became more frustrating towards the end of the program than at the beginning, which could partly explain the lower vocabulary gains in term 2. Another possible explanation for the higher gains in term 1 may be connected to the benefits of repeated reading [6]. As explained in Section 2.3, the students were instructed to read the whole graded reader twice during term 1 but not during term 2.

Additionally, it should be highlighted that, in order to adapt to the learners' proficiency development throughout the year, the graded readers in term 2 were slightly more difficult than those used in the first part of the year. This difference in complexity may have also influenced the results.

Furthermore, the length of the ER program may have played a role in the lack of clear benefits observed in the second half of the school year. Learners may have lost over time their initial motivation for doing something different from their regular English classes, especially in the RO group, which obtained the lowest vocabulary gains in term 2. As noted in Tragant and Vallbona [35], only 31.8% of the students in that group reported an interest in continuing with the ER sessions the following year, compared to 62.5% in the RWL group. It is worth mentioning that the availability of MP3 players for the RWL group

probably served as an additional motivational factor for the young learners. The higher motivation to do RWL as opposed to RO has also been observed in other studies [8,27].

Another aspect that could account for the results is related to the timing of Test 2, which was administered a few weeks before the end of the school year. By this point, students may have been less motivated to take tests, particularly as they had already completed exams in their other courses.

In summary, the ER program in term 2 did not yield results as positive as in term 1, with lower overall vocabulary gains (11.9% gains vs. 50.4% in term 1) that were not significantly different from the control group. Several factors could have contributed to the ER treatment being less supportive of vocabulary learning, such as the employment of more demanding graded readers, the potential decrease in students' motivation during the program's final weeks, or the timing of Test 2 at the end of the school year.

### *4.3. Limitations and Future Research*

The present study has limitations that stem primarily from its classroom-based approach and its emphasis on ecological validity. One limitation is that the divergent outcomes observed between term 1 and term 2 could be attributed to one or more factors that are difficult to disentangle. To differentiate these potentially influential factors, further controlled studies are necessary. First, future research should consider including the same reading procedure as well as the same type of post-reading activities for reading comprehension and vocabulary learning throughout the program, ensuring that students are consistently exposed to effective and engaging exercises. Secondly, it is important to ensure that all books are of comparable difficulty, in order to prevent discrepancies in learning outcomes related to this factor. Thirdly, learners' motivation could be thoroughly and systematically examined throughout the program and statistical analyses could be performed to assess more rigorously the role of motivation in students' vocabulary gains throughout the ER program.

Furthermore, future research could aim to control for exposure to the target words both within and outside of the ER program. The vocabulary-focused activity used in this study may have resulted in varying degrees of engagement with the target words among learners, which were not considered in the analyses. Regarding exposure to the target words outside ER, aiming for complete control is probably an unrealistic goal in a long-term ER program. Previous studies that have attempted to achieve such control used non-words and were more limited in their scope, typically including only a few graded readers. It is crucial to emphasize, however, that ER is not expected to be the sole source of L2 vocabulary learning [2].

Another avenue for research would be to explore how the number of encounters with the target words in the graded readers affect vocabulary gains, as examined in previous studies [20]. A further line of inquiry could involve investigating how text and audio alignment affects vocabulary learning by analyzing learners' eye movements while reading in the RWL mode.

Despite the limitations, the findings of this exploratory classroom-based study hold ecological validity, allowing for more straightforward pedagogical implications to be drawn compared to studies with more controlled learning conditions. The study investigated an authentic ER program, which was implemented based on input from teachers and students and was aimed at serving their interests, rather than being purely research-focused.

It should be mentioned, however, that certain aspects of the current reading program did not adhere to Day and Bamford's recommendations for ER [5]. For instance, the program did not provide a range of literary genres, but rather focused on scientific vocabulary learning, which was a primary objective of the study. Additionally, learners were not allowed to choose their own books in order to facilitate the investigation of vocabulary acquisition. However, some scholars have argued that these factors may not be essential components of ER programs, as defined in the literature [10].

## 5. Conclusions and Pedagogical Implications

The present study provides valuable insights into the effectiveness of an ER program that incorporates science graded readers for improving vocabulary learning among primary school students. Although most previous research on ER has focused on older learners, this study highlights the potential benefits of non-fiction ER for primary school students. However, further research is needed to replicate these findings in different contexts to obtain generalizable results.

The current findings have important pedagogical implications, particularly for schools that offer science instruction in English as a foreign language, as in CLIL contexts. Using ER programs that include science graded readers can facilitate not only the acquisition of new vocabulary but also the learning of scientific content. In fact, when the target students were asked to describe what they had learned during the ER program, most of them referenced information related to the content of the graded readers, rather than language-related aspects [35].

A second recommendation would be for teachers to include audiobooks. Although the vocabulary gains between RWL and RO were not statistically significant in the present study, RWL consistently showed higher gains. Additionally, it was found that the students in the RWL group had significantly more positive attitudes towards ER than their peers in the RO group [35]. Maintaining motivation is crucial, and teachers should regularly assess their students' attitudes towards ER to determine the optimal duration of the program. Additionally, it is important to keep learners engaged while they are being tested by incorporating additional motivational strategies that encourage them to perform their best during tests. Doing so would make it easier to gauge more accurately the learning outcomes after ER.

Another pedagogical recommendation in light of the findings from the present study, and also considering previous findings reported in the literature [20,22], is to include stimulating vocabulary-focused post-reading activities to maximize the degree of students' engagement with the vocabulary that appears in the books.

To conclude, the present study reflects the complexity of classroom-based research and underscores the importance of collaboration among L2 acquisition researchers, practitioners, and L2 learners in generating insights with practical implications for language education. Such collaboration has the potential to bridge the gap between research and teaching practice, ultimately leading to more effective support for L2 learning.

**Funding:** This research was funded by the Spanish Science and Innovation Ministry grants FFI2013-40952-P and PID2019-110536GB-I00.

**Institutional Review Board Statement:** The study adhered to the ethical principles proposed by the Research Ethics Committee at the University of Barcelona (2020-2022) and respected the privacy and confidentiality of the participants involved.

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

**Data Availability Statement:** Data are unavailable due to privacy restrictions.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

| Test 1 | | | | | Test 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Words** | **Freq. Band** | **Freq. Readers** | **Freq. Glossary** | **Freq. Courseb.** | **Words** | **Freq. Band** | **Freq. Readers** | **Freq. Glossary** | **Freq. Courseb.** |
| bark | 2 | 4 | 2 | 0 | arrow | 4 | 3 | 0 | 0 |
| bat | 2 | 14 | 0 | 1 | lane | 2 | 5 | 0 | 0 |
| beetle | 6 | 27 | 0 | 1 | cattle | 3 | 3 | 0 | 0 |
| blood | 1 | 8 | 0 | 0 | stream | 2 | 6 | 1 | 0 |
| bone | 1 | 12 | 3 | 2 | glass | 1 | 32 | 2 | 0 |
| brain | 2 | 8 | 0 | 0 | goods | 3 | 5 | 0 | 0 |
| bud | 5 | 5 | 0 | 0 | race | 1 | 2 | 0 | 0 |
| burrow | 6 | 4 | 1 | 0 | law | 1 | 8 | 0 | 0 |
| claw | 5 | 3 | 1 | 2 | candle | 4 | 2 | 0 | 0 |
| crust | 4 | 15 | 1 | 5 | saddle | 5 | 3 | 0 | 0 |
| eagle | 4 | 8 | 0 | 1 | tar | 7 | 10 | 1 | 0 |
| feather | 2 | 15 | 2 | 2 | fence | 2 | 3 | 0 | 0 |
| field | 1 | 3 | 0 | 0 | tide | 2 | 6 | 0 | 1 |
| flea | 7 | 12 | 0 | 0 | root | 2 | 10 | 0 | 1 |
| flock | 4 | 7 | 0 | 0 | road | 1 | 15 | 2 | 0 |
| flood | 2 | 2 | 0 | 0 | shelter | 2 | 2 | 1 | 0 |
| fox | 2 | 11 | 2 | 0 | shark | 5 | 58 | 0 | 2 |
| fur | 2 | 33 | 3 | 2 | squid | 8 | 10 | 0 | 1 |
| hill | 1 | 5 | 0 | 2 | heat | 1 | 10 | 0 | 6 |
| hole | 1 | 14 | 3 | 0 | cliff | 2 | 15 | 0 | 3 |
| kite | 6 | 14 | 0 | 0 | steam | 2 | 5 | 1 | 0 |
| land | 1 | 26 | 1 | 19 | nostril | 6 | 3 | 1 | 0 |
| moth | 6 | 10 | 0 | 0 | gold | 1 | 4 | 0 | 0 |
| mud | 2 | 13 | 0 | 2 | reef | 5 | 17 | 1 | 0 |
| nest | 2 | 26 | 2 | 0 | layer | 3 | 14 | 0 | 14 |
| paw | 6 | 3 | 0 | 0 | hut | 4 | 5 | 0 | 0 |
| poison | 2 | 9 | 2 | 0 | shield | 3 | 3 | 0 | 0 |
| pond | 4 | 4 | 0 | 1 | path | 2 | 13 | 0 | 0 |
| seal | 2 | 8 | 2 | 0 | army | 2 | 7 | 0 | 0 |
| seed | 2 | 25 | 2 | 1 | earthquake | 5 | 12 | 1 | 4 |
| silk | 3 | 12 | 2 | 0 | shell | 2 | 7 | 1 | 4 |
| skin | 1 | 13 | 1 | 4 | eel | 8 | 3 | 0 | 0 |
| sloth | 11 | 6 | 1 | 0 | pipe | 2 | 10 | 0 | 1 |
| slug | 6 | 6 | 0 | 0 | patern | 2 | 5 | 1 | 0 |
| snail | 6 | 11 | 0 | 1 | health | 1 | 9 | 0 | 0 |
| soil | 2 | 15 | 3 | 2 | stall | 4 | 2 | 0 | 0 |
| stem | 3 | 12 | 0 | 1 | crop | 3 | 2 | 1 | 0 |
| swarm | 6 | 6 | 0 | 0 | fin | 6 | 2 | 0 | 2 |

| Test 1 | | | | | Test 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Words** | **Freq. Band** | **Freq. Readers** | **Freq. Glossary** | **Freq. Courseb.** | **Words** | **Freq. Band** | **Freq. Readers** | **Freq. Glossary** | **Freq. Courseb.** |
| tadpole | 10 | 10 | 2 | 0 | lake | 1 | 36 | 1 | 4 |
| tail | 1 | 12 | 2 | 10 | clay | 4 | 4 | 0 | 0 |
| tooth | 1 | 6 | 1 | 0 | wall | 1 | 14 | 0 | 2 |
| trunk | 2 | 7 | 1 | 5 | coal | 2 | 2 | 0 | 2 |
| tusk | 10 | 12 | 1 | 0 | concrete | 3 | 6 | 0 | 0 |
| walrus | 12 | 2 | 1 | 0 | jellyfish | | 9 | 0 | 1 |
| wasp | 6 | 13 | 0 | 0 | twin | 2 | 5 | 0 | 0 |
| wave | 1 | 7 | 0 | 3 | octopus | 9 | 6 | 0 | 1 |
| web | 4 | 10 | 1 | 0 | limestone | | 19 | 0 | 0 |
| wheel | 1 | 3 | 0 | 0 | cabbage | 6 | 3 | 0 | 0 |
| wing | 2 | 48 | 1 | 6 | danger | 1 | 12 | 3 | 0 |
| wolf | 2 | 6 | 1 | 0 | rope | 2 | 7 | 0 | 0 |
| Mean | 3.66 | 11.3 | 0.9 | 1.46 | Mean | 3.12 | 9.08 | 0.36 | 0.98 |
| Median | 2 | 10 | 1 | 0 | Median | 2 | 6 | 0 | 0 |

## References


1. Nation, P. Is It Worth Teaching Vocabulary? *TESOL J.* **2021**, *12*, e564. [CrossRef]
2. Nation, P. The Four Strands. *Innov. Lang. Learn. Teach.* **2007**, *1*, 2–13. [CrossRef]
3. Nation, P.; Waring, R. *Teaching Extensive Reading in Another Language*; Routledge: New York, NY, USA, 2020.
4. Nakanishi, T. A Meta-Analysis of Extensive Reading Research. *TESOL Q.* **2015**, *49*, 6–37. [CrossRef]
5. Day, R.R.; Bamford, J. Top Ten Principles for Teaching Extensive Reading. *Read. Foreign Lang.* **2002**, *14*, 136–141.
6. Webb, S.; Chang, A.C.-S. Second Language Vocabulary Learning through Extensive Reading with Audio support: How do frequency and distribution of occurrence affect learning? *Lang. Teach. Res.* **2015**, *19*, 667–686. [CrossRef]
7. Webb, S.; Chang, A.C.-S. How Does Mode of Input Affect the Incidental Learning of Collocations? *Stud. Second Lang. Acquis.* **2022**, *44*, 35–56. [CrossRef]
8. Brown, C.; Waring, R.; Donkaewbua, S. Incidental Vocabulary Acquisition from Reading, Reading-While-Listening, and Listening to Stories. *Read. Foreign Lang.* **2008**, *20*, 136–163.
9. Tuzcu, A. The Effects of Input Mode in Learning Declarative and Nondeclarative Knowledge of L2 Collocations. *System* **2023**, *113*, 103006. [CrossRef]
10. Waring, R.; McLean, S. Exploration of the Core and Variable Dimensions of Extensive Reading Research and Pedagogy. *Read. Foreign Lang.* **2015**, *27*, 160–167.
11. Beglar, D.; Hunt, A. Pleasure Reading and Reading Rate Gains. *Read. Foreign Lang.* **2014**, *26*, 29–48.
12. Beglar, D.; Hunt, A.; Kite, Y. The Effect of Pleasure Reading on Japanese University EFL Learners' Reading Rates. *Lang. Learn.* **2012**, *62*, 665–703. [CrossRef]
13. Bui, T.; Macalister, J. Online Extensive Reading in an EFL Context: Investigating Reading Fluency and Perceptions. *Read. Foreign Lang.* **2021**, *33*, 1–29. Available online: http://hdl.handle.net/10125/67391 (accessed on 1 February 2023).
14. McLean, S.; Rouault, G. The Effectiveness and Efficiency of Extensive Reading at Developing Reading Rates. *System* **2017**, *70*, 92–106. [CrossRef]
15. Bell, T. Extensive Reading: Speed and Comprehension. *Read. Matrix* **2001**, *1*, 1–13.
16. Krashen, S. Extensive Reading in English as Foreign Language by Adolescents and Young Adults: A Meta-Analysis. *Int. J. Foreign Lang. Teach.* **2007**, *7*, 23–29.
17. Waring, R.; Takaki, M. At What Rate Do Learners Learn and Retain New Vocabulary from Reading a Graded Reader? *Read. Foreign Lang.* **2003**, *15*, 130–163.
18. Pellicer-Sánchez, A.; Schmitt, N. Incidental Vocabulary Acquisition from an Authentic Novel: Do Things Fall Apart? *Read. Foreign Lang.* **2010**, *22*, 31–55.
19. Pigada, M.; Schmitt, N. Vocabulary Acquisition from Extensive Reading: A Case Study. *Read. Foreign Lang.* **2006**, *18*, 1–28.
20. Webb, S.A.; Chang, A.C.-S. Vocabulary Learning through Assisted and Unassisted Repeated Reading. *Can. Mod. Lang. Rev.* **2012**, *68*, 267–290. [CrossRef]
21. Green, C. Integrating Extensive Reading in the Task-Based Curriculum. *ELT J.* **2005**, *59*, 306–311. [CrossRef]

22. Boutorwick, T.J.; Macalister, J.; Elgort, I. Two Approaches to Extensive Reading and Their Effects on L2 Vocabulary Development. *Read. Foreign Lang.* **2019**, *31*, 150–172.
23. Peters, E.; Hulstijn, J.H.; Sercu, L.; Lutjeharms, M. Learning L2 German Vocabulary through Reading: The Effect of Three Enhancement Techniques Compared. *Lang. Learn.* **2009**, *59*, 113–151. [CrossRef]
24. Sonbul, S.; Schmitt, N. Direct Teaching of Vocabulary after Reading: Is It Worth the Effort? *ELT J.* **2010**, *64*, 253–260. [CrossRef]
25. Tragant, E.; Pellicer-Sánchez, A. Young EFL Learners' Processing of Multimodal Input: Examining Learners' Eye Movements. *System* **2019**, *80*, 212–223. [CrossRef]
26. Pellicer, A.; Tragant, E.; Conklin, K.; Rodgers, M.; Serrano, R.; Llanes, À. Young Learners' Processing of Multimodal Input and Its Impact on Reading Comprehension: An Eye-Tracking Study. *Stud. Second Lang. Acquis.* **2020**, *42*, 577–598. [CrossRef]
27. Serrano, R.; Pellicer, A. Young L2 Learners' Online Processing of Information in a Graded Reader during Reading-Only and Reading-While-Listening Conditions: A Study of Eye-Movements. *Appl. Linguist. Rev.* **2022**, *13*, 49–70. [CrossRef]
28. Montero Perez, M. Second or Foreign Language Learning through Watching Audio-Visual Input and the Role of On-Screen Text. *Lang. Teach.* **2022**, *55*, 163–192. [CrossRef]
29. Vu, D.V.; Peters, E. Incidental Learning of Collocations from Meaningful Input: A Longitudinal Study into Three Reading Modes and Factors that Affect Learning. *Stud. Second Lang. Acquis.* **2021**, *43*, 1–23. [CrossRef]
30. Dang, T.N.Y.; Lu, C.; Webb, S. Incidental Learning of Collocations in an Academic Lecture through Different Input Modes. *Lang. Learn.* **2022**, *72*, 728–764. [CrossRef]
31. Conklin, K.; Alotaibi, S.; Pellicer-Sánchez, A.; Vilkaite-Lozdiene, L. What Eye-Tracking Tells Us about Reading-Only and Reading-While-Listening in a First and Second Language. *Second Lang. Res.* **2020**, *36*, 257–276. [CrossRef]
32. Dang, T.N.Y.; Lu, C.; Webb, S. Open Access Academic Lectures as Sources of Incidental Vocabulary Learning: Examining the Role of Input Mode, Frequency, Type of Vocabulary, and Elaboration. *Appl. Linguist.* **2022**, amac044. [CrossRef]
33. Lightbown, P.M. Can They Do It Themselves? A Comprehension-Based ESL Course for Young Children. In *Comprehension-Based Second Language Teaching*; Courchene, R., St John, J., Therrien, C., Glidden, J., Eds.; University of Ottawa Press: Ottawa, ON, Canada, 1992; pp. 353–370.
34. Tragant, E.; Muñoz, C.; Spada, N. Maximizing Young Learners' Input: An Intervention Program. *Can. Mod. Lang. Rev.* **2016**, *72*, 234–257. [CrossRef]
35. Tragant, E.; Vallbona, A. Reading while Listening to Learn: Young EFL learners' Perceptions. *ELT J.* **2018**, *72*, 395–404. [CrossRef]
36. Tragant, E.; Llanes, À.; Pinyana, À. Linguistic and Non-Linguistic Outcomes of a Reading-while-listening Program for Young Learners of English. *Read. Writ.* **2019**, *32*, 819–838. [CrossRef]
37. Cobb, T.; Lextutor. Vocabprofile [Computer Program]. Available online: https://www.lextutor.ca/vp/comp/ (accessed on 17 March 2023).
38. IBM Corp. *IBM SPSS Statistics for Windows, Version 27.0*; IBM Corp: Armonk, NY, USA, 2020.
39. Laufer, B. Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *Can. Mod. Lang. Rev.* **2003**, *59*, 567–588. [CrossRef]

education sciences

MDPI

*Article*

# From Research in the Lab to Pedagogical Practices in the EFL Classroom: The Case of Task-Based Pronunciation Teaching

**Joan C. Mora * and Ingrid Mora-Plaza**

Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, 08007 Barcelona, Spain; imoraplaza@ub.edu
* Correspondence: mora@ub.edu

**Abstract:** Input and context-related factors identified by research as key success variables in L2 pronunciation development in immersion contexts play a very modest role in instructed foreign language (FL) learning environments. Scarce L2 exposure and use and L1-accented input make pronunciation learning extremely challenging. Current L2 speech learning models attribute difficulties in L2 speech acquisition to L2-to-L1 perceptual sound mappings guided by L1-based perception and poor phonological awareness and noticing of cross-language phonetic differences, which are typically not adequately addressed in instruction through pedagogic tasks. Explicit and incidental pronunciation teaching methods have been found effective at improving learners' pronunciation, but ways to integrate them into communicative approaches to language teaching are still largely unexplored. Thus, language education practices currently lack a research-informed pedagogical approach that incorporates principles of L2 speech learning and task-based language teaching (TBLT) into pronunciation instruction. This article (1) presents an outline of new avenues for research and practice in L2 pronunciation instruction and (2) reports on the findings of an empirical study that implemented a task-based pronunciation teaching (TBPT) approach to teaching a difficult L2 vowel contrast through computerized collaborative map tasks that could be easily integrated into communicative FL classrooms.

**Keywords:** task-based language teaching (TBLT); L2 pronunciation instruction; L2 pronunciation training; task-based pronunciation teaching (TBPT); form-focused communicative instruction; L2 vowel perception and production; map task

**Citation:** Mora, J.C.; Mora-Plaza, I. From Research in the Lab to Pedagogical Practices in the EFL Classroom: The Case of Task-Based Pronunciation Teaching. *Educ. Sci.* **2023**, *13*, 1042. https://doi.org/10.3390/educsci13101042

Academic Editor: Martin Howard

Received: 31 July 2023
Revised: 14 October 2023
Accepted: 15 October 2023
Published: 17 October 2023



## 1. Introduction

Pronunciation appears to be one of the prevailing learning challenges for L2 learners [1] and a teaching challenge for L2 teachers [2,3], especially in instructed second language acquisition (ISLA) contexts [4]. Still, both learners and teachers find pronunciation to be a crucial component of the communicative competence they wish to develop to highly proficient levels [3,5]. Despite the dramatic growth of the field of L2 speech learning and L2 pronunciation teaching and learning in recent years [6–9] and recent research exploring alternative methods to L2 pronunciation teaching beyond explicit pronunciation instruction, such as task-based language teaching (TBLT) approaches to pronunciation-focused instruction [10–15] or computer-assisted pronunciation training [16–18], there is still a dearth of research on the acquisition of L2 pronunciation in ISLA contexts, including English as a foreign language (EFL) learning contexts in Spain.

Despite the challenging aspects of learning L2 pronunciation in an ISLA setting, several meta-analyses have shown that L2 pronunciation instruction is effective [19–22] and that phonetic training of specific L2 speech sounds improves L2 learners' perception and production of difficult L2 sounds and sound contrasts [23–25]. However, enhancing the outcomes of pronunciation instruction through methods that enhance learners' attention to phonetic form and, at the same time, can be fully integrated into communicative teaching

practice remains a real challenge, as communication is primarily a meaning-oriented activity [3]. In addition, research-based approaches to L2 pronunciation instruction based on the intelligibility principle [26,27], in line with current L2 speech learning models [28,29], suggest that pronunciation instruction should focus on target pronunciation features with high functional load that are difficult to acquire due to L2-to-L1 perceptual sound mappings guided by L1-based perception. In other words, pronunciation instruction methods should consider L1-specific learning difficulties (at the segmental and suprasegmental level) and enhance phonological awareness and noticing of cross-language differences in addition to L2-specific phonetic differences between contrasting phonetic features. Implementing these teaching goals within communicative language learning requires large amounts of creativity in task design and implementation to enhance learners' attention to linguistic form (or phonetic form) while learners are performing communicative language learning tasks. A TBLT approach to pronunciation instruction, task-based pronunciation teaching (TBPT), may prove successful at achieving these task design goals. We illustrate this through the empirical study we report on in Section 4. Given the specific low-input conditions associated with ISLA and its well-attested limiting effects on the development of oral skills [30], pronunciation development is unlikely to take place without pronunciation training techniques providing individual intensive practice on specific (e.g., segmental contrasts) and global (e.g., intelligibility) dimensions of pronunciation proficiency. Such techniques (see Section 3 below) could be provided as a supplement to and in combination with TBPT pedagogic interventions.

## 2. L2 Speech Models and FL Pronunciation Learning

The acquisition of L2 phonetics and phonology in adulthood has been shown to be profoundly influenced by learners' L1 perception [31]. Speech learning theories have attributed learners' capacity to categorize L2 sounds accurately to their ability to discern differences between L1 and L2 sounds and their degree of phonological awareness and noticing of cross-language phonetic differences.

For example, Ref. [29]'s revised Speech Learning Model (SLM-r) argues that L2 sounds that are perceptually dissimilar to L1 sounds are easier to acquire and a separate L2 category may be created for them (e.g., L2 English /ɜ:/ having no clear match in L1 Catalan/Spanish) than L2 sounds that are indistinguishable from L1 sounds, for which a distinct category will not be formed (e.g., L2 English /f/ mapped onto L1 Catalan/Spanish /f/). These two scenarios are not predicted to pose any learning problems to L2 learners. However, for L2 sounds that are highly similar perceptually to L1 sounds, learners are unlikely to create separate L2 categories (e.g., L2 English /ʌ/ mapped onto L1 Catalan /a/). In such cases, a composite L1–L2 phonetic category may develop and lead to L1-based foreign-accented productions.

Similarly, the Perceptual Assimilation Model [28] extended to the L2 (PAM-L2) posits that success at detecting L2 phonological contrasts is dependent on how L2 phonemes are initially assimilated to the L1 phonemic inventory. In PAM-L2, when two L2 sounds are equally assimilated to a single L1 sound category (single-category assimilation), learners' ability to acquire the L2 sound contrast will be impaired as discriminating them will be very difficult (e.g., English /r/-/l/ for Japanese learners), whereas when two L2 sounds are unequally mapped onto the same L1 sound category so that one is judged as a better exemplar of the L1 category than the other (i.e., category-goodness assimilation), learners' ability to discriminate the L2 contrast will be moderate, making the acquisition of the contrast easier. Whereas SLM-r and PAM-L2 predictions for L2 sound acquisition have been applied to explaining phonological acquisition for monolingual immigrants living in an environment where L2 is spoken predominantly, neither model was designed with a foreign language (FL) instructional setting in mind, where L2 input is scarce and accented [30,32]. For example, [4] predicted that single-category and category-goodness assimilations were less likely to be acquired in the classroom than via immersion, especially if phonetic differences between the L2 phonemes were lacking in L2-accented spoken input or if L2

learners produced them on the basis of their written form. In fact, oral interactions in the FL classroom are likely to be L1-accented because teachers and students are likely to speak L2 with a shared L1-accent. In addition, unlike in naturalistic language learning, in FL instruction, vocabulary and grammar are primarily taught through written input, and little time is dedicated to communicative pronunciation teaching [4]. Still, models like SLM-r and PAM-L2 can inform L2 pronunciation instruction by explaining why certain L2 contrasts are easier to learn than others, which may be helpful in setting priorities when selecting pronunciation instruction targets and designing pedagogic tasks to learn them.

For example, Ref. [4] proposes that learners should be provided with plenty of opportunities for tuning in to the phonetic differences that distinguish L2 phonological contrasts prior to the acquisition of a large vocabulary so as to avoid homophony of minimal-pair words during L2 word recognition processes. An interactive map task where perceiving and producing a difficult L2 vowel contrast is essential for task resolution, such as the one described in the empirical study we report on in Section 4, can serve this purpose by raising learners' awareness of such phonetic differences and by providing practice in the distinction of the target L2 sounds. In addition, the introduction of written forms should be initially delayed, or orthography should provide one-to-one correspondence between phonemes and graphemes. Last but not least, pronunciation instruction should aim at orienting learners' attention to phonological contrasts in communicative settings so that L2 phonological learning can extend to spontaneous conversations beyond the L2 classroom.

## 3. Pronunciation Training Techniques and Teaching Methods in FL Learning Contexts

### *3.1. Phonetic Training*

High-variability phonetic training (HVPT) is typically a perceptual training paradigm where learners are exposed to L2 sounds produced in a variety of phonetic environments by multiple speakers and need to identify or discriminate the target sounds after receiving individualized feedback [25]. HVPT has been investigated in relation to different segmental and suprasegmental features of L2 speech (e.g., vowels, consonants, syllable structure, tone) and has been found effective in developing L2 speech categories [24,25], leading to gains in L2 speech perception [33], lexical encoding [34] and production [35]. Importantly, phonetic and phonological learning from HVPT has been shown to generalize to novel talkers [36], untrained testing stimuli [37], new phonetic contexts [25], untrained sounds [33], and across perception and production modalities [38]. Additionally, learning gains tend to be retained sometime after the training has ended [18,33,38]. Training outcomes may differ in size as a function of presentation mode and stimulus type. For example, gains in production are greater if learners receive visual articulatory feedback than auditory-only feedback [35], if trained in adverse rather than silent conditions [39], or if trained with non-lexical rather than lexical stimuli [40,41].

Apart from HVPT, other explicit training paradigms, such as phonological specificity training, a paradigm that trains learners on minimal pairs to enhance the distinctiveness of phonolexical representations [42] and training in auditory processing skills [43], have also been shown to impact L2 phonological learning positively. Last, incidental multimodal phonetic training (e.g., playing with a mobile game application that exposes gamers to target sound contrasts when aliens are killed) has been found to be helpful for learners to automate the knowledge of L2 sounds that may have been learned explicitly [17]. Despite the multiple benefits of phonetic training for L2 speech development, one of the potential drawbacks of phonetic training is its limited ecological validity [6]. In other words, HVPT practice, unlike communicative types of practice (e.g., giving directions on a map where identifying street names depends on learners' accurate perception and production of a sound contrast), is often pedagogically decontextualized and can be disassociated with language learning and use in real-life contexts, and, hence, FL teachers may be hesitant to implement methods that have not been proven in real classrooms [44]. Still, empirically validated HVPT computer-assisted systems such as the Golden Speaker [45] or the English

Accent Coach [46] may provide accessible individualized pronunciation learning inside or outside the classroom.

### 3.2. Explicit Instruction

Explicit pronunciation instruction entails providing L2 learners with metalinguistic information about the voicing, place, and manner of articulation of L2 speech sounds and the acoustic as well as prosodic (stress, rhythm, and intonation) characteristics of L2 speech. Several meta-analyses [19,21,22] and individual studies [47,48] have shown the effectiveness of explicit instruction at improving L2 pronunciation and at making L2 speech more intelligible, comprehensible, and less strongly accented in classroom contexts. For example, Ref. [49] examined the impact of short-term explicit pronunciation teaching on both suprasegmental (stress, rhythm, reductions, and linking) and segmental features (/i/, /ɪ/, /æ/, /ɛ/) of L2 learners' speech. The intervention involved three 25 min sessions of explicit pronunciation instruction for three weeks. Compared to a non-explicit pronunciation instruction intervention, explicit phonetic instruction led to improvements in comprehensibility for learners trained on suprasegmental features but not for learners trained on segmentals. Other studies [50] have found little improvement resulting from explicit pronunciation instruction (around 5%) and no improvement for accentedness and comprehensibility.

While explicit pronunciation instruction helps learners notice L1–L2 phonological differences [6], it often involves a decontextualized focus on the accuracy of specific phonological forms, relying mainly on controlled practices. In fact, there is ongoing debate about whether gains obtained from explicit pronunciation instruction can be effectively maintained when using the L2 in real-life conversations. Given the prevailing emphasis on grammar-focused lessons and teachers' limited understanding of which aspects of pronunciation should be taught, researchers have a responsibility to inform teachers and teacher trainers about the key aspects of L2 pronunciation that should be given priority [51,52] and how to integrate them into content-based lessons [3]. One promising approach is engaging learners in interactive tasks that enhance learners' awareness of the communicative impact of pronunciation (e.g., an interaction map task; see Section 4). This way, learners can naturally focus on phonetic form while conveying meaning during communication.

### 3.3. Form-Focused Communicative Instruction

L2 pronunciation research [53–55] has demonstrated the superiority of explicit instruction combined with communicative form-focused instruction (FFI) over explicit instruction alone. FFI entails drawing learners' attention to form in communicative contexts, that is, practicing L2 pronunciation while being engaged in contextualized meaning-oriented communicative activities [20]. Recently, Ref. [56] compared explicit-only pronunciation instruction, consisting of listening and controlled practice of L2 suprasegmental features, to explicit instruction + FFI, which combined explicit teaching of pronunciation features repeatedly with communicative instruction following the Communicative Framework [57] and Automatization in Communicative Contexts of Essential Speech Segments framework (ACCESS) [58]. The results of this study showed that the "explicit-only" learners only improved L2 comprehensibility in controlled tasks, whereas the "explicit + FFI" learners improved both in controlled and, especially, in spontaneous tasks. These findings align with [20]'s synthesis of 15 quasi-experimental studies where FFI was found to contribute to the development of L2 speech in controlled and spontaneous speaking tasks, whereas the benefits of explicit-only instruction were only observable in controlled speech.

Given that the key to L2 phonological learning in the FL classroom is the automaticity of L2 phonological and phonetic processing and generalization from in-class to out-of-class language use, Refs. [3,53] suggest that learners can establish form-meaning mappings and develop L2 accuracy and fluency [59] by using activities that are intently repetitive yet have communicative value and by integrating a focus-on-phonetic-form into meaning-oriented tasks. Implementing a dual focus on form and meaning following the Communicative

Framework [57], ACCESS [58,59] or Strategy-based [60] frameworks has the potential to allow learners to notice and pay attention to L2 pronunciation features and to develop awareness of their own pronunciation problems (e.g., providing corrective feedback). The well-attested positive impact of explicit instruction in L2 phonology [22] can be maximized if it is extended to communicative language use contexts by gradually allowing learners to automate the procedural phonological knowledge they have acquired through form-focused activities when using L2 in contexts where they are primarily attending to meaning. This may in turn facilitate the spreading of L2-specific phonetic features (e.g., the aspiration of /p, t, k/ or discrimination of segmental contrasts /i:-ɪ/, /æ-ʌ/ in L2-English) to the entire lexicon, enhancing phonological acquisition and effectively improving L2 pronunciation while speaking the L2.

While most research has explored the benefits of integrating pronunciation in a communicative task after receiving explicit pronunciation instruction [10,61], practicing L2 pronunciation incidentally during communicative interaction following a TBLT approach [12–14,62] has been considerably less investigated.

### *3.4. New Avenues in L2 Pronunciation Training and Teaching*

Current pedagogical practices in L2 pronunciation teaching and learning do not fully reflect the recent shift in the pedagogic target of L2 pronunciation learning from native-like speech to comprehensible speech. One way to promote the adjustment of pronunciation teaching to this paradigm shift is to make the outcomes of current research on the effectiveness of speech awareness-raising tasks to develop global dimensions of L2 pronunciation proficiency (intelligibility, comprehensibility, accentedness, and fluency) available to the pronunciation instruction community. Empirical research investigating the effectiveness of training tasks to develop L2 speech intelligibility, comprehensibility, accentedness, and fluency globally is scarce and varied in methods, in the level of learners' proficiency, and in its outcomes.

This section outlines a number of pronunciation training techniques whose effectiveness in raising awareness and developing L2 pronunciation has been experimentally proven: accent imitation, multimodal pronunciation training through captioned video, embodied pronunciation training, comprehensibility and accentedness self-assessment, and TBPT.

#### 3.4.1. Accent Imitation

Research on foreign accent imitation training and its benefits for L2 pronunciation development is currently scarce, but findings so far [63,64] support the notion that training learners in imitating an L2 accent (e.g., an English accent) on their native language (L1) is helpful in developing awareness of L1–L2 cross-language phonetic differences and in enhancing the automatization of L2-specific articulatory gestures, leading to improvement of pronunciation accuracy at the segmental level, at least for low-proficiency learners [63]. In foreign accent imitation tasks, learners are asked to speak their L1 (or to produce word- or sentence-long utterances in their L1) with an L2 accent so that measures of phonetic features obtained from the imitated L2 accent (e.g., voice onset time—VOT) may be interpreted as a measure of implicit awareness (or implicit knowledge) of the phonetic properties of the L2 being imitated [65,66]. Learners have been found to imitate an L2 phonetic feature in their L1, such as VOT duration in voiceless oral stops, to the extent that they can produce them accurately in their L2 [67–69]. This technique has been applied to assess the production accuracy of L2-specific segmental phonetic features (mainly VOT) in a delayed accent-imitation paradigm. To the best of our knowledge, Ref. [63] is the only study that has used this technique to train L2-specific phonetic features in low-proficiency young learners. Given appropriate use of imitation training materials such as extended texts and spoken dialogues, accent imitation can be effective in training advanced adult learners' L2 phonetic features, including segmental and suprasegmental properties that may impact global dimensions of L2 speech.

3.4.2. Multimodal Pronunciation Training through Captioned Video

L2-captioned video (intralingual subtitles) can be effectively used to train L2 learners' simultaneous processing of L2 auditory input (speech), orthographic input (dynamic onscreen text), and visual input (onscreen dynamic images). Research has proven the benefits of this type of enriched input for listening comprehension, the incidental acquisition of vocabulary, and grammar. The presence of written word forms as learners process the spoken input has been shown to enhance auditory word recognition and speech segmentation skills [70,71] and can therefore offer interesting pedagogical possibilities for pronunciation instruction if learners' attention is guided to phonetic form while watching. For example, Ref. [72] enhanced learners' attention to phonetic form through pronunciation-related questions popping up occasionally on the screen while learners watched captioned video and found the treatment to improve L2 learners' speech segmentation and speech processing skills. Ref. [73] used audio-synchronized textual input enhancement in captions to promote learners' visual processing of usually mispronounced orthographic word forms immediately before they could be heard in the soundtrack and found benefits in learners' ability to recognize mispronounced forms, supporting the updating of non-target-like phonological representations of words.

3.4.3. Embodied Pronunciation Training

Embodied pronunciation training is based on the notion of multimodal enrichment, which holds that exposure to complementary information across multiple sensory modalities during learning activities can enhance learning benefits [74]. This is a type of multimodal phonetic training that takes advantage of the mutual effects of auditory perception and visual actions on one another to enhance the acquisition of segmental and suprasegmental features of speech [75]. For example, Ref. [76] showed that a group of L2-French learners assigned to an embodied pronunciation training condition (visuospatial hand gestures depicting speech rhythm and intonation during the oral repetition of CV syllables) improved their accentedness and suprasegmental features of L2 French in a dialogue read-aloud task significantly more than a comparable group assigned to a speech-only training condition. Similarly, recent research has also shown that phonetic training using visuospatial hand gestures, such as a fist-to-open-hand burst gesture to visually represent the auditory and articulatory features of Mandarin Chinese aspirated plosives [77], enhances phonetic learning, leading to more accurate production of aspirated stops.

3.4.4. Comprehensibility and Accentedness Self-Assessment

Self-assessment has mainly focused on identifying differences between L2 learners' assessments of their speech and assessments by native speakers (or L2 teachers) focusing primarily on accentedness and comprehensibility [78], but L2-speech self-assessment and peer-assessment tasks can be useful in raising learners' awareness of pronunciation features that make their speech difficult to understand or strongly accented [79,80]. Although L2 learners' ability to assess their own speaking performance is related to their actual speaking performance, the better their speaking skills, the more likely they are to accurately self-assess their performance [81]. Accurate speaking self-assessments indicate a lack of awareness and limitations in noticing the pronunciation features that affect their speech intelligibility, comprehensibility and accentedness [80] and the speech of others, which could have a negative impact on L2 speech development [82].

Research focusing specifically on the benefits of L2 pronunciation self-assessment for L2 pronunciation development [83,84] has generally found it difficult for learners to focus on specific phonological features and shows a mismatch between learners' self-assessments and assessments by native listeners [80,85] so that learners judged to perform well by native listeners tend to underestimate their performance, whereas learners judged to perform poorly tend to overestimate their performance. Overestimation and underestimation in speech-self assessment are indications of learners' difficulties in identifying the underlying pronunciation features that influence their speech, which can be improved with increased

learners' experience in self-assessment [86] and with training in speech-self-assessment [87]. Methods for improving self-assessment skills include discussing learners' own performance, familiarizing learners with rating criteria, using self-testing exercises, benchmarking, and peer assessment. For example, Ref. [88] implemented a treatment based on benchmarking (asking learners to discuss speech evaluation criteria) and peer assessment (evaluating the speaking performance of peers) and found learners increased the alignment between their ratings and those of native listeners.

#### 3.4.5. Task-Based Pronunciation Teaching (TBPT)

An approach to pronunciation teaching that can be integrated into communicative language teaching is task-based pronunciation teaching, or TBPT [11–13]. TBPT makes use of the task-design principles of TBLT to enhance learners' attention to phonetic form in pedagogic tasks that involve meaning-based interaction. TBLT adopts meaning-based communicative tasks as central to defining language learning needs, goals, classroom activities, and assessment, but has not devoted research efforts to investigating pronunciation-focused communicative tasks until recently [89]. Based on [90]'s Framework for Task-Based Learning, a commonly used TBPT task-design implementation procedure [12–14] involves three stages, namely, pre-task, task cycle, and post-task. During the pre-task, the teacher assists L2 learners in learning and recollecting the linguistic resources they will need to perform the interactive task. Learners focus both on phonetic form and meaning through comprehension activities and plan their speech before engaging in the main task. During the meaning-based interactive task, learners apply the language they have encountered in the preceding phase to carry out their interactions. Pronunciation targets (e.g., difficult vowel contrasts) are used incidentally but can be made essential for task completion, thus forcing learners to use the target sounds appropriately to perform the task. Immediately after, learners are asked to plan and report back on how they completed the task. This phase involves a transition from using spontaneous language with an emphasis on fluency to employing planned language that prioritizes fluency, accuracy, and clarity in organizing their public discourse. The last phase consists of orienting learners' attention to relevant pronunciation aspects that naturally arise during the conversations they engage in throughout the task cycle. The objective of the metalinguistic and communicative post-tasks is to help learners consolidate what they have learned and generalize their L2 knowledge to new contexts of L2 use.

TBLT research has shown that, by manipulating communicative task design variables (e.g., repetition, modality, or complexity), it is possible to enhance learners' attention to linguistic form during communicative interaction, helping learners develop their lexical, grammatical, and pragmatic L2 performance. For example, according to the Cognition Hypothesis [91], making a communicative task cognitively more complex by increasing the number of elements and reasoning demands needed to complete the task leads to the production of more lexically and grammatically complex language, thus implicitly enhancing L2 development. Recent research [89] indicates that manipulating task-design variables is effective at enhancing learners' attention to phonetic form in meaning-based tasks, resulting in improvements in the perception and production of L2 segmental [12,13,15] as well as suprasegmental features of L2 speech [92], and speech comprehensibility [10]. As learners tend to engage in such interactive tasks collaboratively, task performance often leads to the occurrence of pronunciation-focused language-related episodes (P-LRE) [93] that serve to raise learners' awareness of pronunciation issues and indicate the extent to which the task design was effective at helping learners focus on pronunciation. In addition, target pronunciation features can be made task-essential [94], forcing learners to pay attention to and use phonological targets to complete the task while focusing on meaning, such as when giving directions on a collaborative map task performed in pairs where the street names contain contrastive sounds that must be distinguished perceptually and in production for the learners to give and understand directions successfully [15,62].

Having reviewed novel methodologies of L2 pronunciation instruction in lab and classroom-based settings, we now illustrate the benefits of one of them, communicative TBPT, through an empirical lab study. TBPT allows teachers to fully integrate pronunciation instruction into communicative language teaching. In this lab-based study, we assess the efficacy of an interactive map task that could effectively be used by teachers in a classroom context to improve the perception and production of difficult L2 sound contrasts.

## 4. A TBPT Empirical Study

Although the benefits of task complexity for linguistic development are well attested in the domain of grammar and lexis [89], such benefits have only begun to be explored for L2 pronunciation. The few studies available to date [10,12–15] suggest that task complexity may be effective in enhancing attention to phonetic form during communicative pronunciation-focused tasks. As a means of illustrating methodological issues related to integrating a focus on phonetic form within communicative tasks, we report on an empirical study that aimed at applying TBPT principles to the design of a computerized map task. This task, which aimed at enhancing accuracy in the perception and production of a difficult L2 vowel contrast for L1-Spanish learners of English, was designed as a pronunciation-focused, meaning-oriented interactive pedagogic task. The design and preliminary outcomes of this study demonstrate the potential of TBPT for L2 pronunciation learning. Alternative interactive TBPT tasks to the ones described here can be found in [95,96].

### *4.1. Materials and Methods*

In the current study, Ref. [90]'s TBLT task-cycle design was followed in that the experimental pedagogic intervention included a pre-task intended to familiarize learners with the phonetic targets to be improved, as well as an interactive communicative task designed around [97]'s definition of task: the primary focus of the task was on meaning; there was some kind of communicative gap; learners had to rely on their linguistic and non-linguistic resources to solve the task; and language was the means to achieve a clearly defined outcome. The task was framed as a two-way, close, convergent, giving-directions task [98] because it was performed in pairs; there was only one possible solution that students had to agree on, and students took it in turns to give and follow directions on a map. Unlike [90]'s framework, for research purposes, the current implementation of the task design lacked a post-task activity usually included in pedagogic tasks to consolidate the knowledge acquired through task performance.

This study followed a pre-test > intervention > post-test design. The pre-test and post-test included an ABX discrimination test to assess the effect of the intervention on L2 learners' accuracy and speed of response in perceptually distinguishing /i:/ from /ɪ/ (e.g., *feet-fit*), as well as delayed nonword repetition (DNWR) and delayed sentence repetition (DSR) tests to assess the effect of the intervention on L2 learners' accuracy in distinguishing /i:/ from /ɪ/ in production.

The intervention consisted of three 30 min computerized tasks in two sessions. Each task consisted of a perception pre-task, a production pre-task, and the corresponding interactive map task. The perception pre-task consisted of an identification task where learners were presented with the nonwords to be used in the corresponding map task and were asked to identify the stressed vowel in them as either English /i:/ or /ɪ/. The production pre-task consisted of an immediate repetition task where learners were presented with the nonwords to be used in the corresponding map task and were asked to repeat them as accurately as they could. Pre-tasks were aimed at familiarizing learners with the auditory and orthographic forms of the street names (English nonwords) they would practice during that map task session (18 minimal-pair nonword pairs). The map tasks were sequenced in order of increasing task complexity (simple > + complex > ++ complex). Learners needed to be able to distinguish /i:/ from /ɪ/ in perception and production to be able to follow

and give instructions on the map and thus complete the task successfully. Map tasks 1 and 2 were performed in session 1, and map task 3 in session 2.

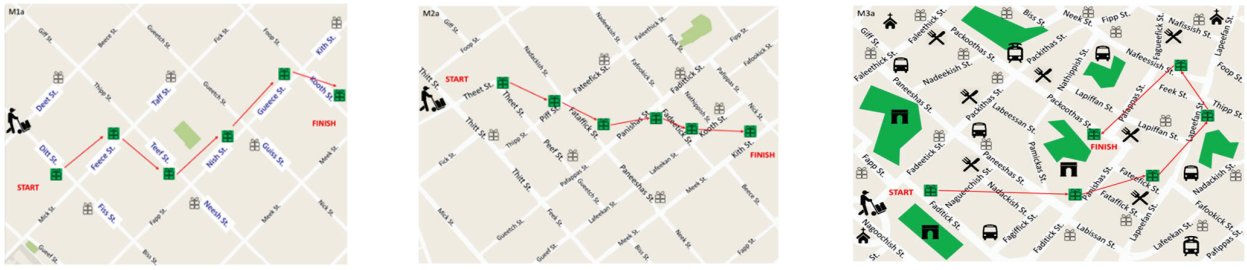
Given the exploratory nature of the current study and the relatively large inter-subject variability in L2 pronunciation proficiency of the participants, they were further randomly assigned to three task difficulty sub-groups. Task difficulty (low, medium, high) was operationalized in terms of how easy or difficult the target contrast was to perceive and produce in the street names based on whether the contrast was embedded in monosyllabic nonwords (easy: peef /pi:f/ vs. piff /pɪf/), trisyllabic nonwords (medium: lapeefan /lə'pi:fən/ vs. lapiffan /lə'pɪfən/), and a mixture of monosyllabic and trisyllabic nonwords (difficult), which might make the target contrast effectively harder for learners to attend to. Perceiving and producing the target vowels /i:/ and /ɪ/ accurately in a trisyllabic nonword was expected to pose greater difficulty and require greater attentional effort for learners than doing so in monosyllabic nonwords. In preliminary analyses of the participants' pre-test and post-test perception scores according to task stimuli difficulty [62], all sub-groups were found to improve sensitivity to the contrast after the intervention, but gains between testing times were smaller and did not reach significance for those participants assigned to the mixed stimuli difficulty condition, suggesting that stimuli variability limited the extent to which learners could benefit from the TBPT intervention and made the task more demanding for learners. As our aim here is to assess the effectiveness of the TBPT map task intervention, we focus on the overall results obtained by all the participants performing the map tasks as the experimental group.

#### 4.1.1. Participants

The participants were 77 upper-intermediate to advanced L1-Spanish learners of English (*Mage* = 20.88, *SD* = 4.30). They had learned English through formal instruction in a school context since around the age of 6 (*SD* = 3.62), and at the time of testing, they were first-year students enrolled in a degree in English studies at a Spanish university. On a 9-point Likert scale (1 = very poor command of English, 9 = near-native command of English), participants rated themselves as fairly proficient (*M* = 6.67, *SD* = 1.17). They were randomly assigned to either an experimental group (*N* = 62) or a control group (*N* = 15) that did the pre-test and post-test but did not perform any of the familiarization pre-tasks or the intervention map tasks.

#### 4.1.2. Map Task Design

Following the SSARC model of task sequencing [99], we designed three map tasks that differed in cognitive task complexity (simple, + complex, ++ complex) in terms of the number elements (streets and crossroads) on the map (see Figure 1) and asked learners to perform them in order of increasing complexity in two intervention sessions.




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

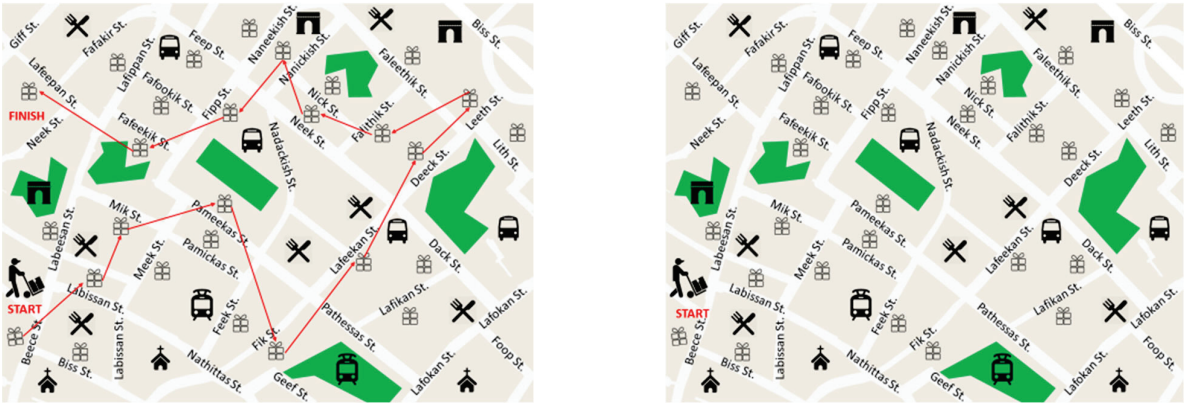
**Figure 1.** Simple (**a**), + Complex (**b**), and ++ Complex (**c**) map tasks.

The pronunciation target was the English vowel contrast /i:/-/ɪ/ (e.g., feet-fit), which Spanish learners of English find difficult to perceive and produce accurately because both English /i:/ and /ɪ/ are perceptually mapped onto the single Spanish front vowel category /i/ [100]. Although the L2 learners that participated in this study were relatively

advanced, the target contrast was embedded in English nonwords and was deemed to pose difficulties in perception and production. The reason why we opted for using nonwords rather than words as stimuli for the street names on the map was twofold. First, using nonwords would avoid learners activating lexical representations for words they might be misrepresenting phonologically in their mental lexicon, which could possibly lead to a less effective intervention, as found in HVPT paradigms [40,41]. Secondly, using nonwords for which lexical representations cannot be activated could be helpful in helping learners concentrate on phonetic form (rather than meaning) when giving and following instructions on the map.

The learners' task was to give (Student A) and follow (Student B) directions using the non-word street names to pick 14 parcels located at streets off crossroads on the map. At each crossroad, a decision had to be made as to which street to take (e.g., lapeefan vs. lapiffan) where a parcel had to be picked. This was conducted in order to make the target vowel contrast task essential [94], forcing learners to focus on the qualitative differences between the contrastive vowels /i:/ and /ɪ/ in production and perception. The map task was collaborative in that each parcel to be picked involved producing and perceiving the contrast accurately by each learner dyad. Getting the wrong parcel would involve negotiating interactively to make the contrast clear to one another.

Student A (giving directions) and Student B (following directions) would be seated in front of each other, and each one of them would be in front of two monitors. Student A's monitor 1 would show a red line indicating the directions to be given according to the path Student B would need to follow to pick the 14 parcels (see Figure 2), whereas monitor 2 would show the same as students' B monitor 2 (the same map without the path). The red path was not visible for Student B, for whom only monitor 2 would be on. This design allowed Student A to monitor at all times what Student B's mouse pointer was doing. In order to pick up a parcel, Student B would need to double-click on it once its location on the map was reached according to the instructions. When clicking on it, the parcel would turn green if the correct street had been taken at the crossroads or red if the wrong street had been taken. The correct (and wrong) streets targeted an equal number of /i:/ and /ɪ/ nonwords. Taking the wrong street implied either Student A not pronouncing the street name correctly or Student B not perceiving the street name correctly, or both, which generated a number of P-LRE as students went back to the crossroads and tried to give and follow directions again to make up for the wrong choice of street. In order for learners to be able to monitor their production accuracy, all of the street names were clickable. When clicking on a street name, the learners would hear the pre-recorded nonword pronounced by either a male or a female native speaker of English. When getting to the end of the path and having picked the last parcel, students A and B changed roles in giving and following directions.




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

**Figure 2.** Student A's monitor 1 (**a**) and monitor 2 (**b**) for the complex map task.

Having correctly picked 14 parcels implies having correctly produced and identified the contrastive vowels in 12 minimal-pair nonwords per map task (12 × 3 map tasks = 36 minimal-pair nonword pairs). The nonword street names were different in versions A and B of each map task, so students A and B had to produce and identify different sets of street names. Each map task also included two control vowel contrasts that posed no difficulty to learners (e.g., /i:/-/æ/, /ɪ/-/æ/, /i:/-/u:/, /ɪ/-/u:/). Students performed the map tasks in pairs while wearing Beyerdynamic DT 990 PRO open headphones, over which they could hear the names of the streets (whenever they clicked on them) and their partners' voices. All of the map tasks were audio-recorded (separately for each speaker) through Shure SM-58 voice microphones onto Marantz PMD660 solid-state digital recorders (44.1 kHz, 16-bit).

#### 4.1.3. Testing and Analyses

In the ABX discrimination test, participants had to decide, as fast and accurately as possible, for every A-B-X trial, whether X contained the same vowel as A or the same vowel as B (e.g., lapeefan-lapiffan-lapiffan = B; lapeefan-lapiffan-lapeefan = A). ABX trials were presented with an inter-stimulus interval of 500 ms and 2000 ms after each response, a new trial was presented, or 2500 ms after the onset of the last item in the trial if no response was provided. In each ABX trial, A and B were spoken in the same voice (either male or female) and X in a different voice. This provided a measure of learners' ability to identify the target vowel across two different voices, an indication of having developed relatively robust, distinct sound categories for the target vowels /i:/ and /ɪ/. The test contained 96 ABX trials, corresponding to 16 test trials and 8 control trials in 4 orders (ABB, ABA, BAA, and BAB). The 96 trials contained an equal number of trained nonwords (nonwords included in the map tasks) and untrained nonwords (new nonwords produced by new voices), an equal number of monosyllabic and trisyllabic nonwords, and an equal number of nonwords produced by a male and a female voice. Accuracy rates and reaction times (RT) in correctly identifying X in ABX trials were used as a measure of learners' sensitivity to the target contrast in perception.

For production, we used DNWR and DSR tests. In the DNWR test, participants were asked to repeat, after a 1500 ms delay followed by a 250 ms beep signal, 64 test nonwords and 32 control nonwords. The 64 test nonwords corresponded to the nonwords in the 16 test trials included in the ABX test (16 /i:/ nonwords + 16 /ɪ/ nonwords = 32 nonwords) presented auditorily in fully randomized order and repeated twice, once after a male voice and once after a female voice. As in the ABX test, the test nonwords included a balanced design in terms of trained and untrained nonwords, monosyllabic and trisyllabic nonwords, and male and female voices. In the DSR test, participants were presented with a set of 8 mini-dialogues involving short prompt-response interactions between two different voices they listened to (e.g., Speaker A: Shall I put the heating on?; Speaker B: Yes, my feet are cold) and were then asked to repeat the response (Speaker A: Shall I put the heating on?; Participant: Yes, my feet are cold). The 8 responses targeted 8 common English words (4 with /i:/ and 4 with /ɪ/: sheep-ship, feet-fit, sit-seat, chips-cheap). Assessing learners' pronunciation of real words (apart from untrained nonwords) allowed us to determine whether the intervention map tasks could be effective at improving the phonological representations of lexical items.

The analysis of the production data involved computing average acoustic measures per vowel (/i:/ and /ɪ/) for each learner to assess changes in vowel quality (degree of vowel height and frontness) in the direction of the four native speakers' vowel productions used as testing stimuli in the ABX and the DNWR tests. Based on the *f*0, F1, and F2 formant frequency measures in Hertz which we extracted from the midpoint of the steady portion of each vowel, we computed a normalized Bark-converted (B) distance metric where B1-B0 represented a normalized height measure and B2-B1 a normalized frontness measure. We also assessed changes in the extent to which learners could make a qualitative distinction between the contrasting vowels in production by computing spectral distance

scores (Euclidean distances) between /i:/ and /ɪ/ calculated on a two-dimensional height (B1-B0) by frontness (B2-B1) space. The larger the spectral distance score (SDS) between /i:/ and /ɪ/, the more distinct the production of the contrasting vowels was. We therefore expected SDSs to be larger at the post-test than at the pre-test.

*4.2. Results*

Having first checked that participants performed at much higher accuracy rates when discriminating control trials (/i:/-/æ/, /ɪ/-/æ/, /i:/-/u:/, /ɪ/-/u:/) than test trials (/i:/-/ɪ/) in the ABX test at pre-test ($M$ = 0.87, $SD$ = 0.332 vs. $M$ = 0.68, $SD$ = 0.468, respectively), we explored the effects of the map task intervention and generalization effects to untrained nonwords based on test nonwords only. The results show that accuracy scores between testing times were slightly larger for the experimental group than the control group and that participants seem to improve similarly between testing times on trained and untrained items (Figure 3). The RT data also showed slightly larger differences in speed between testing times for the experimental group than the control group (Figure 4). However, unexpectedly, the control group obtained considerable perception gains, which can only be explained by a task effect, that is, improvement associated with repeating the perception test twice. In addition, both the experimental and control groups seemed to perform on trained and untrained nonwords equally well, suggesting that improvements in perception could not be attributed to testing trials having been previously trained.

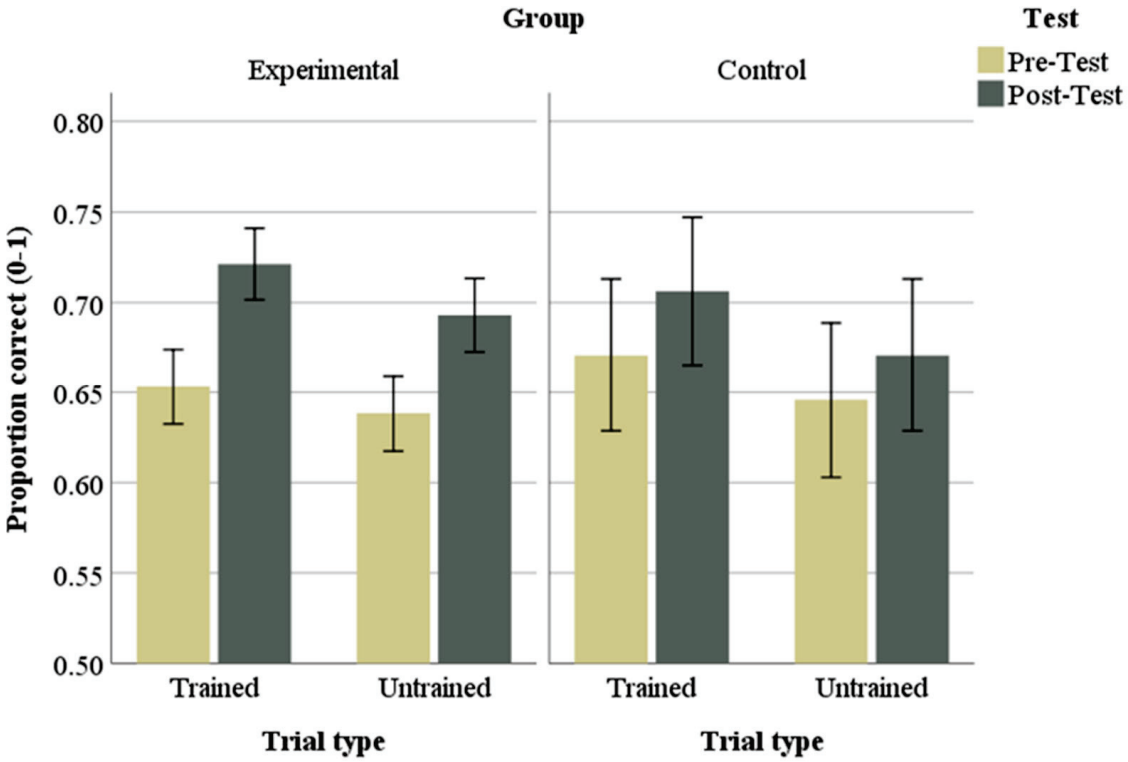



**Figure 3.** ABX accuracy by group, test and trial type (Error bars = 95% CI).

We first assessed whether learning gains could generalize to untrained nonwords by submitting participants' ABX responses to test items (correct or wrong) to a linear mixed-effects model with a binomial logistic regression (in SPSS 25) with test (pre-test, post-test), group (experimental, control), and trial type (trained, untrained) and their interactions as fixed effects, including a random intercept for subject. These analyses showed significant main effects of test ($F$(1, 9912) = 19.75, $p$ < 0.001) and trial type ($F$(1, 9912) = 4.98, $p$ = 0.026), but neither the main effect of group ($F$(1, 9912) = 0.001, $p$ = 0.977) nor any of the interactions reached significance (see Appendix A for the model's parameter estimates). Tests of pairwise contrasts showed that while at pre-test the experimental group

performed equally well on trained and untrained items ($t(9912) = 1.01$, $SE = 0.015$, $p = 0.316$), at post-test they discriminated the target vowels at significantly higher correct rates in trained than in untrained nonwords ($t(9912) = 1.98$, $SE = 0.014$, $p = 0.048$). This indicated, as expected, that training had a larger effect on trained than untrained nonwords, although both improved significantly between testing times. Interestingly, whereas the experimental group discriminated the target vowels significantly more accurately at post-test than at pre-test in both trained ($t(9912) = -5.01$, $SE = 0.015$, $p < 0.001$) and untrained ($t(9912) = -5.01$, $SE = 0.015$, $p < 0.001$) nonwords, the control group did not significantly improve on either trained ($t(9912) = -1.21$, $SE = 0.030$, $p = 0.229$) or untrained ($t(9912) = -0.026$, $SE = 0.031$, $p = 0.406$) nonwords. These findings indicate that for the experimental group (but not for the control group), discrimination accuracy improved for both trained and untrained nonwords, showing that training gains generalized to untrained nonwords. Consequently, in all subsequent analyses, accuracy and RT scores from test items (including trained and untrained nonwords) were used as dependent measures.

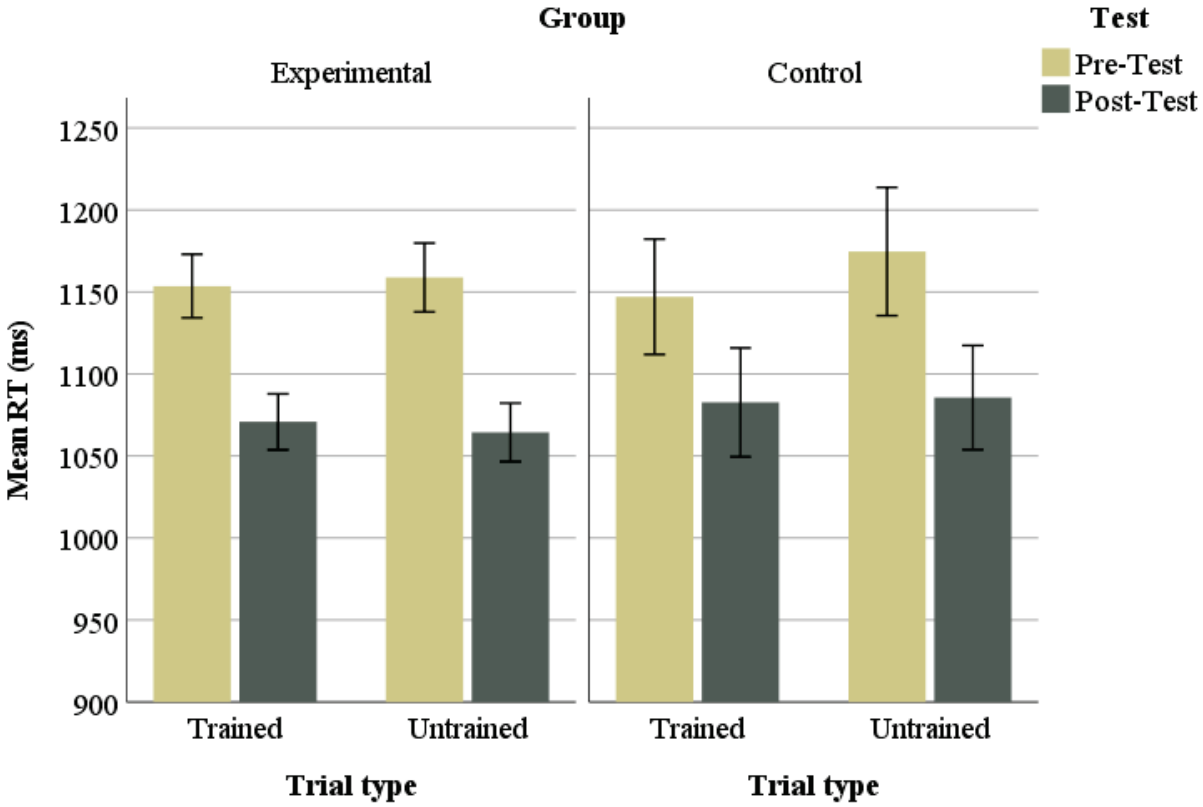



**Figure 4.** ABX response speed by group, test and trial type (Error bars = 95% CI).

In order to assess whether the intervention effects on perception reached significance for the experimental and control groups, we first submitted participants' ABX responses to test items (correct or wrong) to a linear mixed-effects model with a binomial logistic regression with test (pre-test, post-test) and group (experimental, control) and their interaction as fixed effects and random intercepts for subject and item. These analyses showed a significant main effect of test ($F(1, 9912) = 19.75$, $p < 0.001$), but neither the main effect of group ($F(1, 9912) = 0.001$, $p = 0.978$), nor the test $\times$ group interaction ($F(1, 9912) = 2.75$, $p = 0.097$) reached significance (see Appendix A for the model's parameter estimates). According to Bonferroni-adjusted pairwise contrasts, the main effect of the *test* was driven by the gains obtained by the experimental group, who gained a modest but significant 7.6% in accuracy ($t(9916) = -6.52$, $SE = 0.012$, $p < 0.001$), whereas the gains by the control group (3.5%) did not reach significance ($t(9916) = -1.54$, $SE = 0.023$, $p = 0.123$). We then submitted participants' ABX RT on test items, including only those for correct responses and excluding those beyond 2.5 standard deviations from each subject's mean (2.05%)

to a linear mixed-effects model with test (pre-test, post-test) and group (experimental, control) as fixed effects and random intercepts for subject and item. The outcome of these analyses was similar to those we obtained for accuracy: there was a significant main effect of test ($F(1, 6551) = 111.4$, $p < 0.001$), but neither the main effect of group ($F(1, 6551) = 0.152$, $p = 0.697$), nor the test × group interaction ($F(1, 6551) = 0.459$, $p = 0.498$) reached significance (see Appendix A). However, in this case, both the experimental ($t(6551) = 12.74$, $SE = 7.47$, $p < 0.001$) and control ($t(6551) = 5.50$, $SE = 15.22$, $p < 0.001$) groups significantly improved in RT at post-test, with the experimental group only improving 12 ms on average more than the control group (95.2 ms vs. 83.7 ms).

In terms of production accuracy in the DNWR test, improvement between testing times in vowel quality was small and mainly affected /ɪ/ (as English /i:/ is already acoustically very similar to Spanish /i/). Learners' /ɪ/ became lower in height and slightly more centralized, whereas /i:/ became slightly more target-like only in height. However, spectral distances appear to become larger at post-test, suggesting that learners had improved in making a distinction between /i:/ and /ɪ/ in production, although the spectral distance produced between vowels was much smaller than the one the native speakers produced (see Table 1). In order to assess the effectiveness of the map-task intervention in effecting improvement in learners' ability to distinguish /i:/ from /ɪ/ in production, we ran a Paired-samples T-test on the learners' pre-test and post-test spectral distance scores, which confirmed that spectral distances were of a significantly larger magnitude at post-test than at pre-test ($t(61) = -2.59$, $p = 0.012$).

**Table 1.** Normalized Bark distance metrics and spectral distances for natives ($N = 4$) and Learners ($N = 63$) in the DNWR and DSR tests.

| Group | Test | /i:/ | | | | /ɪ/ | | | | /i/-/ɪ/ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B1-B0 (Height) | | B2-B1 (Frontness) | | B1-B0 (Height) | | B2-B1 (Frontness) | | Euclidean Distance | | | |
| | | M | SD | M | SD | M | SD | M | SD | M | SD | Min. | Max. |
| Natives | | 1.96 | 0.24 | 10.87 | 0.55 | 2.98 | 0.42 | 8.65 | 0.85 | 2.45 | 0.79 | 1.85 | 3.52 |
| Learners (DNWR) | Pre-test | 1.76 | 0.53 | 10.50 | 0.84 | 2.48 | 0.44 | 9.32 | 0.64 | 1.58 | 0.60 | 0.59 | 3.37 |
| | Post-test | 1.71 | 0.51 | 10.65 | 0.76 | 2.53 | 0.45 | 9.26 | 0.67 | 1.75 | 0.66 | 0.69 | 3.72 |
| Learners (DSR) | Pre-test | 1.62 | 0.53 | 10.68 | 0.74 | 2.11 | 0.48 | 9.76 | 0.77 | 1.13 | 1.0 | 0.13 | 4.82 |
| | Post-test | 1.53 | 0.56 | 10.73 | 0.80 | 2.06 | 0.49 | 9.82 | 0.78 | 1.14 | 1.0 | 0.07 | 4.50 |

We also measured the vowel productions from the DSR test in the same way as those from the DNWR test and computed spectral distance scores between /i:/ and /ɪ/ to assess improvement in learners' ability to distinguish /i:/ from /ɪ/ in minimal-pair words they knew. Spectral distances were on average much smaller between /i:/ and /ɪ/ on these words than they were on the nonwords from the DNWR test (see Table 1), suggesting that learners made less of a distinction between /i:/ and /ɪ/ in real words embedded in sentences produced from memory than in the production of isolated nonwords. Improvement between pre-test and post-test spectral distances was very small (see Table 1) and non-significant ($t(61) = -0.117$, $p = 0.907$), suggesting that the positive effect of the map-task intervention on production did not generalize to untrained words (sheep-ship, feet-fit, sit-seat, chips-cheap) whose phonological representations were already well-established.

### *4.3. Discussion*

Overall, the effects of the computerized map-task intervention, albeit small, suggest that a two-session TBPT communicative interaction helped learners improve their ability to distinguish a difficult L2 vowel contrast (/i:/-/ɪ/) in perception and production.

On the one hand, the perception results suggest that performing three oral map tasks of increasing cognitive complexity resulted in an improvement in the discrimination accuracy of the target vowels /i:/ and /ɪ/. These findings echo those observed in studies that carried

out implicit high-variability perceptual training of segmentals [17] or those applying form-focused communicative interventions [61]. The fact that discrimination gains were lower in this study than in others obtained through explicit HVPT [33,101] might be due to the short two-session (three map tasks) intervention. Still, generalization to untrained items, in line with other studies [33,37,41] signals robust improvement in L2 vowel discrimination.

On the other hand, the production results show that learners produced larger distances (i.e., less overlap) between contrastive vowels after the TBPT intervention. In accordance with [102] and [15]'s findings, reflection on phonetic form during communicative tasks that make pronunciation targets task-essential and necessitate agreement on a single correct outcome [97] leads to more distinct realizations of L2 confusable sounds, hence, more target-like productions. Our findings for production are in accordance with production gains reported in other short communicative form-focused instruction interventions [55]. Nevertheless, learners were unable to generalize vowel distinctiveness gains to untrained word contexts, contrary to [12,13], which involved learners' performance on many more lexically-based tasks during a longer intervention. This could be attributed to the short two-session intervention as well as to the size of the small data set the DSR test generated (vowel productions from 8 /i:/-/ɪ/ minimal-pair words).

Finally, a number of important limitations need to be considered. First, the lack of L2 production data from the control group suggests the outcome of the analyses in terms of L2 production gains obtained by the experimental group resulting from the TBPT intervention cannot be ascertained and needs to be interpreted with caution. Second, a two-session intervention may not be long enough for learners to develop detectable gains in how distinctly they can produce L2 vowels. Last, the fact that the battery of perceptual and production tasks used at pre-test and post-test to gauge learning gains were very different from the map tasks used in the intervention may have made it difficult for us to observe L2 pronunciation learning gains that might otherwise have shown in an interactive testing task (a map task) more similar to the intervention tasks. Further data analyses on the frequency of P-LRE during learners' interaction and of the performance on the test map task would likely provide further insights into the effectiveness of TBPT for improving L2 pronunciation.

## 5. Conclusions

The current article has outlined and discussed current issues in L2 pronunciation instruction research and practice in ISLA arising from the need to overcome the limitations and challenges FL learning contexts pose to L2 learners' pronunciation development. These include, but are not limited to, learners' scarce L2 exposure and use, the difficulty in applying well-established findings within current L2 speech learning models (SLM-r, PAM-L2) to the FL classroom, the need to seek novel methods and techniques to train L2 pronunciation globally, and the difficulty of integrating a focus on phonetic form in meaning-based tasks in a primarily grammar-centered communicative approach to language teaching. We have highlighted key features of L2 speech models that can inform pronunciation training and instruction and synthesized commonly used as well as novel training and teaching pronunciation instruction methods. Finally, we have exemplified TBPT as a pronunciation instruction method by reporting on the results of an empirical study that used communicative map tasks to teach segmental pronunciation targets.

As a means to incorporate research findings from speech learning models into pronunciation instruction, we suggest that both the design of pronunciation-focused communicative tasks (e.g., TBPT) and the design of individualized pronunciation training techniques (e.g., accent imitation) consider two key aspects of L2 speech learning: namely, (a) that the difficulties in the acquisition of L2 segmental phonology (e.g., L2 sound contrasts) are based on specific L2-to-L1 perceptual sound mappings, and (b) that the target phonetic and phonological features to be acquired are likely to vary in how essential they are to L2 speech intelligibility. Therefore, HVPT paradigms should consider training learners not only on the identification and discrimination of L2 sound contrasts but also on the perception of the

cross-language differences between the segmental phonologies of the L1 and the L2 that determine which L2 sound contrasts are difficult to acquire, as this would help learners overcome L1-based perception. Similarly, pronunciation-focused communicative tasks could be designed to make pronunciation targets (especially those having a larger impact on speech intelligibility) essential to task resolution to enhance attention to phonetic form during communicative interaction.

Most of the pronunciation training and teaching methods outlined in this article exemplify novel creative ways of overcoming the difficulties associated with learning pronunciation in ISLA, but their effectiveness needs further exploration and empirical support. Evaluating their impact on L2 pronunciation development requires establishing consistent measurement standards, which are currently too varied and inconsistent across studies. Research assessing the effectiveness of pronunciation training techniques for L2 pronunciation development makes use of either fine-grained acoustic analyses of target phonetic features (e.g., [12,13] as well as the HVPT and accent imitation studies we have reported on) or perceptual judgments of global dimensions of L2 speech such as intelligibility, comprehensibility, accentedness, and fluency (e.g., [54] and the embodied pronunciation training studies we have reported on). However, the multidimensional nature of L2 speech and the current trend to define the functional load of pronunciation targets in terms of speech intelligibility and comprehensibility [51,52] suggest that pronunciation assessment, especially when aiming at determining the effectiveness of pronunciation training and teaching methods, should be carried out through a combination of acoustic measurements and listeners' judgments of global dimensions of L2 pronunciation proficiency [21].

The TBPT empirical study we report on, together with a few other similar studies [10,12], illustrates a novel approach that incorporates the research findings of L2 speech learning models and TBLT into pronunciation instruction while effectively integrating pronunciation instruction into communicative classroom teaching in ISLA. Although further research is needed to confirm the pedagogic value of TBPT, current research findings already offer preliminary evidence of its effectiveness for L2 pronunciation learning. In addition to integrating pronunciation instruction into communicative language teaching, pronunciation development in instructed SLA needs to combine classroom pedagogical practice with individualized pronunciation training that can provide personalized feedback [48]. A pronunciation teaching approach that combines in-class TBPT tasks with out-of-class pronunciation training through tasks such as those described above may provide a very effective way to learn L2 pronunciation, especially if combined in a structured way, so that individualized pronunciation training tasks serve to raise awareness of challenging L2 phonological features that will be at a later stage practiced communicatively through TBPT tasks in the classroom. Assessing the effectiveness of such a combined approach to L2 pronunciation learning opens exciting research avenues in L2 pronunciation instruction.

**Author Contributions:** Conceptualization, J.C.M. and I.M.-P.; methodology, J.C.M.; formal analysis, J.C.M. and I.M.-P.; investigation, J.C.M.; data curation, J.C.M. and I.M.-P.; writing—original draft preparation, J.C.M. and I.M.-P.; writing—review and editing, J.C.M. and I.M.-P.; project administration, J.C.M.; funding acquisition, J.C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Ministry of Science & Innovation, grant number PID2019-107814GB-I00, and the Catalan Department of Research and Universities (AGAUR), grant number 2021SGR00303.

**Institutional Review Board Statement:** The study protocol adhered to the good practices of data collection, data anonymization, data processing, and data storage of the Institutional Review Board of the University of Barcelona (IRB0003099) and in accordance with the Declaration of Helsinki.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in this 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 part of the data being still under analysis.

**Acknowledgments:** We would like to thank Mayya Levkina for giving us permission to use for the current study part of the data collected and analyzed in Mora, J. C. & Levkina, M. (2018). Training vowel perception through map tasks: The role of linguistic and cognitive complexity. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 151–162). Ames, IA: Iowa State University.

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

## Appendix A

**Table A1.** Parameter estimates of linear mixed-effects models on ABX discrimination scores.

| | | β | SE | t | p | 95% CI | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower | Upper |
| Model 1 [1] | Intercept | 0.74 | 0.15 | 5.03 | <0.001 | 0.45 | 1.03 |
| | Group | 0.10 | 0.16 | 0.64 | 0.526 | −0.22 | 0.43 |
| | Test | −0.12 | 0.14 | −0.83 | 0.406 | −0.39 | 0.16 |
| | Item type | 0.17 | 0.14 | 1.20 | 0.228 | −0.11 | 0.45 |
| | Group × Test | −0.16 | 0.15 | −1.04 | 0.299 | −0.46 | 0.14 |
| | Group × Trial type | −0.03 | 0.16 | −0.20 | 0.845 | −0.34 | 0.28 |
| | Test × Trial type | −0.06 | 0.20 | −0.28 | 0.779 | −0.45 | 0.33 |
| | Group × Test × Trial type | −0.02 | 0.22 | −0.08 | 0.938 | −0.45 | 0.42 |
| Model 2 [2] | Intercept | 0.93 | 0.18 | 5.04 | <0.001 | 0.57 | 1.29 |
| | Group | 0.10 | 0.17 | 0.62 | 0.535 | −0.22 | 0.43 |
| | Test | −0.17 | 0.11 | −1.55 | 0.121 | −0.38 | 0.04 |
| | Group × Test | −0.20 | 0.12 | −1.66 | 0.097 | −0.43 | 0.04 |
| Model 3 [3] | Intercept | 111.48 | 52.28 | 21.24 | <0.001 | 1007.99 | 1212.98 |
| | Group | −27.59 | 56.70 | −0.49 | 0.627 | −138.73 | 83.56 |
| | Test | 83.73 | 15.22 | 5.50 | <0.001 | 53.90 | 113.56 |
| | Group × Test | 11.49 | 16.95 | 0.68 | 0.498 | −21.74 | 44.72 |

[1] ABX accuracy (generalization effects), [2] ABX accuracy (test effects), [3] ABX RT (test effects).

## References

1. Sardegna, V.G.; Lee, J.; Kusey, C. Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Lang. Learn.* **2018**, *68*, 83–114. [CrossRef]
2. Baker, A. Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *Tesol Q.* **2014**, *48*, 136–163. [CrossRef]
3. Darcy, I. Powerful and Effective Pronunciation Instruction: How Can We Achieve It? *Catesol J.* **2018**, *30*, 13–45.
4. Tyler, M.D. PAM-L2 and phonological category acquisition in the foreign language classroom. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*; Nyvad, A.M., Hejná, M., Højen, A., Jespersen, A.B., Hjortshøj Sørensen, M., Eds.; Department of English, School of Communication & Culture, Aarhus University: Aarhus, Denmark, 2019; pp. 607–630.
5. Huensch, A. Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Lang. Teach. Res.* **2019**, *23*, 745–764. [CrossRef]
6. Derwing, T.M.; Munro, M.J. *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*; John Benjamins: Amsterdam, The Netherlands, 2015.
7. Derwing, T.M.; Munro, M.J.; Thomson, R.I. (Eds.) *The Routledge Handbook of Second Language Acquisition and Speaking*; Routledge: New York, NY, USA, 2022.
8. Kang, O.; Thomson, R.; Murphy, J. *The Routledge Handbook of Contemporary English Pronunciation*; Routledge: New York, NY, USA, 2018.
9. Wayland, R. *Second Language Speech Learning*; Cambridge University Press: New York, NY, USA, 2021.
10. Gordon, J. Pronunciation and Task-Based Instruction: Effects of a Classroom Intervention. *Relc. J.* **2021**, *52*, 94–109. [CrossRef]
11. Mora, J.C.; Levkina, M. Task-based pronunciation teaching and research: Key issues and future directions. *SSLA* **2017**, *39*, 381–399. [CrossRef]
12. Mora-Plaza, I. Task-Based Pronunciation Teaching and Learning of L2 Vowels in EFL Learners: Task Complexity Effects. Ph.D. Thesis, Universitat de Barcelona, Barcelona, Spain, June 2023.

13. Mora-Plaza, I. Task-based pronunciation teaching helps improve L2 vowel production: Generalization effects. In Proceedings of the 7th English Pronunciation: Issues & Practices Conference, Grenoble, France, 18–20 May 2022; Henderson, A., Kirkova-Naskova, A., Eds.; Université Grenoble-Alpes: Grenoble, France, 2023; pp. 172–187.
14. Mora-Plaza, I.; Mora, J.C.; Gilabert, R. Focus on phonetic form and cognitive complexity in the acquisition of an L2 vowel contrast. In Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference, Salt Lake City, UT, USA, 1–2 September 2017; Levis, J.M., Ed.; Iowa State University: Ames, IA, USA, 2018; pp. 174–184.
15. Solon, M.; Long, A.Y.; Gurzynski-Weiss, L. Task Complexity, Language-Related Episodes, and Production of L2 Spanish Vowels. *SSLA* **2017**, *39*, 347–380. [CrossRef]
16. Fouz-González, J. Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Lang. Learn. Technol.* **2020**, *24*, 62–85.
17. Saito, K.; Hanzawa, K.; Petrova, K.; Kachlicka, M.; Suzukida, Y.; Tierney, A. Incidental and Multimodal High Variability Phonetic Training: Potential, Limits, and Future Directions. *Lang. Learn.* **2022**, *72*, 1049–1091. [CrossRef]
18. Wang, X.; Munro, M.J. Computer-based training for learning English vowel contrasts. *System* **2004**, *32*, 539–552. [CrossRef]
19. Lee, J.; Jang, J.; Plonsky, L. The effectiveness of second language pronunciation instruction: A meta-analysis. *Appl. Linguist.* **2015**, *36*, 345–366. [CrossRef]
20. Saito, K. Effects of Instruction on L2 Pronunciation Development: A Synthesis of 15 Quasi-Experimental Intervention Studies. *Tesol Q.* **2012**, *46*, 842–854. [CrossRef]
21. Saito, K.; Plonsky, L. Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Lang. Learn.* **2019**, *69*, 652–708. [CrossRef]
22. Thomson, R.I.; Derwing, T.M. The effectiveness of L2 pronunciation instruction: A narrative review. *Appl. Linguist.* **2015**, *36*, 326–344. [CrossRef]
23. Barriuso, T.A.; Hayes-Harb, R. High Variability Phonetic Training as a Bridge from Research to Practice. *Catesol J.* **2018**, *30*, 177–194.
24. Sakai, M.; Moorman, C. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Appl. Psycholinguist.* **2018**, *39*, 187–224. [CrossRef]
25. Thomson, R.I. High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *JSLP* **2018**, *4*, 208–231. [CrossRef]
26. Levis, J.M. Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Q.* **2005**, *39*, 369–377. [CrossRef]
27. Levis, J.M. Revisiting the intelligibility and nativeness principles. *JSLP* **2020**, *6*, 310–328. [CrossRef]
28. Best, C.; Tyler, M. Non-native and second language speech perception. In *Language Experience in Second Language Speech Learning*; Bohn, O.-S., Munro, M.J., Eds.; John Benjamins: Amsterdam, The Netherlands, 2007; pp. 15–34.
29. Flege, J.E.; Bohn, O.-S. The revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*; Wayland, R., Ed.; Cambridge University Press: Cambridge, UK, 2021; pp. 3–83.
30. Muñoz, C. Contrasting effects of starting age and input on the oral performance of foreign language learners. *Appl. Linguist.* **2014**, *35*, 463–482. [CrossRef]
31. Tyler, M.D.; Best, C.T.; Faber, A.; Levitt, A.G. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* **2014**, *71*, 4–21. [CrossRef]
32. Piske, T. Implications of James E. Flege's research for the foreign language classroom. In *Language Experience in Second Language Speech Learning*; Bohn, O.-S., Munro, M.J., Eds.; John Benjamins: Amsterdam, The Netherlands, 2007; pp. 301–314.
33. Carlet, A.; Cebrian, J. Assessing the Effect of Perceptual Training on L2 Vowel Identification, Generalization and Long-term Effects. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*; Nyvad, A.M., Hejná, M., Højen, A., Jespersen, A.B., Hjortshøj Sørensen, M., Eds.; Department of English, School of Communication & Culture, Aarhus University: Aarhus, Denmark, 2019; pp. 91–119.
34. Melnik-Leroy, G.A.; Peperkamp, S. High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Biling. Lang. Cogn.* **2021**, *24*, 497–506. [CrossRef]
35. Kartushina, N.; Hervais-Adelman, A.; Frauenfelder, U.H.; Golestani, N. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *JASA* **2015**, *138*, 817–832. [CrossRef] [PubMed]
36. Lee, D.Y.; Baese-Berk, M.M. Non-native English listeners' adaptation to native English speakers. *JASA Express Lett.* **2021**, *1*, 105201. [CrossRef] [PubMed]
37. Mora, J.C.; Mora-Plaza, I. Contributions of cognitive attention control to L2 speech learning. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*; Nyvad, A.M., Hejná, M., Højen, A., Jespersen, A.B., Hjortshøj Sørensen, M., Eds.; Department of English, School of Communication & Culture, Aarhus University: Aarhus, Denmark, 2019; pp. 477–499.
38. Rato, A.; Rauber, A. The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK, 10–14 August 2015.
39. Cooke, M.; García Lecumberri, M.L. Effects of exposure to noise during perceptual training of non-native language sounds. *JASA* **2018**, *143*, 2602–2610. [CrossRef] [PubMed]
40. Mora, J.C.; Ortega, M.; Mora-Plaza, I.; Aliaga-García, C. Improving the pronunciation of L2 vowels under different training conditions: The use of non-lexical materials and masking noise. *Phonetica.* **2022**, *79*, 1–43. [CrossRef] [PubMed]

41. Ortega, M.; Mora-Plaza, I.; Mora, J.C. Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In *English Pronunciation Instruction: Research-Based Insights*; Kirkova-Naskova, A., Henderson, A., Fouz-González, J., Eds.; John Benjamins: Amsterdam, The Netherlands, 2021; pp. 327–356.
42. Llompart, M.; Reinisch, E. Lexical representations can rapidly be updated in the early stages of second-language word learning. *J. Phon.* **2021**, *88*, 101080. [CrossRef]
43. Saito, K.; Petrova, K.; Suzukida, Y.; Kachlicka, M.; Tierney, A. Training auditory processing promotes second language speech acquisition. *JEPHPP* **2022**, *48*, 1410–1426. [CrossRef] [PubMed]
44. Martin, I.A.; Inceoglu, S. The laboratory, the classroom, and online: What works in each context. In *Second Language Pronunciation: Bridging the Gap between Research and Teaching*; Levis, J.M., Derwing, T.M., Sonsaat-Hegelheimer, S., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2022; pp. 254–272.
45. Ding, S.; Liberatore, C.; Sonsaat, S.; Lučić, I.; Silpachai, A.; Zhao, G.; Chukharev-Hudilainen, E.; Levis, J.; Gutierrez-Osuna, R. Golden speaker builder–An interactive tool for pronunciation training. *Speech Commun.* **2019**, *115*, 51–66. [CrossRef]
46. Thomson, R.I. Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Lang. Learn.* **2012**, *62*, 1231–1258. [CrossRef]
47. Gordon, J.; Darcy, I. Teaching segmentals and suprasegmentals: Effects of explicit pronunciation instruction on comprehensibility, fluency, and accentedness. *JSLP* **2022**, *8*, 168–195. [CrossRef]
48. Saito, K.; Lyster, R. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Lang. Learn.* **2012**, *62*, 595–633. [CrossRef]
49. Gordon, J.; Darcy, I. The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *JSLP* **2016**, *2*, 56–92. [CrossRef]
50. Kissling, E.M. Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *Mod. Lang.* **2013**, *97*, 720–744. [CrossRef]
51. Munro, M.J.; Derwing, T.M. The functional load principle in ESL pronunciation instruction: An exploratory study. *System* **2006**, *34*, 520–531. [CrossRef]
52. Suzukida, Y.; Saito, K. Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Lang. Teach. Res.* **2021**, *25*, 431–450. [CrossRef]
53. Darcy, I.; Hancock, Z.; Lee, J.; Rocca, B. Improving pronunciation in spontaneous speech? A comparison of two instructional methods. In Proceedings of the L2 Pronunciation Research Workshop: Bridging the Gap between Research and Practice, Barcelona, Spain, 6–7 June 2019.
54. Darcy, I.; Rocca, B.; Hancock, Z. A window into the classroom: How teachers integrate pronunciation instruction. *RELC J.* **2021**, *52*, 110–127. [CrossRef]
55. Saito, K. Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Appl. Psycholinguist.* **2015**, *36*, 377–409. [CrossRef]
56. Darcy, I.; Rocca, B. Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. *J. Second. Lang. Pronunciation* **2023**, *8*, 328–362. [CrossRef]
57. Celce-Murcia, M.; Brinton, D.; Goodwin, J. *Teaching Pronunciation: A Reference Book for Teachers*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2010.
58. Gatbonton, E.; Segalowitz, N. Rethinking communicative language teaching: A focus on access to fluency. *Can. Mod. Lang. Rev.* **2005**, *61*, 325–353. [CrossRef]
59. Trofimovich, P.; Gatbonton, E. Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Mod. Lang. J.* **2006**, *90*, 519–535. [CrossRef]
60. Sardegna, V.G. Evidence in favor of a strategy-based model for English pronunciation instruction. *Lang. Teach.* **2022**, *55*, 363–378. [CrossRef]
61. Ruan, Y.; Saito, K. Communicative focus on phonetic form revisited: Less precise auditory processing limits instructed L2 speech learning. *System* **2023**, *114*, 103020. [CrossRef]
62. Mora, J.C.; Levkina, M. Training vowel perception through map tasks: The role of linguistic and cognitive complexity. In Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference, Salt Lake City, UT, USA, 1–2 September 2017; Levis, J.M., Ed.; Iowa State University: Ames, IA, USA, 2018; pp. 151–162.
63. Everitt, C. Accent Imitation on the L1 as a Task to Improve L2 Pronunciation. Master's Thesis, Universitat de Barcelona, Barcelona, Spain, June 2015.
64. Ruellot, V. French stereotypical accent and pronunciation development of /p/,/t/, and /k/. In Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, Ames, IA, USA, 6–8 September 2018; Levis, J.M., Nagle, C., Todey, E., Eds.; Iowa State University: Ames, IA, USA, 2019; pp. 341–353.
65. Kopečková, R.; Wrembel, M.; Gut, U.; Balas, A. Differences in phonological awareness of young L3 learners: An accent mimicry study. *Int. J. Multiling.* **2021**, *20*, 408–424. [CrossRef]
66. Mora, J.C.; Rochdi, Y.; Kivistö-de-Souza, H. Mimicking accented speech as L2 phonological awareness. *Lang. Aware.* **2014**, *23*, 57–75. [CrossRef]
67. Flege, J.E.; Hammond, R. Mimicry of non-distinctive phonetic differences between language varieties. *SSLA* **1982**, *5*, 1–17. [CrossRef]

68. Henderson, A.; Rojczyk, A. Foreign-language accent imitation: Matching production with perception. In *English Pronunciation Teaching: Theory, Practice, and Research Findings*; Sardegna, V., Jarosz, A., Eds.; Multilingual Matters: Bristol, UK, 2023; pp. 102–117.
69. Rojczyk, A. Using FL accent imitation in L1 in foreign-language speech research. In *Teaching and Researching the Pronunciation of English*; Waniek-Klimczak, E., Pawlak, M., Eds.; Springer: Cham, Switzerland, 2015; pp. 223–233.
70. Charles, T.; Trenkic, D. The effect of bi-modal input presentation on second language listening, The focus on speech segmentation. In *Subtitles and Language Learning: Principles, Strategies and Practical Experiences*; Gambier, Y., Caimi, A., Mariotti, C., Eds.; Peter Lang: Bern, Sweden, 2015; pp. 173–198.
71. Mitterer, H.; McQueen, J.M. Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* **2009**, *4*, e7785. [CrossRef] [PubMed]
72. Wisniewska, N.; Mora, J.C. Can captioned video benefit second language pronunciation? *SSLA* **2020**, *42*, 599–624. [CrossRef]
73. Galimberti, V.; Mora, J.C.; Gilabert, R. Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System* **2023**, *116*, 103078. [CrossRef]
74. Mathias, B.; von Kriegstein, K. Enriched learning: Behavior, brain, and computation. *Trends Cogn. Sci.* **2023**, *27*, 81–97. [CrossRef]
75. Chan, M.J. Embodied pronunciation learning: Research and practice. *Catesol J.* **2018**, *30*, 47–68.
76. Baills, F.; Alazard, C.; Prieto, P. Embodied prosodic training helps improve accentedness and suprasegmental accuracy. *Appl. Linguist.* **2022**, *43*, 776–804. [CrossRef]
77. Li, P.; Xi, X.; Baills, F.; Prieto, P. Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Lang. Cogn. Neurosci.* **2021**, *36*, 1313–1328. [CrossRef]
78. Isaacs, T.; Harding, L. Pronunciation assessment. *Lang. Teach.* **2017**, *50*, 347–366. [CrossRef]
79. Isbell, D.R.; Lee, J. Self-Assessment of Comprehensibility and Accentedness in Second Language Korean. *Lang. Learn.* **2022**, *72*, 806–852. [CrossRef]
80. Strachan, L.; Kennedy, S.; Trofimovich, P. Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment. *JSLP* **2019**, *5*, 347–373. [CrossRef]
81. Li, M.; Zhang, X. A meta-analysis of self-assessment and language performance in language testing and assessment. *Lang. Test.* **2021**, *38*, 189–218. [CrossRef]
82. O'Brien, M.G. Attending to second language lexical stress: Exploring the roles of metalinguistic awareness and self-assessment. *Lang. Aware.* **2019**, *28*, 310–328. [CrossRef]
83. Dlaska, A.; Krekeler, C. Self-assessment of pronunciation. *System* **2008**, *36*, 506–516. [CrossRef]
84. Isbell, D.R. Can the test support student learning? Validating the use of a second language pronunciation diagnostic. *Lang. Assess. Q.* **2021**, *18*, 331–356. [CrossRef]
85. Trofimovich, P.; Isaacs, T.; Kennedy, S.; Saito, K.; Crowther, D. Flawed self-assessment: Investigating self-and other-perception of second language speech. *Biling. Lang. Cogn.* **2016**, *19*, 122–140. [CrossRef]
86. Ma, W.; Winke, P. Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Lang. Ann.* **2019**, *52*, 66–86. [CrossRef]
87. Kissling, E.M.; O'Donnell, M.E. Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Lang. Aware.* **2015**, *11*, 283–302. [CrossRef]
88. Tsunemoto, A.; Trofimovich, P.; Blanchet, J.; Bertrand, J.; Kennedy, S. Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Lang. Ann.* **2022**, *55*, 135–154. [CrossRef]
89. Gurzynski-Weiss, L.; Long, A.; Solon, M. TBLT and L2 Pronunciation: Do the Benefits of Tasks Extend Beyond Grammar and Lexis? [Special issue]. *SSLA* **2017**, *39*, 213–224. [CrossRef]
90. Willis, J. *A Framework for Task-Based Learning*; Longman: Harlow, UK, 1996.
91. Robinson, P. Second language task complexity, the cognition hypothesis, language learning, and performance. In *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*; Robinson, P., Ed.; John Benjamins: Amsterdam, The Netherlands, 2011; pp. 3–37.
92. McKinnon, S. TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *SSLA* **2017**, *39*, 287–317. [CrossRef]
93. Loewen, S.; Isbell, D. Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *SSLA* **2017**, *39*, 225–256. [CrossRef]
94. Loschky, L.; Bley-Vroman, R. Grammar and task-based methodology. In *Tasks and Language Learning: Integrating Theory and Practice*; Crookes, G., Gass, S., Eds.; Multilingual Matters: Clevedon, UK, 1993; pp. 123–167.
95. Gurzynski-Weiss, L.; IATBLT. The TBLT Language Learning Task Bank. Available online: https://tblt.indiana.edu (accessed on 1 April 2023).
96. Mora-Plaza, I.; Saito, K.; Suzukida, Y.; Dewaele, J.-M.; Tierney, A. Tools for Second Language Speech Research and Teaching. Available online: http://sla-speech-tools.com (accessed on 15 December 2022).
97. Ellis, R. *Task-Based Language Learning and Teaching*; Oxford University Press: Oxford, UK, 2003.
98. Pica, T.; Kanagy, R.; Falodun, J. Choosing and Using Communicative Tasks for Second Language Instruction. In *Tasks and Language Learning: Integrating Theory and Practice*; Crookes, G., Gass, S., Eds.; Multilingual Matters: Clevedon, UK, 1993; pp. 9–34.

99. Robinson, P. Situation and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In *Cognitive Processing in Second Language Acquisition: Inside the Learner's Mind*; Putz, M., Sicola, L., Eds.; John Benjamins: Amsterdam, The Netherlands, 2010; pp. 243–268.
100. Cebrian, J. Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *J. Acoust. Soc. Am.* **2019**, *145*, EL52–EL58. [CrossRef]
101. Rato, A. Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Copal* **2014**, *5*, 529–546.
102. Sicola, L. "No, They won't 'Just Sound Like Each Other'": NNS-NNS Negotiated Interaction and Attention to Phonological Form on Targeted L2 Pronunciation Tasks. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, January 2008.

education sciences

MDPI

*Article*

# The Effects of Orthography on the Pronunciation of Nasal Vowels by L1 Japanese Learners of L3 French: Evidence from a Longitudinal Study of Speech in Interaction

**Cyrille Granget [1,*,†], Cecilia Gunnarsson [1,†], Inès Saddour [1,†], Clara Solier [2,†], Vera Serrau [1] and Charlotte Alazard [1]**

[1] Laboratory of NeuroPsychoLinguistics, University of Toulouse, 31000 Toulouse, France; cecilia.gunnarsson@univ-tlse2.fr (C.G.); ines.saddour@univ-tlse2.fr (I.S.); vera.serrau@univ-tlse2.fr (V.S.); charlotte.alazard@univ-tlse2.fr (C.A.)
[2] Basque Center on Cognition, Brain and Language, 20009 Donostia-San Sebastián, Spain; c.solier@bcbl.eu
* Correspondence: cyrille.granget@univ-tlse2.fr
† These authors contributed equally to this work.

**Abstract:** In recent decades, a vast literature has documented crosslinguistic influences on the acquisition of L2 phonology and in particular the effects of spelling on pronunciation. However, articulating these research findings in terms of taking into account the effects of L1 phonology and spelling on L2 pronunciation in language teaching remains to be examined. These studies are based on experimental cross-sectional methods and mainly focus on L2 English learning by speakers of languages with an alphabetic system. In French, there are few studies on crosslinguistic influences on the acquisition of the nasal vowels (/ɑ̃/, /ɔ̃/ and /ɛ̃/) and few experimental studies that point to a possible effect of orthography on the pronunciation of these phonemes. The results of experimental studies are difficult to transpose to the language classroom because they are based on word or sentence reading and writing activities, which are quite far-removed from the conversational activities practised in the classroom in interaction with peers and the teacher. Hence, we opted here for a case study of the effect of spelling on the production of nasal vowels in interaction tasks. We conducted a longitudinal study during the first year of extensive learning of French (4 h 30 per week). The results of a perceptive analysis by expert listeners show that (i) learners spell nasal vowels with an <n> or <m> in 98% of the obligatory contexts; (ii) most nasal vowels are perceived as nasal vowels in speech (72%), the others being perceived as vowels followed by a nasal consonant (19.5%) or as oral vowels (8.5%); (iii) consonantisation is stronger when the learner spontaneously produces a word than when (s)he repeats it, (iv) which decreases with time (learning effect) and varies (v) according to the consonant, /ɛ̃/ being less consonantised than /ɔ̃/ and /ɑ̃/. Finaly, we propose a didactic discussion in the light of intelligibility and influence of orthography.

**Keywords:** nasal vowels; L3 French; L1 Japanese; orthographic effect; crosslinguistic influence; longitudinal; spoken and written corpus

**Citation:** Granget, C.; Gunnarsson, C.; Saddour, I.; Solier, C.; Serrau, V.; Alazard, C. The Effects of Orthography on the Pronunciation of Nasal Vowels by L1 Japanese Learners of L3 French: Evidence from a Longitudinal Study of Speech in Interaction. *Educ. Sci.* **2024**, *14*, 234. https://doi.org/10.3390/educsci14030234

Academic Editors: Lawrence Jun Zhang and Martin Howard

Received: 9 June 2023
Revised: 30 January 2024
Accepted: 18 February 2024
Published: 23 February 2024



## 1. Introduction

In second language acquisition, many studies focus on the development of oral or written language. However, if we are interested in acquisition in an instructed environment where the learner is doubly exposed to oral and written language, often from the initial stages, it seems essential to take into account the interaction between oral and written representations of the language. Following from the many studies on the effect of orthography on pronunciation [1–3], this study focuses, in particular, on the effects of the graphemes <n> and <m> on the acquisition of nasal vowels.

The phonological units /sɔ̃/, /ɑ̃kɔʁ/ and /pɛ̃/ have in common that they are words containing a nasal vowel: /ɔ̃, ɑ̃ and ɛ̃/, respectively. Even if the French phonological

inventory is sometimes described as containing four nasal vowels (/ɑ̃/, /ɔ̃/, /ɛ̃/ and /œ̃/), most researchers agree, as Marquez Martinez (2016) points out, that /ɛ̃/ and /œ̃/ have been merged in favour of /ɛ̃/ in Northern Metropolitan French, the variety one usually describes and teaches as standard [4]. Therefore, we only examine these three nasal vowels: /ɑ̃/, /ɔ̃/ and /ɛ̃/, whose spellings includes the grapheme <n> <son> 'his/her/its', <encore> 'still/again', <pain> 'bread'.Since French nasal vowels appear in very frequent words and are distinct phonemes from oral vowels and nasalised allophones found in other languages [4–8], their pronunciation is central in the development and use of an intelligible lexicon. Using one nasal vowel instead of another, an oral vowel instead of a nasal vowel or a vowel followed by a nasal consonant instead of a nasal vowel can lead to a change in meaning: if /ɛ̃/ is pronounced /ɔ̃/ in /pɛ̃/, the meaning of the word is 'bridge' instead of 'bread'; if /ɑ̃/ is pronounced /a/ in /ɑ̃kɔʁ/, then the meaning of the word is 'agreement' instead of 'still/again'; and finally, if /sɔ̃/ 'his/her' is pronounced with a consonantised vowel /sɔn/, then the possessive determiner becomes a verb form meaning 'sound(s)'.

Phonological and phonetic accuracy thus has an impact on lexical acquisition in French as an additional language, and it is therefore relevant to know how nasal vowels in L3 French are perceived by expert listeners. In the light of studies conducted over the last 20 years, French orthography and, in particular, the nasal graphemes <n> and <m> play an important role in the acquisitional process of nasal vowels, as well as in the production process of nasal vowels in which phonological and orthographic representations are involved. Nevertheless, the graphemes <n> and <m> could play contradictory and potentially evolving roles in the acquisitional process: on the one hand, they inscribe in the orthography of the language the nasal feature inherent in the nasal vowel and could, as such, encourage the use of a nasal vowel; on the other hand, they materialise this nasality by means of the grapheme also used to write the nasal consonant /n/ and could, as such, explain the phenomenon of consonantisation of the vowel attested in several studies [4,5,9–11].

However, these effects of 'externalisation of the nasality of the nasal consonant' have been considered by most studies of nasal vowel acquisition from the angle of the influence of the phonology of first languages [4,5,11]. They explain the production of a postvocalic nasal in L2 French under the influence of first languages without nasal vowels but with nasalised vowels in perception or production. We take up the term "L2" used by the authors. In all the studies mentioned above, it refers to a language learned after the first language. In the studies by Montagu (2002) and Martinez-Marquez (2016), French is chronologically the second language learned, but in the case of Detey et al. (2010) and Li et al. (2019), it is probably the third language after English.Three studies point to a possible effect of orthography [8,9,12]. Only one study investigates the specific effect of a postvocalic nasal consonant [12].

The aim of this study is not to decide between these two explanations but to study, in an ecological context of interactive speech, the specific and evolving effect of the graphemes <n> and <m> on the use of nasal vowels by beginner Japanese-speaking learners of L3 French doubly exposed in class to spoken and written French in textbooks, on the blackboard, in exercise books and in personal notebooks.

To do this, we conducted a case study and used a bimodal (oral and written) and longitudinal corpus of oral productions in L3 French from four adult learners of L1 Japanese and L2 English collected at two points during their first year of extensive learning of French at university. We analysed the way in which the phonemes produced by the Japanese-speaking learners in place of nasal vowels are perceived by expert listeners. In particular, we analysed the phenomenon of consonantisation of the nasal vowel and its variation according to three variables: on the one hand, the production context, distinguishing between contexts frequent at the beginning stage of repetition of the interlocutor's word and contexts of spontaneous production and recovery of a lexical unit in the mental lexicon; on the other hand, the type of vowel and the greater or lesser number of graphemic equivalents; and finally, the time of learning, i.e., after 50 h and after 120 h of bimodal

exposure to L3 French. Finally, we supplemented this study of the pronunciation of nasal vowels in L3 French with an analysis of orthographic uses in order to examine whether the orthographic representation of the oral forms of words with a nasal vowel includes the graphemes <n> and <m>.

The results are discussed from the perspective of teaching intelligible speech in an additional language.

## 2. The Role of L1 Phonology and L2 Orthography in the Acquisition of Nasal Vowels in French as an Additional Language

The acquisition of nasal vowels in L2 French has attracted the attention of researchers for several reasons. On the one hand, nasal vowels are attested in only 22.6% of the world's languages. Therefore, they are relatively rare phonemes compared to nasal consonants, which exist in almost all languages, or compared to nasalised vowels, which are present in a large number of languages, particularly in the investigated first languages of adult learners producing nasal vowels in L3 French [4,5,9,11]. Therefore, the acquisition of nasal vowels in adulthood represents an interesting case of the acquisition of second language phonology and of interphonology [9,13]. On the other hand, nasal vowels have the particularity of being spelled using a set of graphemes that vary from word to word but whose invariants are the letters <n>, as in <bo*n*> 'good', <da*n*s> 'in' and <pai*n*> 'bread', or <m> when the letter following the nasal vowel is <p>, <b> or <m>, as in <co*m*prendre> 'understand' or <ja*m*be> 'leg'. In theory, these letters are likely to fix the distinctive nasal feature of the nasal vowel, as they can also promote the production of a postvocalic nasal consonant. In this section, we present the few studies that have focused on the acquisition or use of nasal vowels in French as a second language, a summary of the more numerous studies that have focused on the role of orthography in the acquisition of L2 new phonemes and some proposals for pronunciation teaching of new phonemes taking into account the effect of orthography.

### *2.1. The Acquisition of Nasal Vowels in L2/L3 French*

In several studies mentioned in this section of the state of the art, the learning situations are designed from the point of view of the acquisition of an L2 by learners of a given L1. However, these learners are often multilingual and have learned other languages, such as Mandarin in the study by Li, Yin and Pu (2019) [11] or English in the study conducted by Detey et al. (2010) [9]. Even if the influence on these languages in the acquisition of French is not investigated as such, we can consider that French is an L3. This is why we adopt the term "L2/L3 French". Nasal vowels pronunciation acquisition in L2/L3 French in adulthood is commonly recognised as a particularly complex phenomenon [4,5,9,11]. The specific articulatory, acoustic and auditive properties of French phonemes have been described in several studies [6–9,11]). They differ from oral vowels in that they are produced through the lowering of the velum and in making specific modifications of the articulatory gestures and lip movements, which allow air to escape through the mouth and the nose at the same moment [6]. According to Delvaux et al. (2004) [7], those features, called gravity and compacity, respectively, are necessary for French L1 speakers to perceive the vowel as a nasal. According to Montagu (2002) [5], labiality is also a distinguishing feature between the nasal vowels /ɛ̃/, /ɑ̃/ and /ɔ̃/, the first being characterised by a low degree of labiality [-labial], the second by a higher degree [+labial] and the last by a maximum degree [++ labial]. Dherbey-Chapuy (2021) considers that a main acoustic property of nasal vowels is their second (F2) and third (F3) formants (resonance frequency of the sound wave), which are different from those of nasalised vowels [8].

The contrast between oral vowels and nasal vowels is phonemic in French [14], which means that nasal vowels play a crucial role in differentiating the meaning of words (e.g., *main* [mɛ̃] 'hand' vs. *mais* [mɛ] 'but' vs. *ment* [mɑ̃] 'lies' vs. *mon* [mɔ̃] 'my'). In this respect, nasal vowels must be distinguished from nasalised vowels found in many languages, which are allophones of oral vowels produced in surrounding nasal contexts. Based on the

F2 and F3 values, nasalised vowels are not produced in L1 French [8]. In this study, we use the term nasalised vowel in order to refer to the phonetic phenomenon of coarticulation with a nasal consonant in languages other than French. Therefore, we make therefore an a priori distinction between coarticulation as a phonetic phenomenon in a language used as a first language and the nasal consonantisation or nasalisation of the vowel as a phonetic phenomenon that has been observed in second languages or learner varieties: the use of a postvocalic nasal consonant.

The perception and production of French nasal vowels can be challenging for L2 learners whose L1 lacks equivalent phonemic nasal vowels: American English in studies conducted by Montagu (2002) [5] and Marquez-Martinez [4]; Japanese and Spanish in a study conducted by Detey et al. (2010) [9]; and Cantonese a study conducted by Li, Yin and Pu (2019) [11] Indeed, difficulties in acquiring these phonemes might lie both at a phonological level and at a phonetic level. For example, L2 learners whose L1 exhibits nasal vowels in a different phonemic contrast compared to French nasal vowels [15] or whose L1 does not exhibit phonemic nasal vowels at all need to develop new phonological representations in order to distinguish the target nasal vowels from nasalised or oral vowels available in their L1 or L2 phonemic repertoire [4,16]. This is the case for Japanese L2 English learners: although these languages are characterised by different vocalic systems, with English exhibiting a rich variety of vowels, while Japanese exhibits five vowels only. In both English and Japanese, nasal vocalisation is a non-phonemic feature, and the articulatory movements involved differ from the articulation of French nasal vowel' articulation [4,5,17]. In English, for instance, vowels can be nasalised through anticipatory coarticulation when they are followed by a nasal consonant (sank [sæ̃ŋk]). These nasalised vowels always co-occur with a nasal consonant and do not have distinct phonemic representations from their corresponding oral vowels (e.g., [18,19]). Furthermore, L2 learners whose L1 does not exhibit nasal vowels in their phonetic repertoire or whose L1 nasal vowels are articulated differently from French nasal vowels need to learn and coordinate specific aspects.

The intermediate pronunciation of nasal vowels in L2 French has been categorised in different terms according to theoretical frameworks and methodological approaches: rate of nasality [11], nasal unpacking or stripping [4], non-nasalisation of the vowel, residual [n] or [m], vowel substitution [20] (cited from ref. [4]) and degrees of consonantisation [9]. By means of aerodynamic, acoustic and lip movement measures, Li, Yin and Pu (2019) analysed the different physical properties of segments corresponding to the expected nasal vowel in a reading-aloud corpus of L1 Cantonese speakers of L2 French [11]. A comparison of nasality rates in L2 French and L1 Cantonese reveals a nasalisation rate that peaks at the end of the segment 'nasal vowel' in L2 French, which corresponds to the phonetic characteristics of vowels coarticulated with the nasal consonant in L1 Cantonese. In their study of the quality of nasal vowel realisation by Japanese L1 learners in L2 French, Detey et al. (2010) perceptively evaluated three degrees of nasal consonantisation of the nasal vowels: a first degree corresponding to the absence of a postvocalic nasal consonant, a third degree of clear presence of the same consonant and an intermediate second degree of presumed nasal consonantisation [9]. In her study of the acquisition of L2 French nasal vowels by adult speakers of L1 English, Marquez Martinez (2016) adopted a different categorisation partly linked to the theoretical framework she used, according to which the pronunciation of a nasal vowel by a beginner speaker of a language with nasalised vowels corresponds phonetically to the integration of the nasal feature into the vowel and the elimination of the time unit between the oral vowel and the nasal consonant that follows [4]. According to the author, the failure of this process in L2 results in two distinct 'strategies'. The first strategy is to 'unpack' or divide the nasal vowel into two segments, i.e., an oral vowel and a nasal consonant (*maison* 'house' is pronounced /mezɔn/ instead of /mezɔ̃/), called nasal unpacking. This strategy was first observed as the most frequent loanword adaptation strategy in languages without nasal vowels but borrowing words with nasal vowels, for instance, Lingala borrowing words from French [21]. The second strategy attributed to faulty perception is nasal stripping (*maison* is perceived as /mezɔ/,

then produced /mezɔ/). Phonetic parallels can be drawn between these different categories. Marquez Martinez's (2016) [4] nasal unpacking category represents a subcategory of Detey et al.'s (2010) [9] category 3: 'clear presence of a nasal consonant'. However, it cannot be reduced to this. In their typology of mispronunciations of the nasal vowel in L2 French in a corpus of spontaneous reading and speaking, Kamiyama et al. (2016) also identified, among intermediate pronunciations with a nasal consonant, cases of pronunciation of a nasal vowel followed by a nasal consonant (/mezɔ̃n/) [13]. Similarly, nasal stripping is a subcategory of the absence of a postvocalic nasal consonant but cannot be reduced to it either. The production of a nasal vowel, whether expected or not, is another case of zero degrees of nasal consonancy. Liddiard (1994) also noted cases of English speakers who substitute one vowel for another, producing /ɑ̃bʁ/ instead of /ɔ̃bʁ/[20]. Finally, we note the importance of the intermediate category corresponding to category 2 of Detey et al. (2010) [9] and to the 'residual [n] or [m]' category of Liddiard (1994) [20], in which the learner produces a nasal vowel followed by a slightly audible nasal feature. Even if the presence of a postvocalic nasal consonant seems characteristic of the intermediate pronunciation of the nasal vowel by speakers of first languages without nasal vowels, a detailed characterisation of phonetic realisations cannot be performed without considering the degree of nasal consonantisation (absence, residual or clear) and the shape of the vowel phoneme (oral or nasal).

In her research conducted on the perception of L2 French nasal vowels by English-speaking learners of L2 French, Marquez Martinez (2016) showed different strategies at different stages of development [4]. According to her analysis, at the initial stage, native English speakers of French split a nasal vowel into segments that already exist in their L1 phonological inventory: an oral vowel followed by a nasal consonant. However, as their exposure to French increases, intermediate learners tend to apply the nasal stripping strategy, perceiving French nasal vowels as oral vowels. Studies on the perception of L2 French nasal vowels by Japanese-speaking learners also reveal progressive discrimination of the nasal vowels, with beginners showing some difficulties in the identification of /ɑ̃/ and/or /ɔ̃/ and intermediate learners showing difficulties in perceiving the nasal vowel only /ɔ̃/ [10,22]. In a phonological discrimination task carried out by 124 students after their first year of acquiring French as a foreign language at university, Sauzedde (2018) reported that the mean score of discrimination was 36.2% for /ɑ̃/, 48.6% for /ɔ̃/ and 76.2% for /ɛ̃/. After the second year, the same task was performed again and this mean rate increased up to 61% for /ɑ̃/ and up to 83.8% for /ɛ̃/ but not for /ɔ̃/ (50.5%) [23].

As for the production of L2 nasal vowels, results from recent studies show some variation according to the stage of development, the task and the type of nasal vowel. English-speaking L2 learners [24] and Japanese-speaking L2 learners [10] pronounce phonemic nasal vowels from the earliest acquisitional stages. Differences between beginners and advanced learners involve more allophonic than phonemic changes [17]. In other words, the phonemic feature of nasality seems to be rapidly mastered in production, and difficulties seem to lie at the identification and articulation levels [21,24]. Kamiyama, Detey and Kawaguchi [13] found that Japanese learners have difficulties in pronouncing French nasal vowels, /ɛ̃/ being pronounced [ãɴ], /ɑ̃/ [ãɴ] or [õɴ] and /ɔ̃/ [õɴ]. In production, Detey et al. (2010) found that the average degree of nasal consonantisation of vowels in the Japanese-speaking group varied according to several variables [9]. Consonantisation varies for instance according to the L1, the position of the nasal vowel in the word and the type of nasal vowel. Their study has shown that the pronunciation of the nasal vowel /ɔ̃/ is more accurate than that of /ɑ̃/ and /ɛ̃/. Similarly, consonantisation varies according to the task: the rate was higher in the reading task than in the word repetition task, which was confirmed by an acoustic analysis and may suggest an effect of visuo-orthographic input.

In summary, the reviewed studies show that the production of a postvocalic nasal consonant instead of an expected nasal vowel in L2 French is a well-documented phenomenon. Although the term coarticulation is not used to refer to the phenomenon observed in L2 and, indeed, would not be sufficient to describe the aerodynamic and labial properties of nasal

vowels in L2, it seems to describe fairly well their phonetic characteristics as perceived by expert listeners. However, we use the terms nasalisation and nasal consonantisation to refer to the use of a postvocalic nasal consonant in an L2 instead of a nasal vowel in an L2.

An explanation often put forward in the studies presented above to account for the production of a postvocalic nasal consonant instead of an expected nasal vowel is the lack of nasal vowels and the existence of nasalised vowels in the first languages studied (English, Spanish, Japanese and Cantonese). This leads to the attribution of the nasal feature from the nasal consonant to the oral vowel, as is the case in the first language. Thus, there is no mention of the influence of other previously learned languages, in particular L2 English, on the acquisition of L2/L3 French by speakers of L1 Japanese. However, if the L1 plays a role in the interphonology of L2 French, we cannot rule out a role of an L2 in the acquisition of nasal vowels in L3 French and a reinforced use of a nasal consonant after a vowel instead of a nasal vowel by Japanese learners of L3 French due to the presence of nasalised vowels in L2 English. This hypothetical reinforcement of postvocalic consonancy needs further investigation.

Similarly, with the exception of Detey et al. (2010) [9], another influence is overlooked in most studies, namely the effect of orthography and the Latin letters <n> or <m> pronounced as the nasal consonants /n/ and /m/, respectively, in some positions. However, there is a large body of literature suggesting that the orthography of nasal vowels is likely to influence their pronunciation, as we observe in the next section.

### *2.2. The Effect of Orthography on Pronunciation*

Examining the acquisition of French nasal vowels by Japanese learners, Detey and Nespoulous (2008) highlighted the role of literacy in phonological awareness and the activation of orthographic representation by both auditory and visual stimuli [25].

A number of studies [1–3,26] have reported, based mostly on L2 English, that orthography plays a major role in shaping L2 phonology in speech perception and production. In fact orthographic forms can have a positive influence on speech perception (e.g., [27,28]) by providing cues that help learners discriminate L2 lexical items differentiated by a new phonological contrast. However, the positive influence of spelling is conditioned by the congruence of grapheme-to-phoneme correspondences (GPCs) between L1 and L2 (e.g., [29–31]). Regarding speech production, there are also contradictory results. Some studies have shown a positive effect of spelling, with orthographic input disambiguating auditory input [12,29,32,33]. Other studies have shown a negative effect of spelling in the presence of incongruent GPCs between L1 and L2, resulting in an orthographic pronunciation, e.g., the pronunciation of silent letters or double consonants or vowels [34–37]. In a recent review of the state of the art, Bassetti (to appear) considers sound additions as one type of orthographic effect, along with deletions and substitutions [38]. A frequent motivation for phonetic addition is that L2 users pronounce a silent letter that is part of a larger orthographic unit, such as a multi-letter grapheme. This is the case for English past morpheme <ed> being pronounced /t,d/ in some contexts in L1 English but sometimes /ɛd/ in L2 English because of the pronunciation of the silent <e> grapheme [35].

Some recent studies have shown that exposure to orthographic and auditory input, compared to auditory input alone, promotes lexical learning [39–42]. Indeed, learners are faster and more accurate in naming pictures after being exposed to the orthographic and auditory modality compared to the auditory modality alone. However, the presence of orthographic input during learning has a detrimental effect on the quality of production, leading to a non-native-like pronunciation of the target phonemes. As such, the impact of L1 GPCs on L2 pronunciation has been observed, even for non-alphabetic languages on the pronunciation of L2 alphabetic languages (i.e., [43], L1 Japanese on L2 English). Sokolović-Perović et al. (2020) [43] has shown that advanced Japanese L1 learners of L2 English are, indeed, influenced by L2 orthography and L1 pronunciation rules when it comes to consonant and vowel duration. As these findings have not been reported in studies on orthographic effects on phonology across scripts in beginner learners of

English [28,44,45], a possible interpretation for this orthographic effect is that the L2 learners already knew the orthographic form of the word.

However, the influence of orthography on production may be limited to the lexical stage of speech processing, since Ventura et al. (2004) found no orthographic effect for non-words and no effect in a repetition task. Since the repetition task does not necessarily rely on the participants' lexical representations, it could explain why their pronunciation was less impacted by orthography in a repetition compared to spontaneous speech [46]. Detey et al. (2010) found nasal consonantisation in repetition and reading tasks but attributed the more accurate pronunciation of nasal vowels by Japanese learners of L2 French in repetition than in reading to the effect of orthography [9].

To summarise the effects of spelling on L2 phonology, Hayes-Harb and Barrios [3] proposed four variables influencing the effect of spelling on learners' L2 phonological development. The first variable relates to the systematicity of the relationships between phonemes and graphemes in the target language (transparency vs. opacity of a writing system). If a new phonological contrast is systematically represented by the same graphemes in the L2, then learners will be able to rely on orthography to make inferences about the phonological structure of words. The second variable concerns the familiarity of one or more L2 graphemes and is modulated by the third variable: the congruence between L1/L2 GPCs. For L2 graphemes familiar in the L1, they may or may not have the same GPCs in the L1 and L2. If the GPCs are different, grapheme familiarity does not help and may even shape the L2 phonological development. Finally, the last variable is perceptibility, i.e., the learners' ability to perceive a new contrast.

Applied to to the acquisition of nasal vowels in L3 French by Japanese-speaking learners, the findings of previous studies on the effect of orthography on L2 phoneme pronunciation [47] suggest that the orthography a nasal vowel, a plurigrapheme with one or more graphemes <a, e, i, o, u> followed by <n> or <m> and, optionally, other silent or non-silent graphemes, could recall the nasal feature inherent in the nasal vowel because of the systematic use of <n> or <m> in writing in the target language and promote its pronunciation. However, in order to investigate the effects of orthography on pronunciation in spontaneous speech, it seems necessary to examine if learners have those orthographic representations, especially the <n> or <m> grapheme. Another possible effect is the use of the GPC rules of L2 English in L3 French. Even though the graphemes <n> and <m> do not exist in the syllabic alphabets of Japanese, Japanese-speaking learners are familiar with the Latin alphabet from a very early age, and we can therefore postulate a familiarity with these graphemes at the time of learning new contrastive phonemes in French due, in particular, to exposure to this graphic system during the often early learning of English. Nevertheless, there is no congruence between the French GPCs and those of previously acquired languages, as we show in the next section. This is why the orthography of nasal vowels can also have a negative effect on pronunciation, namely the production of a nasalised vowel instead of the pronunciation of the nasal vowel. Effects linked to the task and the stage of development are also to be expected. However, they require further research.

### *2.3. Awareness of Orthographic Effects on Pronunciation: Some Didactic Proposals*

Is the aim of teaching the pronunciation of nasal vowel phonetic correction and the production of the acoustic, articulatory and auditive properties of the French nasal vowel or simply the production of one or more phonemes that can be interpreted as a nasal vowel?

A number of tips or suggestions for teaching nasal vowel pronunciation in French have emerged from L2 studies that have been carried out. Li, Yin and Pu (2019) advised that L1 interference should be taken into account in teaching [11]. Montagu (2002) suggested emphasising the role of labiality [5] in phonetic correction lessons. Detey et al. (2010) emphasised the need to offer a variety of activities in written and oral modalities in order to develop balanced phonetic–phonological and phonographic competence [9]. These interesting proposals focus on the pronunciation of nasal vowels and do not specify the

degree of intelligibility of a French word containing a nasalised vowel instead of a nasal vowel. Do these intermediate pronunciations disturb the intelligibility of an utterance? This question seems all the more important, since, according to Dherbey-Chapuis (2021), there is, at the phonetic level, a "high variability in the pronunciation of nasal vowels (NVs) among French speakers", which "makes it hard to compare the formant values of the NVs pronounced by learners with a native-like norm" [8].

Recent research on the acquisition of L2 pronunciation has been mainly conducted on L2 English as a lingua franca in international communication, suggesting that the aim of pronunciation teaching in English should not be the mastery of a native norm and accent reduction but word intelligibility, fluency and comprehensibility of discourse in interaction [3,38,48–50]. Following [49], intelligibility may be defined as "the extent to which a speaker's message is actually understood by a listener" and may be operationalised using a technique of word-by-word sentence transcriptions made by listeners, as reported by Munro and Derwing (2020) [51]. This means that a phoneme like a nasal vowel may be transcribed as a nasal vowel even if it does not have all the acoustic properties of the nasal vowel in the target language. According to Levis (2018), teaching intelligible pronunciation does recognise the importance of acquiring the contrastive phonemes of a language, since an error in a word's phoneme can impair speech intelligibility [49]. If we look at nasal vowels, we can ask ourselves what the conditions for a vowel to be intelligible in a given context are and which activities or interactional feedback could favour the use of intelligible nasal vowels.

To build a curriculum with the aim of intelligibility of pronunciation, one proposal has emerged: focus on distinctive phonemes or phonemes with a high functional value (past time morphemes, for instance) [48,50] common to a set of varieties of the language [50]. Even if Derwing (2017) [48] and Colantoni et al. (2021) [50] agree on the importance of setting the goal of intelligibility from the earliest stages, the type of activities needed to achieve these goals is less clear. For example, integrated activities are proposed, either aiming at showing the lexical or grammatical functions of a phoneme (minimal pairs like *pain* 'bread' vs. *pont* 'bridge') or focusing on its articulation with other components.

However, these studies do not address the question of the impact of exposing learners to the written forms of words in addition to their exposure to the spoken forms. To the best of our knowledge, only four works have transposed the results of studies on orthographic effects into didactic proposals [3,8,40,52]. Among these, one focused on experimental didactics [8]. Some studies have, nevertheless, experimentally used unfamiliar scripts to explore the effect of unfamiliar orthographic forms on word learning and syllable discrimination, since L2 orthography and L1 GPC rules may influence L2 phonology. These studies have shown contradictory results, a facilitative effect [28,44], no effect [45] or a negative effect [53,54].

Another way to look into supporting the phonological development of L2/L3 learners at initial stages is to identify (un)intelligible pronunciations in conversational contexts, such as those occurring in the classroom, and to analyse the variable use of nasal vowels according to the amount of exposure, contexts of production and type of nasal vowel.

## 3. The Phonetic vs. Phonemic Role of Nasality in French, Japanese and English

In this section, we present some descriptive elements of nasality and its orthography in languages in contact, which is essential for discussing phonological and orthographic influences, as well as the role of nasal vowels graphemes <n> or <m> in acquisition of L3 French by Japanese-speaking learners in the initial stages: the presence of nasal vowels in the phonological inventory or allophone nasalised vowels of oral vowels, their frequencies, the writing system(s), grapheme-to-phoneme rules and vice-versa.

### *3.1. Nasality in French: Three Nasal Vowels and More Than Twenty Corresponding Graphemes*

In French, at the phonemic-level nasal vowels are contrastive phonemes that need to be learned because this phoneme may determine the lexical or grammatical meaning of

the word. At the phonetic level, the vocal nasality is a very complex phenomenon at the articulatory, acoustic and aerodynamic levels [7]. However, nasal vowels are produced variably and tendentiously differently by the majority groups in northern and southern France. In southern French, from the Basque Country to Provence via Gascony, speakers "have a variable degree of nasalisation depending on the speaker and are followed by a more or less prominent nasal appendage" [55]. Thus, in southern French, there are four distinctive nasal vowels: /ɛ̃ⁿ/, /œ̃ⁿ /, /ɔ̃ⁿ /, /ɑ̃ⁿ /, with /ⁿ/ being the variable nasal appendage [13]. Even though nasal vowels are contrastive phonemes in French, they are realised differently depending on the region and the speaker. This case of variation shows the limits of the notion of accuracy or native-likeness and the value of a measure of intelligibility that takes into account the speaker's perception.

The French nasal vowels ɑ̃, ɔ̃ and ɛ̃ are not as frequent as their vocal counterparts. According to Planton's investigation of the Lexique Database (2014: xviii) [56], the vocal vowels in initial and final positions in words are more frequent than the nasal vowels for two of them: /a/ (18334) vs. /ɑ̃/ (15344) and /ɛ/ (18907) vs. /ɛ̃/ (6816), except for /ɔ̃/ 8943 vs. /ɔ/ (992) vs. /o/ (4450), although they are still frequent in French words (see Table 1). In both the Gougenheim and Lexique 3.83 corpora [57], words with a nasal vowel are very frequent and the three most frequent words with a nasal vowel are the monophonemic words *on* /ɔ̃/, *en* /ɑ̃/ and *un* /ɛ̃/ (Table 1).The Gougenheim corpus is a corpus of spoken French collected in the early 1950s to provide the lexicon of elementary French, i.e., 1500 frequent French words. Lexique 3.83 is a vast corpus of spoken and written French. We have used the sub-corpus "Sous-titres de films populaires", which includes the French subtitles of 9474 films or series, representing a total of 50 million words. The subtitles come from four categories of films: French films (1.9 million words) (e.g., Camille Claudel), Anglo-Saxon films (26.5 million words) (e.g., Arizona Dream, Schindler's List), Anglo-Saxon films and series, (19.5 million words) (e.g., Friends, Ally McBeal) and non-English European films (2.5 million words) (e.g., Cria Cuervos, Good Bye Lenin!).

**Table 1.** Words with a nasal vowel among the 200 most frequent words in the corpora of Gougenheim and Lexique 3.83.

| **Gougenheim Corpus (1954)**<br>**Unit = Type** | **Lexique 3.83 Corpus**<br>**Unit = Token** |
|---|---|
| on- /ɔ̃/ 'we', un-/ɛ̃/'a/one', en-//ɑ̃// 'in', dans-/ɑ̃/ 'in', en-/ɑ̃/ 'of it' -pronoun, non-nɔ̃ 'no', enfin-/ɑ̃/,/ɛ̃/'at least', quand-/ɑ̃/ 'when', mon-mɔ̃ 'my', ben-/ɛ̃/ discourse particle, prendre-/ɑ̃/ 'to take', rien-/ɛ̃/ 'nothing', un peu-/œ̃/ 'a litle bit', encore-/ɑ̃/ 'still/again', hein-/ɛ̃/ interjection, grand-/ɑ̃/ 'big', temps-/ɑ̃/ 'time/weather', eh bien-/ɛ̃/ - interjection, an-/ɑ̃/ 'year', son-ɔ̃ 'his/her', cent-/ɑ̃/ 'hundred', comprendre-/ɔ̃, ɑ̃/ 'to understand', maintenant-ɛ̃, /ɑ̃/ 'now', bon-ɔ̃ 'good', matin-/ɛ̃/ 'morning', évidemment-/ɑ̃/ 'obviously', avant-/ɑ̃/ 'before', seulement-/ɑ̃/ 'just/only', pendant-/ɑ̃, ɑ̃/ 'during', français-/ɑ̃/ 'French', entendre-/ɑ̃, ɑ̃/'to hear', un-/ɛ̃/ 'one'-pronoun, commencer-/ɑ̃/ 'to start', un-/ɛ̃/ 'one'-numeral, rendre-/ɑ̃/ 'to give back', tellement-/ɑ̃/'so', trente-/ɑ̃/ 'thirty', quand même-/ɑ̃/ 'still/anyway', moment-/ɑ̃/ 'moment', vingt /ɛ̃/ 'twenty', comment-/ɑ̃/ 'how', moins-/ɛ̃/'less', vraiment-/ɑ̃/ 'really', franc-/ɑ̃/ -old French money, cinq-/ɛ̃/ 'five', enfant-/ɑ̃, ɑ̃/ 'child', demander-/ɑ̃/ 'to ask' | un-/ɛ̃/ 'a/one', on /ɔ̃/ 'we', en-/ɑ̃/ 'in', dans-/ɑ̃/ 'in', bien-/ɛ̃/ 'good', non-/ɔ̃/ 'no', mon-/ɔ̃/ 'my', en-/ɑ̃/ 'in', rien-/ɛ̃/ 'nothing', quand-/ɑ̃/ 'when', son-/ɔ̃/ 'his/her', ton-/ɔ̃/ 'your', sont-/ɔ̃/ 'are', encore-/ɑ̃/ 'still', temps-/ɑ̃/ 'time/wheather', maintenant-/ɛ̃, ɑ̃/ 'now', sans-/ɑ̃/ 'whithout', vraiment-/ɑ̃/ 'really', viens-/ɛ̃/ 'come', comment-/ɑ̃/ 'how', bon-/ɔ̃/ 'good', monde-/ɔ̃/ 'world/people', besoin-/ɛ̃/ 'need', ans-/ɑ̃/ 'years', quelqu'un-/ɛ̃/'someone', donc-/ɔ̃/ 'then', gens-/ɑ̃/ 'people', maison-/ɔ̃/ 'house', bonjour-/ɔ̃/ 'hello', comment-/ɑ̃/'how' |

Among the 200 most frequent lemmas in the Gougenheim corpus, there are 46 lemmas with a nasal vowel, and among the 200 most frequent occurrences in the corpus of French subtitles of popular American films in the Lexique 3.83 database [57], 30 occurrences contain at least one nasal vowel. These frequent words with a nasal vowel are grammatical words (*on* 'we', *dans* 'in', *mon* 'my'); discourse particles (*ben, hein, eh bien*); and lexical words, including nouns *(bonjour* 'hello', *matin* 'morning', *maison* 'house'), adjectives (*bon*'good',*français* 'French'), verbs (*prendre* 'to take', *demander* 'to ask', *rendre* 'to give back') and adverbs (*seulement* 'just/only', *évidemment* 'obviously', *encore* 'still/again'). Some of these frequent words often contain two nasal vowels, such as *enfin* 'at least', *enfant* 'child', *pendant* 'during', *comprendre* 'to understand', *maintenant* 'now' and *entendre* 'to listen'. To sum up, all listeners of French are immediately and recurrently exposed to nasal vowels.

According to Hayes-Harb Barrios (2021), a crucial point in predicting the effect of orthography on pronunciation is the consistency of phoneme-to-grapheme and grapheme-to-phoneme correspondences. In this respect, French has a deep orthographic system, and it is not easy to establish a measure of consistency in languages with deep orthography. In his research, Planton (2014: xx) took up this challenge and listed the phoneme-to-grapheme correspondences of 142,000 orthographic forms in the Lexique 3.8 database, given their position in the words (initial, median and final). He included inflected forms, for example, the feminine and plural forms of nouns, and excluded monophonemic words such as <un>, <on> and <en>. His aim was to account for the phoneme-to-grapheme consistency of the French phonemes, a so-called opaque language, in order to model as reliably as possible the consistency of the correspondences between phonological sequences and orthographic units in the French language. If we take the three nasal vowels, his analyses show that /ɑ̃/ has 7 possible spellings in word-initial position (by frequency range, en (most frequent), em, an, am, han, ham and hen) and 26 in word-final position (by frequency range, ant (most frequent), ent, ans, and, emps, ents, ants, an, end, ens, ends, ands, anc, ang, ancs, amp, amps, angs, aon, engs, eng, aons, am, empt, en and ams), giving a total of 33 possible spellings (not including phonemes in the median position). The nasal vowel /ɔ̃/ has 5 possible spellings in word-initial position (on, om, hon, hom and un) and 17 in word-final position, for a total of 22 different spellings, and the nasal vowel /ɛ̃/ has 7 possible spellings in word-initial position (in, im, ain, un, hin, hum and ein) and 28 in word-final position, for a total of 33 different spellings. In other words, even though the nasal vowel system is generally very inconsistent in French, the nasal vowel /ɔ̃/ is the one whose spelling varies the least.

### *3.2. Japanese: Nasalised Vowels*

Japanese speakers have a repertoire of five vowels /a, e, i, ɔ, u/, all of which can be short or long, making it possible to distinguish between two lexical words. Japanese also has two nasal units: /n/ and /N/. /n/ is a nasal consonant, whereas /N/ is a so-called mora, a basic longer phonological sequence. The Japanese nasal mora /N/ has a variable realisation, at least in coda position, including coronal, velar-to-uvular and even as a nasalised glide [58,59] (We thank an anonymous reviewer for this precision). Youngberg (2021) even proposed that "in the place of a nasal consonantal coda, the Tōkyō variety has a nasal vowel, Ōsaka has a syllabic nasal and Kagoshima has a nasal coda" [60]. In Japanese, vowels may be nasalised, but this phonetic feature does not change the lexical meaning of the word. Vowels may be nasalised in front of a vowel, in front of the phonemes /s/ or /ʃ/ or at the end of a word [61]. Japanese is otherwise a moraic language with essentially open and simple syllables and no consonant clusters.

Three writing systems are used by Japanese speakers: the kanji system, with ideographs, which are in a deep relationship to phonology; the hiragana system, which is a moraic system in a transparent relationship with phonology (one mora CV, one character); the katakana system, which is a transparent moraic system as well but devoted to the writing of foreign words; and the Romaji system, which is alphabetic, corresponding to the Latin alphabet

and used for specific purposes (proper names of foreigners, locations in the public transport network, labels, etc.)

*3.3. English: Nasalised Vowels*

The vowel inventory of English is larger (there are usually 7 short vowels and 10 long vowels in the standard American English variety). But vowels are vocalic, and nasal vowels do not belong to English phonemes. There are nasalised vowels when a vocalic vowel precedes or follows a nasal consonant /n, m, ŋ/, but they are just considered allophones of the oral English vowels. The English writing system is alphabetic, and the phoneme-to-grapheme and grapheme-to-phoneme correspondences are known to be deep. Nevertheless, the graphemes <n>, <ng> and <m> regularly encode nasal consonants. Note that a number of words in English and French are homographs, such as <France>, pronounced [fʁa:ns] in standard English (as prescribed in a dictionary) and [fʁɑ̃s] in standard French or [fʁɑ̃$^{n}$s] in meridional French.

In summary, Japanese learners of L3 French have to learn at a phonological level that nasal vowels are contrastive phonemes, and they have to learn to pronounce new sounds (articulations, lip movements and duration).

## 4. Research Questions and Hypotheses

Our study aims to examine whether the orthographic representation of the nasal vowels in L3 French has an impact on their pronunciation. More specifically, it attempts to verify whether the plurigraphemic representation of nasal vowels including the grapheme <n> or <m> (<V(V)n/m(CC)>) leads to the use of a postvocalic nasal consonant by Japanese learners using L3 French in interaction at the initial stages. The influence of orthography has been observed in experimental reading-aloud and repetition tasks (e.g., [2,34,36,43,47]). According to previous studies on the effect of orthography on pronunciation, the grapheme <n> could favour the retention of the nasality feature but favour a nasalisation of the vowel, especially at the initial stages [4].

In the light of previous studies and the phonological and orthographic properties of the languages involved (L1 Japanese, L2 English and L3 French) our hypotheses are outlines as follows:

1. H1: The graphemes <n> and <m> influence the pronunciation of nasal vowels and give way to nasal consonantisation or nasalisation, i.e., the use of a vowel followed by a nasal consonant, as observed in other studies [4,9,11].
2. H2: This effect of the graphemes <n> and <m> decreases with time of exposure to L3 French, mainly due to increasing exposure to the acoustic forms of nasal vowels.
3. H3: Nasalisation of vowels in L3 French is more frequent in spontaneous speech—a word retrieved from the mental lexicon, where the oral and written forms co-exist—than in repeated speech—a word repeated from the feedback of the interlocutor [9,46].
4. H4: Nasalisation should vary according to the phoneme-to-grapheme consistency and to the frequency of the nasal vowel in the production context; the nasal vowel /ɔ̃/ will thus be less nasalised than the vowels /ɑ̃/ and /ɛ̃/ because of its relatively stronger phoneme-to-grapheme consistency [3,56].
5. H5: Japanese learners of L3 French use the graphemes <n> or <m> to write words with nasal vowels at the initial stages.

The results of this study will be discussed in order to contribute to the didactic proposals already made with respect to nasal vowel teaching and to outline some possible avenues for a research agenda that could be relevant for teaching practices related to the interfaces between phonology and discourse, the effects of orthography and translinguistic influences.

## 5. Methodology

To answer the two questions and test the five hypotheses, we carried out an analysis of perception of nasal vowels in a longitudinal corpus of conversational and narrative speech

produced by 4 Japanese learners of L3 French interacting with a speaker of L1 French. We also analysed the learners' orthographic productions. Although the sample is limited to 4 participants, their productions present two important advantages with regard to our research question and hypotheses. First, the productions were collected in a face-to-face conversational setting that has not been extensively studied until now. This oral corpus thus enables us to observe two production situations: when the learner produces a known word and when he repeats an unknown word given by his interlocutor. These two situations allow us to ecologically observe the effect of orthographic knowledge of the word in the production process. The second reason is that the corpus includes both written and oral productions, enabling comparative analysis of the phonology and orthography of the same learners. To the best of our knowledge, such an ecological, bimodal corpus has not yet been analysed. We come back later and in the conclusion to the limitations raised by the sample size.

### *5.1. Participants*

Four Japanese students aged 19–20 (one male and three females) who were enrolled in an optional French course at a Japanese university participated in this study. All of them are multilingual, having learned Japanese in the Kanto area from birth, then English when entering secondary school, with a tested level of at least 500 points in TOEFL up on entering university, where they began to learn French as a third language (Table 2).

**Table 2.** Participants.

| ID | Gender | Age | L1, L2 |
|---|---|---|---|
| JL1 | female | 19 | Japanese, English |
| JL2 | female | 19 | Japanese, English |
| JL3 | male | 19 | Japanese, English |
| JL4 | female | 20 | Japanese, English |

### *5.2. Exposure to French*

The first recording (T1) took place in July, three months after the beginning of the French course. The four young adults had then benefited from about 50 h of institutional exposure to spoken and written French. The second data collection time (T2) took place five months later, which corresponded to a further 70 h of exposure to French, i.e., a total of 120 h of instruction (4.5 h a week (3 × 90 min)).

During the first year of French learning, the weekly exposure to input took place mainly in the classroom with two bilingual teachers, a teacher of L1 Japanese and L2 French and a teacher of L1 French and L2 Japanese. The L1 French teacher speaks a standard variety of French and gave listening comprehension and oral production classes that can be described as oralographic. The term oralographic, as used by Bouchard [62], refers to the fact that the learner in a language class is exposed both to the flow of phonemes and to the orthographic chain, whether it be the text in the handbook or the words or sentences written on the board, such as the words /ãglɛ/, /ãglɛz/ that are repeated orally by the teacher and then written <anglais(e)> on the board (turn a) in (1). Learners are therefore simultaneously exposed to the phonemic and graphemic forms of the words.

(1) Ishikawa Corpus [63], 1 week before T1: Interactive correction of an exercise in the handbook *Spirale. Méthode de français pour débutants*. T = teacher; A8 and A1 are identified students, unlike An. Transcription conventions:/ small pause; (3 s) pause whose length is 3 s; syllables perceived as stressed because of their higher volume or longer duration are transcribed in capitals.

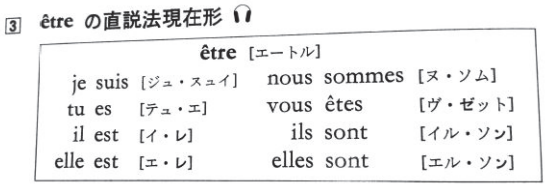
a. T: BRAVO !/ très très bien anglais hein ? (T. writes on the black board <anglais(e)>)/ anglais anglaiSE !/ adjectif (3s) BONJOUR ! (*GREAT!/ very very good, anglais, no?/ (P. writes on the black board <anglais(e)>)/ anglais anglaiSE !/ adjective (3s) HELLO !*)

b. A8: bonjour (*hello*)

c. T: ON CONTINUE !/ on continue s'il vous plaît !/ « le mont Blanc » ! (2s) (*Let's go on!/ let's go onplease !/ « le mont Blanc » ! (2s)* )
d. A1 : c'est/ c'est en France (*it's/ it's in France*)
e. T: très très BIEN !/ c'est en France/ À ? (*very very GOOD!/ it's in France/ in?*)
f. A1: À Chamonix (*in Chamonix*)
g. T: BRAVO !/ VOUS CONNAISSEZ BIEN !/ OUI :::!/ à Chamonix !/ à Chamonix (*GREAT!/ YOU KNOW WELL!/ YES:::!/ in Chamonix!/ in Chamonix*)
h. An: Chamonix ?/ Chamonix ?
i. T: Chamonix/ Chamonix/ on écrit comme ça/ regardez/ à Chamonix (2s) (P. writes on the black board <Chamonix>) / c'est près de Genève ! (*Chamonix/ Chamonix/ you write it like this/ look/ in Chamonix (2s) (P. writes on the black board <Chamonix>) / it's near Genève !*)

These interactions in the language classroom show the omnipresence of words with nasal vowels right from the start of the learning process, either in the lexicon used in the read aloud dialogues or in the different activities (*en anglais 'in English'*, *Mont Blanc*, *en France 'in France'*). Nasal vowels were also present in the frequent and more or less emphasised feedback given by the teacher (*très bien bien* 'very very good' turn a ; *très très BIEN* turn e, *VOUS CONNAISSEZ BIEN 'YOU KNOW WELL'!* in turn g) or in interactional management (*BONJOUR 'HELLO'* to greet a student entering the room in a and *ON CONTINUE !/ on continue s'il vous plaît 'WE GO ON !/ we're going on please'!* in c).

An analysis of the two textbooks used in the course, *Spirale* [64] and *Bonjour, Paris* [65], shows that the written forms of words in French are also transliterated in the first lessons. In other words, their pronunciation is transcribed in one of the two syllabaries used in Japanese: *katakana*, which is often used to write foreign words [23,66]. The inflected forms of the verb *être* that the learner can listen to are thus doubly transcribed in the textbook: orthographically with the Latin alphabet system usually used for French and between square brackets moraicly with the *katakana*, which is usually used to write foreign loanwords (Figure 1).


3 être の直説法現在形

être [エートル]

| | | | |
|---|---|---|---|
| je suis | [ジュ・スュイ] | nous sommes | [ヌ・ソム] |
| tu es | [テュ・エ] | vous êtes | [ヴ・ゼット] |
| il est | [イ・レ] | ils sont | [イル・ソン] |
| elle est | [エ・レ] | elles sont | [エル・ソン] |

**Figure 1.** Inflectional paradigm of the verb *to be* in *Bonjour Paris*, p. 12.

The transliteration of French with the katakana syllabary was very frequent in textbooks published until 2008 (Sauzedde 2014: 112). It associates an adapted, simplified phonological sequence with the graphemic word. For instance, the 3rd person plural form /sɔ̃/ is doubly transcribed with the Latin alphabet as <sont> and with two kanas respectively pronounced /sɔ/ and /ɴ/. This leads to two graphemic and phonological representations for the same word: /sɔ̃/, which might be heard in the classroom, and /sɔ.ɴ/, which might be read in the textbook. Moraic transliteration as in Figure 1 thus favours the use of a postvocalic nasal consonant.

In our study, participants were exposed to phonemic and graphemic representations of a variety of words containing nasal vowels at T1 and T2, namely to their bi- or trigraphemic forms provided in the textbook or on the blackboard and to bi-moraic written forms provided in the textbook and possibly used by learners in their own notebooks.

### *5.3. Corpus*

The investigated data set is a longitudinal corpus of productions elicited through three tasks (two spoken and one written production task) carried out at two data collection times, T1 and T2, after 50 and 120 h of exposure (+70 h) to French in the university, respectively.

The recordings took place in an ordinary room and allow for perceptive analysis but do not have the quality required for acoustic analysis.

The two oral tasks consist of :

1. A semi-guided interview conducted between the interviewer and each participant;
2. A picture story in the form of vignettes about a boy who loses his cat and goes looking for it.

The picture story features situations whose descriptions are likely to trigger the production of frequent lexical units with nasal vowels: /garsɔ̃/, written <garçon>, meaning 'boy'; /ɑ̃fɑ̃/ <enfant> 'child'; /ʃɑ̃bʁ/, <chambre> 'room'; /parɑ̃/ <parent> 'parent'; /mɛzɔ̃/, <maison> 'house'; /vwazɛ̃/ <voisin> 'neighbour'; /pwasɔ̃/ <poisson> 'fish'; ǀ/ʃjɛ̃/ <chien> 'dog'; /jardɛ̃/ <jardin> 'garden', to mention but a few examples of names used to refer to the main protagonists and places in the story as in (2).

(2) JL3, Cat Story, T1.

a. et /garsɔn/ (.) /tuvə/ a CHAT /tuvə/ a chat (..) avec /pwazõn/
and boy (.) find a CAT find a cat (..) with fish
'and the boy finds a cat finds a cat with a fish'

The spoken longitudinal corpus of conversational and narrative speech in interaction makes it possible to analyse the variable pronunciation of nasal vowels by Japanese learners according to a set of variables. These two oral tasks allow for examination of the following factors:

- How the L1 Japanese learners produce expected nasal vowels in L3 French: as a nasal vowel, an oral vowel or a nasalised vowel, as observed in previous studies;
- How the pronunciation of expected nasal vowels evolves with increased exposure to French between T1 and T2;
- How the pronunciation of expected nasal vowels varies according to the micro-context of production: if the nasal vowel is spontaneously produced in a word retrieved from the mental lexicon such as JL4's *bonjour*, /bõʒuʁ/, 'hello' in (3) or repeated as a word immediately given before by the interlocutor such as JL4's /fwɑ̃sɛ/ 'French' in 4f;
- How the pronunciation of expected nasal vowels varies according to the type of nasal vowel, i.e., the consistency of their phoneme-to-grapheme correspondences.

(3) Interview, T1, JL4 enters the room where the interviewer is.

a. JL4: xx bonjour (*hello*)
b. INT: bonjourE mm asseyez-vous j'vous en prie (*hello mm sit down please*)
c. JL4: mercé (*than you*)

(4) Interview, T1, presentation phase.

a. INT: d'accord d'accord j'ai compris et euh JL4 vous étudiez le français? (*ok ok I understand and euh JL4 do you study French?*)
b. JL4: vousétu (*doyoustu?*)
c. INT: vous vous étudiez le français? et quoi? (*do you study French and?*) (*Let's go !/ let's go please !/ « le mont Blanc » ! (2s)* )
d. A4: et quoi? kanshoïzume (*and kanshoïzume*)
e. INT: oui nononon à XX [=name of the town] YY [=name of the university] vous étudiez le français avec ZZ [= name of the teacher] (*yes nonono in XX YY do you study French with ZZ?*)
f. JL4: lel fwançais and they anglais (*the Flench and they English*)
g. INT: anglais aussi très bien (*English as well very good*)
h. JL4: oui yeh yeh anglais (*yes yeh yeh English*)
i. INT: très bien français et anglais (*very very good, French and English*)

The written task is a narrative task involving the retelling of two extracts from the movie *Modern Times* that the student watches in a sequence. They watch the first extract (approx. 1 min) in which a woman steals a loaf of bread, runs away and is arrested by

the police—the participants have 10 min to write about this passage. Then they watch the second extract, which lasts about a minute and, again, have 10 min to recount the scenes they have seen. The participant writes by hand and does not use any lexical or grammatical aids. This task is also likely to trigger the use of plurigraphemic units <a,e,u,i + n> to encode nasal vowels such as <pain> 'bread', <prend> 'takes', <entre> 'enters', <mange> 'eats' and <restaurant> 'restaurant', in addition to grammatical units such as <un> 'a'/'one', <dans> 'in', etc. The written corpus makes it possible to check the orthographic knowledge of the nasal vowels /ɔ̃/, /ɑ̃/, /ɛ̃/ at the initial stage.

*5.4. Coding and Analysis*

In the spoken corpus, coders identified the units with nasal vowels, and each unit with an expected nasal vowel in French was considered a target token and analysed by four expert French listeners who coded the following elements:

- The expected type of nasal vowel: /ɑ̃/, /ɔ̃/, /ɛ̃/;
- The perceived phonemes as corresponding to the target nasal vowel or not;
- The context of production: repeated versus spontaneously produced and retrieved;

The expert listeners were four female, multilingual expert teachers of L2 French and researchers in L2 acquisition living in the south of France at the time of coding and who socialised in various places. Table 3 shows the distribution of the coding task according to coders.

**Table 3.** Distribution of coders.

| Coder | Participant |
|---|---|
| C1 | JL1, JL2 |
| C2 | JL2, JL1 |
| C3 | JL3, JL4 |
| C4 | JL4, JL3 |

The categories used for coding were identified by the four coders based on data analysis. Overall, seven categories of nasal vowel realisations were found: (1) NV, nasal vowel: /pɑ̃s/, *pense*, 'think'; (2) NV + N, nasal vowel followed by a nasal consonant: /ʃɑ̃t/ *chante*, 'sing'; (3) OV + N, oral vowel followed by a nasal consonant /n/, /dan/ *dans*, 'in'; (4) OV + M, oral vowel followed by a nasal consonant /m/, /kɔmpʁi/, *compris*, 'understood'; (5) OVS + N, substitution of the oral vowel (/e/ instead of /ɛ/ followed by a nasal consonant /n/, /ʃen/ for *chien*, 'dog'; (6) OV, oral vowel, /ʃabɾ/ *chambre*, 'room'; (7) OVS, substituted oral vowel, /a/ *un*, 'a'.

In the quantitative analyses, the seven categories used in the data-driven coding are merged into three pronunciation categories of the three nasal vowels ɑ̃, ɔ̃ and ɛ̃ as follows:

- NV: nasal vowel;
- V + N : vocalic or nasal vowel followed by nasal consonant (merging categories NVN, OVN, OVM and OVSN);
- OV: oral vowel (merging categories OV and OVS).

In the written narratives, we selected the target words, e.g., <mange> 'eats', <numbreux> 'numerous' and <restrant> 'restaurant', as in (5), and coded the accuracy of the graphic encoding of the nasal vowel, regardless of the other surrounding syllables and consonants. Thus, we considered that the nasal vowel in the word <restrant> was correctly spelled, using the two graphemes <a> and <n>, unlike in the word <numbreux>, where the bigraph <um> does not match the expected bigraph <om>.

(5) JL2, *Modern Times*, Written retelling, Restaurant scene (The examples provided are the learners' spellings).

a. <Il mange numbreux déjunéur à réstrant.>

The analyses of the pronunciation and spelling of the French nasal vowels by the four beginner Japanese-speaking learners are presented in the following section.

## 6. Results

The analysed corpus includes 520 tokens (i.e., syllables) with an expected nasal vowel: 467 in oral production (Table 4) and 53 tokens in written production. Despite a small sample limited to four speakers, the number of analysed occurrences is higher or equivalent to that of previous experimental studies [9,11]. Since our research question concerns the effect of the graphemes <n> and <m> on the pronunciation of nasal vowels, our analyses focus first and foremost on the spoken corpus and the pronunciation of the nasal vowels. Analysis of the learners' written corpus and the learners' orthography will be addressed at the end of this section.

**Table 4.** Interactional Corpus: duration and number of expected nasal vowels (NV).

| ID | Duration (mn) | NV T1 + T2 |
|---|---|---|
| All | 172 | 467 |
| JL1 | 36 | 90 |
| JL2 | 53 | 114 |
| JL3 | 40 | 143 |
| JL4 | 43 | 120 |

There are individual differences in the use of syllables, with expected nasal vowels ranging from 1 to 1.5 between JL1, who uses 90 syllables with an expected nasal vowel, and JL3, who uses 140.

### *6.1. Pronunciation*

Each syllable with an expected nasal vowel was evaluated by two raters in terms of correction and type of error where applicable. The raters have congruent perceptions in correction and type of error in 335 of the 467 occurrences (71.7%). A closer look into the most frequent words in the learner's corpus (*non* 'no', *un* 'a/one', *français* 'French', *chien* 'dog', *garçon* 'boy', *en* 'in', *son* 'his', *poisson* 'fish', *dans* 'in' and *content* 'happy') reveals an inter-rater rate over 90 %, except for *son*.

Furthermore, the raters agree on perceived nasalisation in one form or another: nasal vowel (NV) or vowel + nasal consonant (V + NC) in 109 cases (Table 5). In the remaining 23 cases, they disagree on a perceived oral vowel (OV) and perceived nasalisation of some type (NV or V + NC). In total, there is agreement on perceived nasalisation in 444/467 (95.1%) of the syllables with expected nasal vowels.

The pedagogical consequences in terms of intelligibility of this agreement on nasalisation but disagreement on the type of nasalisation will be further discussed in Section 7.

Here, we concentrate on the 335 cases where both raters agree on the type of pronunciation. The main reason for this is that we do not investigate the pronunciation of the nasal vowels per se but whether it is possible to observe influences of orthography on the oral production in a corpus study, as it has been shown in experimental studies (e.g., [2,34,36,43,47]), and when the L1 has a non-alphabetic script [43]. The results are presented in the order of the four hypotheses.

**Table 5.** Common perception of the expected nasal vowels.

| Expected Nasal Vowels (N = 467) | Perceived Phoneme(s) | (Dis)agreement |
|---|---|---|
| 335 | Nasal or Oral Vowels<br>O, NV | |
| 109 | Nasal(ised) Vowels<br>NV vs OV + N | |
| 444 | | Agreement on orality or nasality (95.1%) |
| 23 | Oral or Nasal(ised) Vowel<br>OV vs NV or OV + N | Disagreement on orality or nasality (4.9%) |

6.1.1. First Hypothesis: The <n> Spelling Leads to the Pronunciation of the Nasal Consonant following the Vowel (V + N)

The first hypothesis is that the pronunciation of French nasal vowels is affected by their spellings, leading to the pronunciation of the nasal consonant following the vowel to be nasalised.

The analysis of the 335 congruently perceived sounds shows that 72% (241/335) of them are perceived as expected nasal vowels, 19.5% (66/335) as a vowel followed by a nasal consonant and 8.5% (28/335) as oral vowels (Table 6).

**Table 6.** Distribution of the perceived phonemes (n = 335) in % and number of occurrences per participant and category (#).

| ID | NV | V + N | OV | Total |
|---|---|---|---|---|
| All | **72**<br>(241) | **19.5**<br>(66) | **8.5**<br>(28) | 100<br>(335) |
| JL1 | 84<br>(61) | 11<br>(8) | 5<br>(4) | 100<br>(73) |
| JL2 | 79<br>(61) | 21<br>(16) | 0<br>(0) | 100<br>(77) |
| JL3 | 62<br>(65) | 30.5<br>(32) | 7.5<br>(8) | 100<br>(105) |
| JL4 | 67.5<br>(54) | 12.5<br>(10) | 20<br>(16) | 100<br>(80) |

When an NV is not perceived, a postvocalic nasal consonant (V + N) is the most frequent perception, i.e., 19.5% (66/335). Nevertheless, there is an important inter-individual variation. The perceived production of JL1 contains fewer postvocalic nasal consonants (V + NC) than JL3' (11 vs 30.5%). Expected nasal vowels are frequently perceived as oral vowels by both raters in JL4's production (20%), whereas they are absent in JL2's.

6.1.2. Second Hypothesis: Pronunciation of V + N Is Less Frequent at T2

The second hypothesis is developmental and predicts a trade-off effect: the impact of orthography on pronunciation decreases when French oral exposure increases. Out of the 247 (53%) expected vowels in the first recording and 220 (47%) in the second, both raters agreed in perception on 181 vowels at T1 and 154 at T2. If we look, in particular, at the rate of perceived postvocalic nasal consonants between T1 and T2, it decreases overall from 23% to 16% (Table 7).

**Table 7.** Part of postvocalic nasal consonants in % and raw values at T1 and T2.

| | **T1 n = 181** | | **T2 n = 154** | |
|---|---|---|---|---|
| All | 23 | (41/181) | 16 | (25/154) |
| JL1 | 2 | (1/44) | 24 | (7/29) |
| JL2 | 26 | (13/49) | 11 | (3/28) |
| JL3 | 43 | (21/49) | 19.5 | (11/56) |
| JL4 | 15.25 | (6/39) | 10 | (4/41) |

In three of the four participants (JL2–4), the nasalised pronunciation vowel + N decreases, while it increases in JL1. A proportion of 43% of the phonemes produced by JL3 are perceived as vowels with a nasal consonant at T1 but only 19.5% at T2. The opposite trend is observed for vowels produced by JL1: only 2% are perceived as followed by a nasal consonant at T1 compared to 24% at T2.

#### 6.1.3. Third Hypothesis: The V + N Pronunciation Is More Frequent in Retrieved Words Compared to Repeated Words

If we consider the third hypothesis, the nasal consonantisation should be more frequent in retrieved than in repeated words.

If we look into the production of the postvocalic /n/, this tendency seems to be confirmed. A proportion of 77% (51/66) of the V + NCs are pronounced in spontaneous speech (Table 8).

**Table 8.** Distribution of the perceived postvocalic nasal consonants in %.

| | **Retrieved** | **Repeated** |
|---|---|---|
| V + N<br>n = 66 | 77.5<br>(51) | 22.5<br>(15) |

In retrieved words (n = 188), the postvocalic /n/ is present in 27% (51/188) of the occurrences, while it is only present in 10% (15/147) of the repeated occurrences. As a comparison, NVs are pronounced in 64% (121/188) of the spontaneous productions and in 82% (120/147) of the repeated occurrences (Table 9). A chi-square test of independence reveals that there is a significant difference between the retrieved and repeated words ($X2(2) = 15.425$, $p < 0.001$). The pronunciation of French nasal vowels seems better in the repetition than the spontaneous context.

**Table 9.** Distribution of the perceived postvocalic nasal consonants (N = 335) in %.

| | **Retrieved n = 188** | **Repeated n = 147** |
|---|---|---|
| NV | 64<br>(121) | 82<br>(120) |
| V + N | 27<br>(51) | 10<br>(15) |
| OV | 9<br>(16) | 8<br>(12) |

The repetition of a spoken model leads to the pronunciation of a nasal vowel. On the contrary, retrieving a word from the lexicon leads to a consonantised pronunciation of the nasal vowel.

6.1.4. Fourth Hypothesis: Pronunciation of the Nasal Consonant following the Vowel

The fourth hypothesis is that graphemic consistency has an impact on the pronunciation of the nasal vowels. According to Planton (2014) [56], a nasal vowel should be less consonantised when it is more consistent. Therefore the nasal vowel /ɔ̃/ should be less consonantized because it has fewer graphemic correspondences than the other nasal vowels (/ɑ̃/ and /ɛ̃/).

Overall, 19.5% of the expected nasal vowels are perceived as consonantised vowels and the nasal consonantisation depends on the type of vowel. Among these, 7% of instances of the nasal vowel /ɛ̃/ are perceived as a vowel followed by a nasal consonant by both raters, whereas 25% and 23% of the nasal vowels /ɔ̃/ and /ɑ̃/ are perceived as consonantised (TAble 10). Nasal consonantisation depends of the type of nasal vowel but not as previous studies hypothesised.

**Table 10.** Distribution of perceived postvocalic nasal consonants according to the type of nasal vowel.

| Type of nasal vowel | /ɔ̃/<br>n = 123 | /ɑ̃/<br>n = 128 | /ɛ̃/<br>n = 84 |
|---|---|---|---|
| V + N perception (%) | 25<br>(31) | 23<br>(29) | 7<br>(6) |

*6.2. Orthographic Use of Words with Nasal Vowels*

The written texts are quite short and contain 53 syllables that should be pronounced as a nasal vowel and written with a graphic vowel and the <n> or <m> grapheme.

Almost all syllables (54 out of 55, i.e., 98%) are written as expected bigraphs (vowel + the correct nasal consonant <n> or <m>) (Table 11). In a few cases, the vowel in the bigraph is misspelled : <pan> for <pain> (6/8 occurrences), <en> for <o> (2/2), <en> for <un> (1/5) and <numbreux> for <nombreux> (1/2). The only occurrence where the bigraph (vowel + nasal consonant) is not respected is <a> for the indefinite article <un>.

**Table 11.** Orthographic corpus size.

| **ID** | **<n>-Segments**<br>**T1 + T2** | **T1** | **T2** |
|---|---|---|---|
| All | **53** | 17 | 36 |
| JL1 | 21 | 7 | 14 |
| JL2 | 17 | 7 | 10 |
| JL3 | 12 | 2 | 10 |
| JL4 | 3 | 1 | 2 |

Some of these misspellings could be explained by crosslinguistic influence, either the use of the Japanese L1' word /pan/, which is closely related to the French word /pɛ̃/ and <pain>, as in 6a, or the use of the English (L2) indefinite monophonemic article <a> as in 6b. In 6c, the spelling <numbreux> may be explained by the words <numerous> having the same meaning as <nombreux> but starting with <num> instead of <nom>.

(6) a. Elle prend un *pan.
She take.3SG a pan.JAP?
'She takes a pan.'

b. Une femme veut *a pain.
A.FEM woman want.3SG a.EN? bread.
'A woman wants un bread.'

c. Il mange *numbreux déjunéur à réstrant.
He eat.3SG numberous.EN? lunch in restrant.
'He is eating a lot of dishes in the restaurant.'

The analysis of this small corpus of written retellings provides evidence of the stability of orthographic representations of words containing nasal vowels, which allows us to conclude that these are overall target-like. Even at T1, the expected bigraphs are used. Most of the misspellings illustrate the multilingual competence of the participants and can be explained by crosslinguistic effects from their L1 or L2.

## 7. Discussion

To sum up, the results of this analysis of expert listeners perceptions show that (i) beginner Japanese learners of L3 French spell French nasal vowels with an <n> or <m> in 98% of the obligatory contexts; (ii) in speech, most expected nasal vowels are perceived by expert listeners from the initial stages as nasal vowels (72%), the others being perceived as vowels followed by a nasal consonant (19.5%) and as oral vowels (8.5%); (iii) consonantisation is stronger when the learner spontaneously produces a word than when (s)he repeats it; and (iv) it decreases with time (learning effect) and (v) varies according to the consonant, /ɛ̃/ being less consonantised than /ɔ̃/ and /ɑ̃/.

The state of the art reveals that overall, researchers have studied the influence of previously acquired languages on the pronunciation of nasal vowels without necessarily taking into account the effects of orthography. Even if these effects are suggested by the findings of those studies, they were not at the center of the investigation, hence Detey et al.'s (2010) call to explore the effect of orthography on the acquisition of L2 French nasal vowels [9]. Particularly, in research on French as a second language, crosslinguistic phonological influences and orthography effects are poorly articulated. In fact, the phenomenon of nasal consonantisation observed in L2 French is either attributed to the phonology and phonetics of the first languages or to the orthography of the newly acquired language. In our study, we cannot decide between the two possibilities because the data do not allow us to distinguish between the effects of L1 Japanese phonology and those of L3 French orthography, except for supposed new words that the learner hears and uses orally during interaction (without exposure to their written forms). Nevertheless, we hope to have provided answers regarding whether Japanese learners of L3 French produce nasal or nasalised vowels in the initial stages of acquisition specifically in an ecological situation of interaction with an expert speaker.

Before discussing the results, it is noteworthy from previous studies that the nasal consonantisation in the production of L2 French nasal vowels is mainly observed in learners whose L1 lacks nasal vowels and has nasalised vowels, like English, Spanish, Japanese and Cantonese. Conducting comparative studies with learners whose L1 has nasal vowels would bring insights into whether L2 nasal vowel' production is facilitated by the existence of equivalent phonemes in the L1. If not, the use of a postvocalic consonant could be explained by a common path of development in initial stages: first, the use of a vowel followed by a consonant, then a nasal vowel. Such a system would make it possible to distinguish between L1, phonological development in L2 and even the influence of another L2. In any case, it would be interesting to also consider the influence of other languages previously acquired by the learner, like L2 English. Indeed, the possible influence of English on the appropriation of the phonological system of L3 French cannot be excluded, especially given that French and English share a common writing system and a number of similar words (e.g., those ending in 'tion', like construction). Historically, English has borrowed a good proportion (varying between half and a third depending on the authors) of its vocabulary from French [67]. This explains a set of homographs but different grapheme-to-phoneme correspondence rules, i.e., for a certain number of graphemes orthographically common to both languages (the <an> of <restaurant>) pronounced using nasalised vowels in English and nasal vowels in French. This lexical proximity could favour a nasalised pronunciation of the nasal vowel in L3 French. A comparison of the pronunciation of nasal vowels in L3 French by various speaker profiles (Group A: L1 with nasal vowel, L2 English, L3 French; Group B: L1 with nasalised vowel, L2 English, L3 French) would make

it possible to measure, all other things being equal, the impact of the nasalised vowel in L1 learners knowing L2 English.

Furthermore, the comparative studies of nasal vowel acquisition in L2 French carried out to date have mainly focused on single-speaker tasks (reading aloud or repetition of sentences or words) conducted, for acoustic reasons, in a quiet room. From a didactic perspective, one limitation of these studies is that the machines produce diagnoses that are partly inaccessible to human perception, which generates results that are difficult to transpose into pedagogical scenarios. To the best of our knowledge, there are no studies on conversational or narrative data by learners of French in ecological interaction situations. In this respect, our study makes it possible to account for pronunciation in situations close to typical classroom conversational activities. Moreover, the data illustrate the different production possibilities in interaction for beginner learners, i.e., the repetition of words given with the expert speaker's support or the spontaneous production of a word retrieved from the mental lexicon without mediation by the expert. Another advantage of analysing pronunciation in verbal interactions is that it allows us to account for the intelligibility of speech, understood as the ability to identify the lexical units produced based on, among other things, phonemes' recognition.

Moreover, given that this is a longitudinal and bimodal study, it has provided insights into different phenomena involved in the production of nasal vowels. In fact, our findings suggest that, despite the fact that previously acquired languages lack nasal consonants, the L3 French Japanese learners produce, from the earliest stages, phonemes that are identified by expert listeners as nasal vowels in 72% of cases and as nasalised vowels in 19.5% of cases. The first hypothesis claims that the graphemes <n> or <m> influence the pronunciation of the nasal vowels and give way to the use of a vowel followed by a nasal consonant, as observed in other studies [4,9,11]. Since learners produce phonemes that are identified as nasal vowels in two-thirds of the obligatory cases, this consonantisation hypothesis is thus rejected. This figure is higher than the rate of transcription accuracy observed in Detey et al.'s experimental study with more advanced learners (B2–C1) [9]. French non-expert listeners had to hear words pronounced by advanced learners and write them orthographically. The accuracy rate was 64.5%. The analysis of nasal vowels in L2 speech in interaction differs from the results observed experimentally, since a larger part of L2 phonemes are identified as nasal vowels in the initial stages. These differences confirm the already mentioned importance of varying the tasks, particularly from a didactic perspective.

But can we completely rule out an effect of spelling on the production data? Such an influence is possible, but as an anonymous reviewer mentioned it, an equally likely possibility is that the postvocalic nasal consonant results from mistiming of the oral and nasal (i.e., velum closing) gestures. One way of looking at the effect of spelling is to compare cases of oral production with and without spelling knowledge of the word. Spontaneous conversation at the initial stage in the L2 presents these conditions because the learners use the lexicon they learned in its written and oral forms and because the interlocutor constantly provides the lexicon orally at the learner's request, enabling him/her to reuse it and elaborate his/her discourse. The comparative analysis of words and nasal vowel perception in these two contexts confirms an effect of spelling on the oral production process (H3), which corroborates the differences observed in other studies [9,46], according to which word repetition leads to a better pronunciation of nasal vowels than reading aloud because repetition is a phenomenon that does not call up the mental lexicon and the interference of orthographic representations in phonological encoding. Consonantisation is more frequent in spontaneous production when an orthographic representation of the word can be assumed, even though it remains marginal compared to the production of a nasal vowel. Most of the postvocalic nasal consonants are pronounced in spontaneous speech (51/66), indicating a possible impact of the written form of the retrieved word from the mental lexicon. It is sometimes possible to follow this orthographic influence when the learner repeats the word given by his interlocutor and understands which word is

pronounced by the interviewer, then retrieves it from his/her mental lexicon, adjusting the pronunciation to the following nasal consonant (see the following examples 7 and 8).

(7) JL3, Cat story.

a. JL3: boy
b. INT: OUI un GARÇON GARÇON *(YES a BOY BOY*)
c. JL3: /garson/ ah:: OUI *(BOI ah yes*)
d. INT: garçon *(boy)*
e. JL3: /garson/ *(boi*)

In (7), JL3 first spontaneously repeats the word *garçon*, pronounced [garsɔ̃], with the nasal vowel given by his interlocutor in the oral retelling task, then understands the word and reproduces the word [garsɔn] and thus pronounces all the graphemes in the word. This behavior seems to indicate a retrieval of the orthographic form in the mental lexicon in order to understand which word has been given by the interviewer. The retrieved visual form of the word (inner vision) then impacts the following pronunciation of the word, and this is observed in all of the recordings (in total, 25 occurrences), even though [garsɔ̃] with a nasal vowel is used at two other times by the interviewer. Sometimes, when the learner pronounces a word that has the same graphic form in English and French, like <restaurant>, and uses, in French, the postvocalic nasal consonant, as in (8), it is difficult to know whether this is a direct phonological influence of English or if the learner encodes the homographic form phonologically with the English grapheme-to-phoneme correspondence rules. Analysis of the production of nasal vowels in an interactional context makes it possible to isolate a few very clear cases of the influence of a phonological decoding mechanism by mental graphic transcription and registration in the multilingual repertoire. This is the case in (8) in the oral retelling task. Initially, in 8a, the interlocutor provides the word 'poisson' (fish) [pwasɔ̃] to the learner, who does not know it and repeats it identically. Then, after a few statements, in 8b, she describes the final image in which the child finds his cat with a fish in its mouth. The narrator then retrieves the word 'fish' given earlier, which we consider to be a spontaneous use, and produces a transformed series of phonemes /pɔj/ /pɔjszən/ /pɔjzən/, which cannot be explained by the translinguistic influence of English alone.

(8) a. i. INT: ça c'est /pwasɔ̃/ *(this it's fish)*
ii. JL4: /pwas/, /pwasɔ̃/ a /pwasɔ̃/

b. i. JL4: and euh: da /pɔj/ /pɔjszɔn/ /pɔjzn/ *(and euh da fri frish fris)*
ii. INT: /pwasɔ̃/ *(fish)*
iii. JL4: /pwasɔ̃/ une /pwasɔ̃/ *(fish a fish)*

This English-like pronunciation is only made possible by the existence of a mental graphic form made up of several graphemes. The interaction thus reveals an automatic processing of mental transcription. As in the study by Detey et al. (2010) [9], orthography is one factor among others influencing phonology in the L3. Particularly, the task also seems to affect the production of nasal vowels by learners or their identification by listeners.

Furthermore, our study shows that it makes sense from the beginner stages onward to postulate an effect of orthography and, in particular, of the graphemes <n> and <m> because the analysis of written productions shows that learners use them correctly in 98% of cases (H5). The written forms of the words are present from the very beginning in instructed contexts and could favour spelling and influence L2 pronunciation [37,68]. This may be because of the 'transparent' storage of an opaque language like French in the mental lexicon (one grapheme = one phoneme). This may also have the positive consequence of retaining silent consonants in orthographical memory; in other words, the association of /dans/ with <dans> instead of /dã/ leads to a better spelling of the word. This is an argument in favour of teaching suggestions tailored to the real needs of learners. But if mastery of oral language takes precedence over mastery of written language, then other suggestions may apply, like learning the International Alphabetic Alphabet, even if experimental studies exploring the effect of unfamiliar orthographic forms

on word learning and syllable discrimination have shown contradictory results (positive effect [28,44], no effect [45] or a negative effect [53,54]). More research on this topic would help to understand what is going on at the grapheme/phoneme interface.

Above all, our analyses have shown that nasal consonantisation decreases over time for most learners (H2), and this may be explained by greater exposure to nasal vowels and better control of orthographic interference over time. This is a real contribution to research on L3 French pronunciation, given the absence of longitudinal studies.

Lastly, this study has shown differences in consonantisation between vowels in a direction that does not correspond to the predicted one: the nasal vowel /ɔ̃/ is the most consonantised, although it has the fewest orthographic corespondents (H4). Nevertheless, there are 22 of them, and at this stage of irregularity, the difference between nasal vowels may not be very significant. Following the example of Detey et al. (2010), another possible study would be to test phonotactic hypotheses, i.e., whether the position of the vowel in the word (initial, medial or final) and its consonantal environment modifies its pronunciation. As this parameter was not controlled, it could explain the differences observed between vowels.

Our study has mainly shown that lexical or grammatical items with an expected nasal vowel produced by Japanese learners of L3 French were identified by listeners of interactive discourse, regardless of the phonemes perceived. Whether the listeners perceive a nasal vowel, a nasalised vowel or an oral vowel instead of an expected nasal vowel, the word in which this or these phonemes are found is identifiable in speech. In other words, the nasalised or even oralised pronunciation of the nasal vowel does not affect intelligibility, even in the initial stages, after a few dozen hours of exposure to French. Several explanations may be put forward, starting with the predominance of target pronunciation (72%), which makes words interpretable. In addition, the words are produced in a context which contributes to meaning construction. Furthermore, the interactional situation, particularly during the image description task, creates a shared experience where a certain number of words can be interpreted with reference to the context. Finally, even though nasalised vowels are not part of the phonological inventory of French, it cannot be ruled out that the nasal feature conveyed in the nasalised vowel is helping the expert listener to reconstruct the nasal vowel and identify the word produced. However, according to previous studies, these results raise questions about learners' perception and articulation, as well as about the role of phonotactic constraints. Indeed, Marquez Martinez's (2016) perceptual study shows that English speakers adopt perceptual strategies when they are naïve or beginners. The results of her study indicate that naïve listeners mostly heard French nasal vowels as sequences of oral vowel+nasal consonant (perceptual unpacking). But after exposure to French instruction, learners initially heard French nasal vowels as oral, thus applying the nasal stripping strategy. Keeping nasality in the vowel and adding a residual nasal consonant is the third stage of development according to Marquez-Martinez (2016). The fact that the beginner learners studied are at stage 3 of perception may be surprising, and it would be interesting, in a future study, to compare ecological oral production data and perception data for the same sample of learners in order to understand whether the perception of nasal vowels in learner speech by expert speakers is correlated with a perception of the inherent nasality feature by these learners and a target articulation of nasal vowels or whether it is a matter of phonological reconstruction of the nasal vowel by expert listeners in the speech context. Further research is needed to better understand the relationship between perception, production and intelligibility.

## 8. Conclusions

This longitudinal study conducted on ecological and bimodal data contributes to an articulation of phonological and orthographic explanations of pronunciation in L3 French and to a better understanding of the interaction of these effects in production tasks, which are close to conversational activities carried out in the language classroom. Our results provide, with intelligibility in mind, to insights into how to adjust teacher

feedback and to adapt activities to different learner profiles in the language classroom. The fact that mispronunciation of nasal vowels does not interfere with the intelligibility of the words is an incentive to evaluate pronunciation in conversational contexts. In this respect, this study has shown that L2 speech in interaction is an interesting context in which to investigate pronunciation. The classroom may also be an interesting laboratory to investigate crosslinguistic influences at the interface between phonology, orthography, lexicon and discourse. However, these results are based on a corpus of four speakers' productions, and individual differences were observed in terms of nasalised pronunciation frequency and development over time. The intelligibility of nasalised pronunciation of nasal vowels in context therefore needs to be investigated in a larger sample. Similarly, the higher frequency of nasalised pronunciation in retrieved words than in repeated words is an interesting clue as to the role of orthography in the lexical production-and-retrieval process, but would require confirmation in a separate more controlled study.

**Author Contributions:** Conceptualisation, C.S., I.S., C.G. (Cecilia Gunnarsson), V.S. and C.G. (Cyrille Granget); methodology, I.S., C.G. (Cecilia Gunnarsson), C.G. (Cyrille Granget); validation, I.S. and C.G. (Cecilia Gunnarsson); formal analysis, I.S. and C.G. (Cecilia Gunnarsson); resources, C.G. (Cyrille Granget); data curation, C.A., I.S., C.G. (Cecilia Gunnarsson) and C.G. (Cyrille Granget); writing—original draft preparation, C.S., C.G. (Cecilia Gunnarsson), V.S. and C.G. (Cyrille Granget); writing—review and editing, C.S., C.G. (Cecilia Gunnarsson), I.S. and C.G. (Cyrille Granget); project administration, I.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The data were collected in a Japan–France Integrated Action Program (Sakura Program) with the financial support of the Fonds pour la Recherche Scientifique (France) and the Japanese Society for the Promotion of Science (JSPS, Japan).

**Institutional Review Board Statement:** Not applicable.

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

**Data Availability Statement:** The Cat Story used for oral retellings is free available in the IRIS deposit, https://www.iris-database.org/details/7i0ag-5qK6e (accessed on 1 December 2023).

**Acknowledgments:** We wish to thank Fumiya Ishikawa (Rykkyo University), Dominique Klingler (Avignon University), Ako Nishimura (Shirayuri University) and Pascale El Haj for their help in collecting and transcribing the data. We would like to express our sincere gratitude to the reviewers who provided us with valuable feedback on a preliminary version of this article. Any errors are ours, of course.

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

## References

1. Bassetti, B. Orthographic input and second language phonology. In *Input Matters in SLA* ; Thorsten, P., Young-Scholten, M., Eds.; Multilingual Matters: Bristol, UK, 2008; pp. 191–206. [CrossRef]
2. Bassetti, B.; Escudero, P.; Hayes-Harb, R. Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist.* **2015**, *36*, 1–6. [CrossRef]
3. Hayes-Harb, R.; Barrios, S. The influence of orthography in second languagephonological acquisition. *Lang. Teach.* **2021**, *54*, 297–326. [CrossRef]
4. Marquez Martinez, M.A. The Acquisition of French Nasal Vowels: From First Language Allophony to Second Language Phonolgical Contrast. Ph.D. Thesis, Indiana University, Bloomington, IN, USA, 2016.
5. Montagu, J. L'articulation labiale des voyelles nasales postérieures du français: Comparaison entre locuteurs français et anglo-américains. In *Actes des XXIVè Journées d'Etudes de la Parole (JEP)*; Association Francophone de la Communication Parlée: Nancy, France, 2002; pp. 253–256.
6. Delvaux, V. Contrôle et Connaissance Phonétique: Les Voyelles Nasales du Français. Ph.D. Thesis, Université de Mons, Mons, Belgium, 2003.
7. Delvaux, V.; Demolin, D.; Soquet, A.; Kingston, J. La perception des voyelles nasales du français. In *Actes des XXVè Journées d'Etudes de la Parole (JEP)*; Association Francophone our la Communication Parlée: Fès, Morocco, 2004; pp. 54–57.
8. Dherbey Chapuis, N. Enseignement Ciblé Implicite ou Explicite: Etude Comparative de L'apprentissage des Représentations Phonétiques et Phonologiques et du Vocabulaire Associé. Ph.D. Thesis, Université de Fribourg, Fribourg, Switzerland, 2021.

9. Detey, S.; Racine, I.; Kawaguchi, Y.; Zay, F.; Buehler, N. Evaluation des voyelles nasales en français L2 en production: De la nécessité d'un corpus multitâches. In *Actes du Congrès Mondial de Linguistique Française, New Orleans, USA*; EDP Sciences: Paris, France, 2010; 13p. [CrossRef]
10. Detey, S.; Racine, I. Does perception precede production in the initial stage of French nasal vowel quality acqusition by Japanese learners? A corpus-based discrimination experiment. In Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS2015), Glasgow, UK, 10–14 August 2015; p. 894.
11. Li, J.; Yin, Y.; Pu, Z. Acquisition des voyelles nasales du français et interférences du cantonaisn [Acquisition of French Nasal Vowels and Interference of Cantonese]. *Rev. Linguist. Didact. Langues* **2019**, *59*. [CrossRef]
12. Solier, C.; Perret, C.; Baqué, L.; Soum-Favaro, C. Written training tasks are better than oral training tasks at improving L2 learners' speech production. *Appl. Psycholinguist.* **2019**, *40*, 1455–1480. [CrossRef]
13. Kamiyama, T.; Detey, S.; Kawaguchi, Y. Les japonophones. In *La Prononciation du Français Dans le Monde, du Natif à L'apprenant*; Detey, S., Eychelle, J., Kawaguchi, Y., Racine, I., Eds.; Clé International: Paris, France, 2017; pp. 155–162.
14. Maeda, S. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, Nasalization, and the Velum*; Academic Press: Cambridge, MA, USA, 1993; pp. 147–167.
15. Berri, A.; Pagel, D. Réalisation des voyelles nasales du français par des locuteurs brésiliens. *Trav. L'Institut Phon. Strasbg.* **2003**, *33*, 105–117.
16. Peperkamp, S. La perception de la parole et l'acquisition de la phonologie. *Rééducat. Orthophoniq.* **2007**, *229*, 3–16.
17. Takeuchi, K.; Arai, T. *Strategy for the Production of French Nasal Vowels by Japanese Students*; PTLC: London, UK, 2007.
18. Billières, M. Codage Phonique et Boucle Articulatoire en Mémoire de Travail. *Corela* **2005**, *HS-1*. Available online: http://corela.revues.org/1110 (accessed on 1 December 2023).
19. Solé, M.-J. Phonetic and Phonological Processes: The Case of Nasalization. *Lang. Speech* **1992**, *35*, 29–43. [CrossRef]
20. Liddiard, B. Does nasal development have an order? A study of English Speakers learning French. In *Accent, Intonation and Phonological Models*; Ferguson, B., Gezundhait H., Phillips, M., Eds.; Éditions Mélodie: Toronto, ON, Canada, 1994; pp. 55–70.
21. Paradis, C.; Prunet, J.-F. Nasal Vowels as Two Segments: Evidence from Borrowings. *Language* **2000**, *76*, 324. [CrossRef]
22. Sauzedde, B. Difficulté des phonèmes vocaliques du français auprès des étudiants japonais. *Rev. Jpn. Didact. Fr.* **2014**, 9, 97–112.
23. Sauzedde, B. Acquisition de la Structure Syllabique du Français par des Apprenants Japonais. Unpublished Ph.D. Thesis, University of Nantes, Nantes, France, 2018. Available online: https://www.theses.fr/2018NANT2039 (accessed on 1 December 2023).
24. Petersen, S. Categorical nasal vowel acquisition in L2 French learners. *Univ. Pa. Work. Pap. Linguist.* **2015**, *21*, 25. Available online: http://repository.upenn.edu/pwpl/vol21/iss1/25 (accessed on 1 December 2023).
25. Detey, S.; Nespoulous, J.L. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* **2008**, *118*, 66–81. [CrossRef]
26. Simon, E.; Van Herreweghe, M. The Relation between orthography and phonology from different angles: Insights from Psycholinguistics and Second Language Acquisition. *Lang. Speech* **2010**, *53*, 303–306. [CrossRef]
27. Escudero, P.; Hayes-Harb, R.; Mitterer, H. Novel second-language words andasymmetric lexical access. *J. Phon.* **2008**, *36*, 345–360. [CrossRef]
28. Showalter, C.E.; Hayes-Harb, R. Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second. Lang. Res.* **2013**,*29*, 185–200. [CrossRef]
29. Escudero, P.; Simon, E.; Mulak, K.E. Learning words in a new language: Orthography doesn't always help. *Biling. Lang. Cogn.* **2014**, *17*, 384–395. [CrossRef]
30. Escudero, P.; Wanrooij, K. The effect of L1 orthography on non-native vowel perception. *Lang. Speech* **2010**, *53*, 343–365. [CrossRef]
31. Stoehr, A.; Martin, C.D. Orthography affects L1 and L2 speech perception but not production in early bilinguals. *Biling. Lang. Cogn.* **2022**, *25*, 108–120. [CrossRef]
32. Rafat, Y. The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative. *Appl. Psycholinguist.* **2015**, *36*, 43–66. [CrossRef]
33. Steele, J. Assessing the role of orthographic versus uniquely auditory input in acquiring new L2 segments. In Proceedings of the 7th Rencontres Internationales du Réseau Français de Phonologie, Aix en Provence, France, 2–4 June 2005.
34. Bassetti, B. Orthography affects second language speech: Double letters and geminate production in English. *J. Exp. Psychol. Learn. Mem. Cogn.* **2017**, *43*, 1835–1842. [CrossRef]
35. Bassetti, B.; Sokolović-Perović, M.; Mairano, P.; Cerni, T. Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Lang. Speech* **2018**, *61*, 577–597. [CrossRef]
36. Bassetti, B.; Atkinson, N. Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Appl. Psycholinguist.* **2015**, *36*, 67–91. [CrossRef]
37. Cerni, T.; Bassetti, B.; Masterson, J. Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Front. Commun.* **2019**, *4*, 31. [CrossRef]
38. Bassetti, B. Orthographic effects in the phonetics and phonology of second language learners and users. In *Cambridge Handbook of Bilingual Phonetics and Phonology*; Amengual, M., Ed.; Cambridge University Press: Cambridge, UK, 2023.
39. Bürki, A.; Welby, P.; Clément, M.; Spinelli, E. Orthography and second language word learning: Moving beyond "friend or foe?" *J. Acoust. Soc. Am.* **2019**, *145*, 265–271. [CrossRef]

40. Rafat, Y.; Stevenson, R.A. Auditory-orthographic integration at the onset of L2 speech acquisition. *Lang. Speech*, **2019**, *62*, 427–451. [CrossRef]
41. Stoehr, A.; Martin, C. D. The impact of orthographic forms on speech production and perception: An artificial vowel-learning study. *J. Phon.* **2022**, *94*, 101180. [CrossRef]
42. Welby, P.; Spinelli, E.; Bürki, A. Spelling provides a precise (but sometimes misplaced) phonological target. Orthography and acoustic variability in second language word learning. *J. Phon.* **2022**, *94*, 101172. [CrossRef]
43. Sokolović-Perović, M.; Bassetti, B.; Dillon, S. English orthographic forms affect L2 English speech production in native users of a non-alphabetic writing system. *Biling. Lang. Cogn.* **2020**, *23*, 591–601. [CrossRef]
44. Hayes-Harb, R.; Cheng, H.W. The Influence of the Pinyin and Zhuyin writing systems on the Acquisition of Mandarin word forms by native English speakers. *Front. Psychol.* **2016**, *7*, 785. [CrossRef]
45. Pytlyk, C. Shared orthography: Do shared written symbols influence the perception of sounds? *Mod. Lang. J.* **2011**, *95*, 541–557. [CrossRef]
46. Ventura, P.; Morais, J.; Pattamadilok, C.; Kolinsky, R. The locus of the orthographic consistency effect in auditory word recognition. *Lang. Cogn. Process.* **2004**, *19*, 57–95. [CrossRef]
47. Bassetti, B. *Effects of Orthography on Second Language Phonology: Learning Awareness, Perception and Production*, 1st ed.; Routledge: London, UK, 2023. [CrossRef]
48. Derwing, T. The role of phonological awareness on language learning. In *The Routledge Handbook of Language Awareness*; Garrett, P., Cots, J.M., Eds.; Routledge: London, UK, 2017; pp. 339–353.
49. Levis, J.M. *Intelligibility, Oral Communication, and the Teaching of Pronunciation*; Cambridge University Press: Cambridge, UK, 2018.
50. Colantoni, L.; Escudero, P.; Marrero, V.; Steele, J. Evidence-based design principles for Spanish pronunciation teaching. *Front. Commun.* **2021**, *6*, 639889. [CrossRef]
51. Munro, M.J.; Derwing, T.M. Foreign accent, comprehensibility and intelligibility, redux. *J. Second. Lang. Pronunciation* **2020**, *6*, 283–309. [CrossRef]
52. Bassetti, B.; Cerni, T.; Masterson, J. The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Appl. Psycholinguist.* **2022**, *43*, 683–705. [CrossRef]
53. Mathieu, L. The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Lang. Res.* **2018**, *32*, 145–170. [CrossRef]
54. Showalter, C.E. Impact of Cyrillic on native English speakers' phono-lexical acquisition of Russian. *Lang. Speech* **2018**, *61*, 565–576. [CrossRef] [PubMed]
55. Durand, J.; Eychenne, J. Le français méridional. In *La Prononciation du Français Dans le Monde: Du Natif à L'apprenant*; Detey, S., Racine, I., Kawaguchi, Y., Eychenne, J., Eds.; CLE International: Paris, France, 2016; pp. 34–38.
56. Planton, S. Processus Centraux et Périphériques en Production Écrite de Mots:études Comportementales, en Neuroimagerie Fonctionnelle et par Stimulation Magnétique Transcrânienne. Unpublished Ph.D. Thesis, Université Paul Sabatier, Toulouse, France, 2014.
57. New, B.; Pallier, C.; Brysbaert, M.; Ferrand, L. Lexique 2: A new French lexical database. *Behav. Res. Methods Instrum. Comput.* **2004**, *36*, 516–524. [CrossRef]
58. Vance, T. *Introduction to Japanese Phonology*; State University of New York Press: Albany, NY, USA, 1987.
59. Maekawa, K. Production of the utterance-final moraic nasal in Japanese: A real-time MRI study. *J. Int. Phon. Assoc.* **2023**, *53*, 189–212. [CrossRef]
60. Youngberg, C. Representing the moraic nasal in Japanese: Evidence from Tōkyō, Ōsaka and Kagoshima. *Glossa J. Gen. Linguist.* **2021**, *6*, 63. [CrossRef]
61. Shimamori, R. *Grammaire Systématique du Japonais*; J. Maisonneuve: Paris, France, 2000.
62. Bouchard, R. Le "cours", un événement oralographique structuré: Etude des interactions pédagogiques en classe de langue et au delà. *Le Fr. Monde Rech. Appl.* **2005**, *38*, 64–74.
63. Ishikawa, F. Impact des motivations sur le développement de la L2 en interaction didactique : Représentations en français langue étrangère. *Linguist. Didact. Langu.* **2009**, *40*, 49–69. [CrossRef]
64. Crépieux, G.; Callens, P. *Spirale*; Hachette: Paris, France, 2006.
65. Nakayama, M.; Sugiyama, R. *Bonjour, Paris*; Hakusuisha: Tokyo, Japan, 2002.
66. Detey, S.; Racine, I. Learning to read liaisons in French as a Foreign Language: A longitudinal study of beginner Japanese learners in Japan. *Lang. Interact. Acquis.* **2019**, *10*, 97–117. [CrossRef]
67. Robert, J.-M. L'anglais comme langue proche du français ? *Études Linguist. Appl.* **2008**, *149*, 9–20. [CrossRef]
68. Gunnarsson-Largy, C.; Dherbey, N.; Largy, P. How do L2 learners and L1 writers differ in their reliance on working memory during the formulation subprocess? *Read. Writ.* **2019**, *32*, 2083–2110. [CrossRef]

education sciences

MDPI

*Article*

# An Analysis of the Type of Questions Posed by Teachers in English-Medium Instruction at University Level

**Aintzane Doiz * and David Lasagabaster**

Department of English and German Philology, University of the Basque Country UPV/EHU, 01004 Vitoria-Gasteiz, Spain
* Correspondence: aintzane.doiz@ehu.eus

**Abstract:** Teacher-led questions not only guide meaning-making interactions but they also scaffold students' learning, and this is especially important in English-medium instruction (EMI). Due to the scant literature on this topic in higher education, this article analyses what type of questions EMI history lecturers pose and whether they are subject to individual differences. The study is based on 12 two-hour lectures whose transcriptions were analysed by three researchers. The results showed that instructional or content question types were much more commonplace than regulative questions (related to classroom procedures). Confirmation check, display and referential questions, which belong in the instructional category, were not posed to fulfil their intended pedagogic goals, a limitation accentuated by students' trend to provide short responses. These results reveal the need to design teacher training courses aimed at developing teachers' interactional abilities. Since questioning practices varied considerably between lecturers, customized training sessions should also be considered.

**Keywords:** English-medium instruction; questions; interaction; higher education; teacher training

**Citation:** Doiz, A.; Lasagabaster, D. An Analysis of the Type of Questions Posed by Teachers in English-Medium Instruction at University Level. *Educ. Sci.* **2023**, *13*, 82. https://doi.org/10.3390/educsci13010082

Academic Editor: Martin Howard

Received: 2 December 2022
Revised: 9 January 2023
Accepted: 10 January 2023
Published: 12 January 2023



## 1. Introduction

Teacher-led questions play a paramount role when it comes to boosting students' comprehension of subject matter, and this irrespective of the teaching language (be it the first language (L1) or a foreign language). Teachers ask questions to guide meaning-making interactions and to scaffold their students' learning. This practice helps narrow the gap between students' actual knowledge and the knowledge they are expected to gain in collaboration with the lecturer, who we presume to be more knowledgeable regarding the subject and who should facilitate this process. As Kawalkar and Vijapurkar [1] put it, teacher questions are important because they affect the nature of students' thinking and reasoning while they determine the quality and the level of students' participation. That is, they can become indices of quality teaching. In addition, research has documented that high dialogic teacher talk positively predicts academic outcomes [2]. For all these reasons, the role of questioning exchanges becomes "a fruitful area to explore" [3] (p. 816).

Although research abounds on the impact of asking students questions in L1 and second language (L2) EFL learning contexts, little has been carried out regarding its impact in the field of English-medium instruction (EMI) at university, this paper's context. EMI is "an educational system where content is taught through English in contexts where English is not used as the primary, first, or official language" [4] (p. 114). Since EMI programmes are mushrooming at universities all over the world [5], this is an issue well worth investigating. Our study aims to help fill the gap and it is particularly innovative for the following reasons: (i) Nearly all previous research has taken place in primary and secondary education [3,6–9] whereas little attention has been paid to tertiary education, which is our focus; (ii) Most studies have involved small-group student interaction, whereas whole class teacher/student interaction, which is the focus of our study, has been overlooked [6,10];

(iii) The bulk of research has been in mathematics (see [11] for a systematic review of 15 studies; [12]), business administration [13], and science classrooms [1,3,14–16], while few papers have dealt with humanities, as is the case of the present study.

## 2. Questioning Practices in EMI in Higher Education

Previous EMI research has analysed the role of questions within a broader approach to EMI discourse and teaching. Dafouz et al. [17] delved into *disciplinary reasoning episodes* (DREs) in order to analyse the role played by language-related negotiations in explicit reasoning episodes. The authors concluded that "[t]he benefits of engaging students in question-and-answer formats for the quality of disciplinary reasoning align with findings from L1 science education" (p. 556). Via a questionnaire Suviniitty [18] asked EMI students how teacher questions affected comprehensibility and concluded that lectures with a higher degree of interaction and questions were judged to be easier to understand.

Other studies indicate that students' lack of English skills constrains their willingness to ask questions. Recent research reveals that university students tend to feel uncomfortable when asked to contribute orally in class, but their qualms are even greater in classes delivered in their L2 [19]. Tsou [20] points out that in some contexts, such as Taiwan, EMI students are not linguistically ready to ask questions, and this forces them to consult the teacher privately during the break. The reason is that they are afraid of asking questions in front of the whole class, which is why they are often allowed to ask questions in their L1. In fact, Tsou observed that "the teachers appeared to be used to the lack of interaction because most of the time they answered their own questions without waiting for a response from a student" [20] (p. 83). In Sweden Airey [14] found that students believed they learnt equally well in Swedish and English, but, after watching video footage of actual lectures, they acknowledged that fewer questions were asked and answered in EMI classes. Airey and Linder [15] also observed that among Swedish students the traditional reluctance to ask questions was exacerbated in EMI classes which they find "all the more worrying when we take into account the fact that *lecturers* see a strong correlation between asking questions and student understanding" (p. 556; emphasis in the original).

Sánchez-García [13], who undertook research in Spain, explored two business administration teachers' practices in Spanish-medium and EMI lectures. Sánchez-García distinguished two main question categories: instructional questions (related to the content being learnt) and regulative questions (related to classroom management and organisation). Interestingly the number of questions was largely similar in Spanish (L1) and EMI, although there were almost twice as many questions regarding classroom management in EMI lectures, revealing lecturers' concern about students' understanding of lecture organisation in the latter. Conversely, the use of self-answered questions was three times higher in Spanish, which seems to indicate that these types of questions are not yet part of lecturers' repertoire in English. The author concludes that "teachers may not be asking as many eliciting questions as they often believe they do" (p. 46), which is why there is a need to raise teachers' awareness of the role of questions in students' learning process, as questions could be better exploited to support pedagogical objectives.

Finally, Chang [21] investigated whether disciplinary cultures influenced the patterns of questions in fifteen small-class lectures with no more than 40 students. Chang compared three different disciplines, namely Humanities and Arts, Social Science and Education, and Physical Sciences and Engineering. The results showed "far more similarities than differences between the soft and hard fields with regard to the use of questions in academic lectures" (p. 112), which leads the author to conclude that the influence of genre (the lecture as a genre) outweighed that of disciplinary culture.

The traditional lecture format still seems to be the main mode of teaching at undergraduate level [13,22] and it seems that few studies focus on the role that questions play in this format. Our study aims to help fill that gap. We analyse teacher-fronted questioning during whole-class discussions in history classes delivered by four EMI teachers at a Spanish university. We believe there is a need to know how questions dovetail with the

interactional exchanges that take place in EMI classes, as well as what the most common types of questions used by EMI teachers are. With this analysis we can propose measures to boost balanced and effective teaching practices.

## 3. Research Questions

Based on Sánchez-García's [23] taxonomy of questions (see Table 2 below), this paper addresses the following two research questions:

RQ1: Are there any general tendencies in the types of questions posed by the EMI history teachers?

RQ2: Are the questioning tendencies subject to individual differences among the lecturers?

## 4. The Study

This work is part of a longitudinal research project whose main goal is the study of teacher-student interaction in an EMI context. The study was conducted at the Department of History of the University of the Basque Country in Spain (UPV/EHU). The UPV/EHU is a public university in which classes are taught in Spanish and Basque, the two official languages and, since 2005, in non-official languages, primarily English, under the Multilingualism Program (MP). There are currently more than 750 undergraduate and more than 300 Master's subjects taught through English.

### *4.1. The Participants and the Courses*

In order to recruit participants for our study, we contacted the seven EMI teachers of the Department of History via email and explained the study's objective. Five lecturers showed interest in participating, but one of them was excluded because his classes mainly consisted of student interactions which were not appropriate for this paper's object of study (i.e., teacher-fronted questions). Thus, after the screening, the participants for the study were four male lecturers that will be referred to as T1, T2, T3 and T4. As required by the MP of the UPV/EHU, the four lecturers had an equivalent of the C1 level of the Common European Framework of Reference for Languages (CEFR) in English.

Table 1 contains information on the subjects the lecturers taught, their teaching experience in general and teaching experience in EMI. It also includes information on the twelve lectures analysed (three per lecturer) which were randomly selected from a pool of 29 observed lectures. These lectures, like the rest of the lectures, were primarily teacher-fronted. For ease of reference in the discussion of the data, each lecture has a code number which is provided in the second to last column to the right. The total number of words uttered in the lessons per teacher is given in the last column. This figure refers to the actual words uttered, including repetitions of the same word, but unfinished words were not taken into account. Each of the lectures lasted approximately two hours and the corpus consists of a total of 91,904 words.

**Table 1.** The participants and their classes.

| | Subject | Years of Teaching Experience | Years of Teaching in EMI | Lectures Observed and Recorded | Number of Words Recorded |
|---|---|---|---|---|---|
| **T1** | America in the modern age | 25 | 7 | 1, 4, 16 | 27,105 |
| **T2** | Early modern history I | 16 | 2 | 2, 8, 9 | 21,027 |
| **T3** | World economic history | 21 | 5 | 3, 6, 10 | 27,155 |
| **T4** | Contemporary history of the Basque Country | 30 | 2 | 11, 13, 15 | 16,617 |

The cohorts of students were small, ranging from six to 20 students. The groups were heterogeneous in terms of their level of proficiency in English, as is usually the case in EMI courses irrespective of the country [24,25]. However, the majority of the learners would be at the B2 level of CEFR [26], and homogeneous in terms of cultural background since, with

the exception of two international exchange students per course, the rest of the students were local.

### 4.2. Data Collection and Coding Process

Once the lecturers and the students granted us the necessary permissions, we recorded one lecture every fortnight during the whole semester. We randomly selected three sessions per teacher for analysis. These were then transcribed by a research assistant and were revised by the authors for accuracy. The transcriptions reflect what was said word-for-word in the classes to the best of our and the research assistant's ability. Ungrammaticalities, inaccuracies and repeated words have not been eliminated or resolved (see the nomenclature for the transcription conventions).

The coding process of the questions was as follows. First, the research assistant was instructed to identify all the teacher-fronted questions, that is to say, those instances in which the utterance's intonation pattern and/or its syntactic pattern was that of a question, and to categorise them into one of the eleven categories proposed by Sánchez-García [13,23] and presented in Table 2.

**Table 2.** Taxonomy of questions [13] (p. 32).

| **Instructional Questions (Related to Content)** | |
|---|---|
| Display | Those to which the answer is known by the teacher. |
| Referential | Those to which the answer is not known by the teacher. |
| Repetition | Those seeking repetition of the last word, idea, utterance, etc. |
| Language | Those seeking assistance as regards language matters. |
| Confirmation Checks | Those aimed at ensuring understanding of the topic/lecture. |
| Retrospective | Those which require the students to recall previous information. |
| Self-answered | Those which are immediately answered by the teacher. |
| Rhetorical | Those to which no answer is expected. |
| Indirect | Those which are not uttered to get a response but to exemplify some situation. |
| Regulative questions (related to classroom procedures) | |
| Procedural | Those which refer to the development of the lesson and do not focus on the content/language, but on the lecture itself or a particular activity. |
| Off-task | Those which refer to a topic that departs from the main subject. |

Then one of the authors of this study analysed the questions individually in order to verify that they performed the function that had been assigned to them by the research assistant. There were no significant discrepancies between the research assistant's and the author's classification of the question categories of referential, repetition, language, confirmation check, retrospective, rhetorical, procedural and off-task. However, they both found it difficult to determine whether some of the questions belonged in the display category (i.e., "those to which the answer is known by the teacher") and which were answered by the lecturers themselves or the self-answered category ("those which are **immediately** answered by the teacher, preventing other participants from providing any response", emphasis placed by the authors of this paper) [13] (p. 32). The questions in extract 1 illustrate the categorisation dilemma we faced:

Extract 1: Lecture 1

T1: what is an *audiencia* in # in Spanish # eh # system? There were *audiencias* in Spain # there were were one in Valladolid # there were one in Granada.
T1: what ## which is an *audiencia*?
T1: in Spain # it was # and it is # a xxx tribunal # a high court.

In extract 1, T1 asked what *audiencia* (hearing) was in the past twice and then he provided the answer. As transcribed, the two questions could be classified as display since the lecturer knows the answer. However, alternatively, they could also be referred to as self-answered since the lecturer himself provided the answer.

In order to determine which question type captured what occurred in this and other similar instances, we concluded that it was necessary to know whether the lecturers gave the students a chance to reply; that is to say, whether they gave the students enough wait time, and to include this information in the transcripts. Wait time is the time the teacher waits for the student or students to provide an answer. Following Tobin [27] and Mujis and Reynolds [28], we decided on three seconds or more to be the ideal wait time in which the students can provide their answer. If the lecturer answered his own question immediately after he posed it without allowing any wait time, the question was categorised as self-answered. On the other hand, if the lecturer provided the answer to the question himself after the wait time was over, the question was classified as display. In the case of the questions in extract 1, we noted that T1 did not give the students any wait time, therefore these questions were classified as self-answered. Some other minor discrepancies between the research assistant's and one of the author's classifications of rhetorical and indirect questions occurred and were discussed with the second author until inter-coder agreement was reached. The patterns and tendencies are presented next.

## 5. Results

### *5.1. First Research Question: Are There Any General Tendencies in the Types of Questions Posed by the EMI History Teachers?*

Table 3 presents the distribution patterns of all the teacher-fronted questions found in our data from the most frequent to the least frequent. The data have been normalised to 1000 words. That is to say, the second column of Table 3 refers to the number of questions per 1000 words. The total number of questions for each category is also provided.

**Table 3.** Overall distribution of question categories.

| **Total Number of Words: 91,904** | **‰** | **Number of Questions** |
|---|---|---|
| Confirmation check | 20.44 | 1879 |
| | | |
| Display | 4.11 | 378 |
| Referential | 2.57 | 237 |
| Self-answered | 1.33 | 123 |
| | | |
| Repetition | 0.44 | 41 |
| Retrospective | 0.34 | 32 |
| Procedural | 0.19 | 18 |
| Off-task | 0.18 | 17 |
| Indirect | 0.10 | 10 |
| Rhetorical | 0.02 | 2 |
| Language | 0.01 | 1 |
| TOTAL | 29.79 | 2.738 |

Table 3 reveals that the questions posed by the lecturers fall into three categories: those which are by far the most common, namely, confirmation check questions with an average of 20.44 per 1000 words, a second group of questions with considerably fewer tokens (e.g., display (4.11‰), referential (2.57‰) and self-answered questions (1.33‰)), and finally, a third group of questions with an average lower than 0.44 occurrences per 1000 words,

such as, repetition, retrospective, procedural, off-task, indirect, rhetorical and language questions. Next, we turn to the individual analysis from the most frequently asked type of questions to the rarest ones.

The high number of **confirmation check** questions in our data in comparison with other question categories means that the main goal of the lecturers' interactions with their students was to check comprehension, as clearly illustrated in extract 2, where T1 solicits the students' confirmation that they have understood what he said.

Extract 2: Lecture 1

T1: We can assume that in Spain # or all the provinces # or the *diputaciones* # are # the same ## they have the same basic eh competences # even in # I don't know # in eh # Andalusia # or Cataluña # or whatever # we can assume that.
T1: yes # yes # or not? [con] mm? [con]

These questions were formulated with the expression "Okay?" in 58.22% of all the comprehension check questions, and the interjections "mm?" and "eh?" Other forms such as "Yes or no?", "yes?", "No?", "Any question?", "Have you followed me?", "Have you understood"? were also found, but were less frequent. However, a closer look at the confirmation check questions revealed that many of the questions reflected "the mechanized use of [ . . . ] apparently instinctive structures belonging to the linguistic repertoire of the lecturer as filler expressions" [23] (p. 195), and, in fact, most of the time, the students did not reply to them.

The second most widely used type of questions was **display**, with a quarter of the average of comprehension checks per 1000 words (4.11‰). The display questions of our corpus tended to test the learners' knowledge of specific aspects of the content of the lesson rather than broader aspects, indicating "that students' limited answers are good enough for teacher's questioning purposes and that, in fact, a minimal response is what lecturers are looking for" [23] (p. 199). This is why the lecturers frequently helped the students to formulate their answer by providing the first word of their response (extract 3) or limited the students' answer to the minimum, as the students simply needed to fill in a missing word (extract 4).

Extract 3: Lecture 15

T4: bombing of the civil # of the Spanish civil war and probably one of the most famous: in spite of that # however # the bombing of Guernica is the most famous # in the world. Okay? [con] Why? [dis] (*two students answer at the same time*)
T4: okay # wait wait wait.
T4: well # first *(pointing at S2)* # because? [dis]
S2: it was the first civil bombing.
T4: no # Durango was first.

Extract 4: Lecture 3

T3: and # in this kind of system # the peasants # or the farmers # had to provide the government with? [dis]
S2: labour work? [con]
T3: labour services.
S: eh.
T3: okay. Labour services.

Furthermore, the lecturers' tendency to elicit succinct answers from the students was also observed in the so-called chain questions, in which lecturers produced a succession of questions in which the scope of the question narrowed down from an initially open question to a closed-ended question (see [23] (p. 42), for a similar observation). This is illustrated in extract 5 where the more general question, "What happened with England" is followed by the very specific question "When was England # the major power in Europe?"

Extract 5: Lecture 2

T2: What happened with # eeeh # England? [dis]

T2: When was England # the major power in Europe? [dis]
T2: when? [dis]
S: Seventeen # eighteen.
T2: Seventeen # seventeen century # from the seventeen century # as a result of the first # civil war # remember eeeeh # Cromwell # okay? [con]
T2: and # as a result of the # various # revolution in 1688 # okay? [con]

Whether the students were conditioned by the nature of the lecturers' questions, or it was the students' natural tendency not to intervene with long utterances, students' replies to display questions, and in general to all kinds of questions, were brief [29,30]. Extract 6 is representative of the students' replies, where the contrast between the student's one-word answer on the one hand, and the lecturers' rephrasing of it, on the other, is very illustrative.

Extract 6: Lecture 1

T1: So # which was the decision? [dis]
S: Vicekingdoms.
T1: Vicekingdoms.
T1: So # to create # two figures # of alter egos of the king ### whose seat will be # two American cities # and they will # be # at the top # of that system # eh in # in in in a America.

**Referential** questions are questions whose answer is not known by the teacher [13] (p. 32). In our corpus, their main function was to make the content of the lecture more relevant to the students by several means such as shifting the focus to the students themselves, enquiring about the student's opinion rather than the content itself (extract 7), establishing connections between the lectures and the students' opinions (8), and referring to their personal life (extract 9) and experiences (extract 10).

Extract 7: Lecture 2

T2: Why is important the public debt? [dis]
T2: What do you think it's important? [ref]

Extract 8: Lecture 3

T3: And the problem is # what would happen # if # once you have invested # a lot # of your work # of your human effort # and even of your money in the improvement of your plot # [ . . . ] at the end of the year # the emperor tells # perfect # now # we are going to distribute # the land # for the next year # which is going to begin # right now # what would happen then? [dis]
T3: What would you feel? [ref]
T3: What would you think in that case? [ref]
T3: Wouldn't you feel cheated? [ref]
T3: What would you think? [ref]
T3: Would you like that? [ref]

Extract 9: Lecture 3

T3: So # what do you think # rural industries mean? [ref]
S: eh # a *(??)* industry that provides # em ### xxx to the # to the # rural world.
T3: Not really.
T3: Where do you come from # xxx? [ref]

Extract 10: Lecture 3

T3: Have you ever heard # the expression *(writing on the board)* paddy fields [ref]?
T3: Have you seen # the movies about Vietnam? [ref]
T3: eh? [con]
T3: Have you ever seen # movies such as "Rambo" # "Apocalypse now" # and so on? [ref]

Referential questions were also a means to create a relaxed atmosphere in which the lecturer does not make any assumptions about the students' knowledge, and therefore, any reply is welcome, as illustrated in extract 11.

Extract 11: Lecture 6

T3: Do you know the # theee # history of the discoveries # no # Ana? [ref]
T3: Why they say that Columbus was probably a xxx idiot? [dis]
S: Because # er # he thought he found # India.
T3: Yeah but is more # is even more interesting than that.
T3: You know the # you know # the problems Columbus had with the Portuguese crown? [ref]
T3: You know the history? [ref]
T3: Is really interesting # because Columbus thought . . .

Sánchez-García [23] noted that most of the **self-answered** questions occur within long monologues and serve to "introduce a new topic or explain and deliver new discipline content" (p. 208) and organise the discourse. The self-answered questions we found in our corpus were also part of long teacher-fronted interventions, and fulfilled the functions stated by Sánchez-García. For example, the self-answered question introduced a new concept in extract 12, it served to mark the relevance of some aspect of the message in extract 13, and facilitated student comprehension by breaking the message into chunks that are easier for the students to process in extract 14. Among the self-answered questions, those headed by the interrogative pronoun "why" were quite prevalent (extract 15), which reveals the importance that lecturers attributed to the cognitive discourse function "explain," according to which knowing why a historical event happened is as important as telling what happened [31].

Extract 12: Lecture 4

T1: Eh # what is the astrolabe? [sel]
T1: This is an astrolabe. *(showing a picture)*
T1: xxx device made of xxx metal # eeeeeh # xxx # xxx # object *(??)* # but # always # have these features # with different # eeeh # eh # elements # that are used for what? [sel]
T1: Well # basically # for measuring the time.

Extract 13: Lecture 3

T3: What I am trying # to tell you with this? [sel]
T3: We are # at the end # of # an economy # of a *(¿?)* # kind of Chinese economy # based # in the closely # in the close involvement # of the state # in the economic affairs

Extract 14: Lecture 4

T1: so as you can see # changes in technology # some of them imported from # outside Europe # others # created in Europe # that led to what? [sel]
T1: to huge development # during the fourteenth and sixteenth centuries # in # the art # of ship making # mm? [con]

Extract 15: Lecture 16

T1: to a certain extent # the main # element # of contact # not only contact # of control # to the different empires in the Americas was made at the beginning # through # the activity of piracy.
T1: why? [sel]
T1: because something interesting is that # no # European # country # had # a permanent army # in the Americas # up to the eighteenth century.

The lecturers' questions posed as requests for **repetition** were usually triggered by the physical conditions of the classroom (e.g., street noise, bad acoustics), students not speaking loud enough or the need to clarify what the student had said. The lecturers used the interjections "Eh?", "What?", Mm?", "Sorry?" or rising intonation as an invitation to the student to provide the expression that was not understood (extract 16). Fully fledged interrogative questions were not used probably as a result of the Spanish L1 influence. However, since the students' participation was low, requests for repetitions were rare.

Extract 16: Lecture 10

T3: Let's see the economics students.

S: the consumers # eeeh # xxx better # but # in contrast # eh another countries # eh # thee they can lost their # their own # mm # economy system # like Argentina is a # very good # country # but
T3: very? [rep]
S: is a very # good country.
T3: very? # sorry? [rep]
S: good.
T3: well # yeah.

As stated above, **retrospective** questions were used sparingly, and were asked to revise the course materials, to dwell on the subject content that had already been covered in previous classes (extract 17), and to refer to the general knowledge the lecturer believed the students to have. All the instances of retrospective questions were affirmative statements with the verb *remember* uttered with rising intonation, as opposed to an interrogative yes or no question (e.g., "do you remember . . . ?").

Extract 17: Lecture 11

T4: and # eeeem # well # mm # remember the # the scene # in which # the camera ## lingers # on the # on the # plates? [ret]

**Procedural** questions referred "to the development of the lesson" [13] (p. 32) and included questions posed to: (1) Check whether the students had done the assignment (e.g., watch a video, write an essay; extract 18); (2) Ask volunteers to elicit an answer to a question asked by the lecturer; (3) Pool the students' opinion on the timing for a break during class (extract 19); (4) Ask the students to speak louder; and (5) Ask questions, occasionally in the lecturers' L1, while facing technical problems.

Extract 18: Lecture 11

T4: Watch # the movie? [pro]
T4: and you? [pro]
T4: okay.
T4: have you # written a # paragraph or something # or not? [pro]
S: no.
T4: just # no.
T4: okay only Ricardo? [pro]

Extract 19: Lecture 3

T3: but we will deal with that # eh # in the third part of the lesson # okay? [con]
T3: so # what about a little break? [pro]

Most **off-task** questions were related to time keeping, other courses the students may be enrolled in, and a bit of chit chatting. Questions asked to enquire the students' names were also considered off-task, but since the lecturers did not normally address the students by their names, the use of off-task questions for this purpose was rare (extract 20).

Extract 20: Lecture 16

T1: and actually we are going to start # a long long weekend. [ . . . ] Do you # do you have any # any class # any lesson tomorrow? [off]
S: yes.
T1: yes # ah # all right. What time? [off] what time? [off]
S: three o'clock.

**Indirect** questions, the questions designed to exemplify some situation but which are not expected to obtain a response, were scarce in our corpus. Extract 21 illustrates this kind of question type.

Extract 21: Lecture 16

T1: is like . . . actually # who is the best # what is the best # way # of stopping # the activity # of # such a powerful man # that can ## create # the whole army # and the whole ## navy # not not not army # but navy # against you in the Caribbean? [dis]

T1: well # kill him? [ind]

Similarly, **rhetorical** questions were very rare (0.02‰), and the two tokens we found were posed by one of the lecturers, T3, in the same utterance (extract 22). These questions were used to introduce a new topic for the students, however, given the small sample of the rhetorical questions in the data, these questions may carry out other functions that we are not aware of.

Extract 22: Lecture 10

T3: so why did Britain remain committed # eeeh # to free trade? [rhe]
T3: why did Britain remain # focused # on the promotion of free trade # until 1931? [rhe]
T3: and this is what we are going to address right now.
T3: okay? [con]
T3: and to do that ## we will have to take a glance ## sorry *(going through slides)* ## to this lovely table *(going through slides)* ### which is probably the *Capilla Sixtina* [Sistine Chapel] # of the # economic history temples.

There was one single instance of a language question from the lecturers to the students (extract 23). This occurrence represents 0.01‰ of all the questions in the corpus. Consequently, the lecturers did not seem to face language difficulties in general which may have prompted a question or at least it seems that they had the necessary resources to avoid any language-related issues.

Extract 23: Lecture 13

T4: Not to # eh how do you say this # to *paños calientes*? (the English equivalent to "not to go in for half measures") [lan]

### *5.2. Second Research Question: Are the Questioning Tendencies Subject to Individual Differences among the Lecturers?*

A breakdown of the questions asked by each of the lecturers revealed the existence of individual differences, as shown in Table 4. Due to the variability in word count of the lectures, and in order to facilitate the comparisons among them, the data have been normalised to 1000 words. The number of tokens for each question uttered by the lecturers is provided between parentheses.

**Table 4.** Overall distribution of questions by lecturer.

| | **T1**<br>**27,105 Words**<br>**‰ (Questions)** | **T2**<br>**21,027 Words**<br>**‰ (Questions)** | **T3**<br>**27,155 Words**<br>**‰ (Questions)** | **T4**<br>**16,617 Words**<br>**‰ (Questions)** |
|---|---|---|---|---|
| **Confirmation check** | **10.77 (292)** | **50.83 (1069)** | **13.40 (364)** | **9.26 (154)** |
| | | | | |
| **Display** | **3.06 (83)** | **4.04 (85)** | **6.59 (179)** | **1.86 (31)** |
| **Referential** | 0.81 (22) | 0.99 (21) | **5.81 (158)** | **2.16 (36)** |
| **Self-answered** | **1.62 (44)** | **1.23 (26)** | 1.32 (36) | 1.02 (17) |
| | | | | |
| **Repetition** | 0.07 (2) | 0.23 (5) | 0.92 (25) | 0.54 (9) |
| **Retrospective** | 0.18 (5) | 0.38 (8) | 0.18 (5) | 0.84 (14) |
| **Procedural** | 0.03 (1) | 0.14 (3) | 0.33 (9) | 0.30 (5) |
| **Off-task** | **2.58** (7) | 0.09 (2) | 0.18 (5) | 0.18 (3) |
| **Indirect** | 0.11 (3) | 0.04 (1) | 0.11 (3) | 0.18 (3) |
| **Rhetorical** | | | 0.07 (2) | |
| **Language** | | | | 0.06 (1) |
| **TOTAL** | 16.93 (459) | 57. 87 (1220) | 28.94 (786) | 16.42 (273) |

**T1's** questions were mainly confirmation checks usually formulated with the interjection "mm?" (10.77‰), and display questions as the second most widely used question category (3.06‰). He produced the highest number of self-answered questions of the four lecturers (1.62‰) and of off-task questions (2.58‰), although these questions represented a very small percentage of the total amount of questions. If confirmation checks are disregarded, he asked the least number of questions of all the lecturers.

**T2** asked considerably more questions than the other lecturers. His main trait was the use of confirmation checks, in particular those formulated with the expression "okay?" (50.83‰), although these questions did not receive an answer for the most part. If confirmation checks are disregarded, he used display and self-answered questions the most, like T1. Hence, T1 and T2 showed similar behaviours in the interactions with the students.

**T3** produced twice as many questions as the rest of the lecturers, if the confirmation check questions are disregarded. In particular he used display questions (6.59‰) and referential questions mainly. Repetition questions were somewhat more frequent in his case as he adopted more forward strategies to promote student participation than the other lecturers. In fact, he was the only lecturer who called on specific students by name to respond to a question. Perhaps there is a relationship between the fact that he called on the students by their names and the fact that he asked the most questions, in particular referential questions, which were not content related strictly speaking (e.g., when he asked them about students' personal view on a particular issue).

**T4** asked the lowest amount of questions. If confirmation checks are disregarded, his speech is characterised by the use of referential questions the most and display questions next. He also used procedural questions more than T1 and T2, although they were only 0.30‰ of all his questions. T3 and T4 obtained the highest rates of student participation according to our class observations and the recordings.

## 6. Discussion

The first guiding research question of this study addressed the existence of general tendencies in the questions posed by the EMI history teachers during lessons. We based the analysis on the taxonomy put forward by Sánchez-García [23]. The first conclusion of the study is that, although lecturers did not ask many questions, we did observe some tendencies in their behaviour. In particular, they resorted to instructional question types (i.e., questions related to content) primarily, while regulative questions, those related to classroom procedures such as procedural and off-task questions, were extremely rare (see [13] for similar results in EMI teacher-fronted classes in Madrid). Out of the nine different instructional question categories, confirmation checks were clearly the most widely used question type, followed at a considerable distance by display, referential and self-answered questions. The rest of the instructional question categories posed by Sánchez-García [23] were extremely rare. It is noteworthy that within the regulative questions, procedural questions having to do with technical issues were slightly more likely to elicit the use of the participants' L1, a language which was never used in class. This may be due to an understanding that, as in these situations the content-matter was not the focus of the question, perhaps a relaxation of the use of the L2 was justified, not to mention the propensity to release one's emotions such as the anxiety caused by the technological complications in one's L1.

Two additional conclusions can be drawn from the first research question on the question categories. The second conclusion is that the three most prevalent question categories, namely, confirmation check, display and referential questions, do not seem to be posed to fulfil their intended pedagogic goals. Thus, while it might appear that the lecturers constantly checked student comprehension, most of the confirmation questions were mere mechanical filler expressions devoid of meaning. Most of the display questions in our corpus were closed-ended, or started out as open-ended in a chain of questions that led to a closed-ended question, requiring a simple short answer by the students. Furthermore, despite the lack of complexity of the potential student response and the fact

that the lecturers quite frequently provided cues to the answer, the students' responses to the display questions were short or were not provided, as found in other EMI settings where students' unwillingness to ask questions seems to be exacerbated by linguistic limitations and their fear of speaking in front of the class [14,15,19,20,29]. In addition, finally, referential questions which normally dealt with students' opinions, personal habits and thoughts did not trigger more lengthy contributions in our study either. As stated by Sánchez-García [23] (p. 199) on referential questions:

> more sophisticated, extensive and lengthy contributions would be expected, precisely because these questions address students' personally and allow more assorted and tailor-made replies promoting on many occasions out-of-the-box critical thinking. Surprisingly, students' output remains stagnant concerning referential enquiries and do not entail any observable change in terms of length or verbal complexity.

Hence it appears that confirmation checks, display and referential questions are unlikely to achieve learning gains in our context. The behaviour associated with these question categories could be derived from a number of reasons some of which lay outside the lecturer's zone of control such as the lack of knowledge on the student's part, or the closed nature of the display questions which may not be conducive to students' participation [23] (p. 200) counter to the lecturers' beliefs (more on this below when dealing with our study's limitations), or other factors attributable to the EMI context (e.g., such as the students' insecurity in the productive skills in the L2). However, EMI lecturers in general can modify certain aspects to make these questions more effective. One aspect has to do with the wait time provided by the lecturers. It has been shown that teachers do not usually provide enough wait time for the students to respond (Smith et al., 2003, cited in [28]). This aspect is very important, especially in an EMI context, where students may require longer wait time due to the language barrier.

The third conclusion derived from the results of our study is that lecturers would benefit from receiving some linguistic and methodological (e.g., the kinds of questions that contribute more to learning) training, in particular in association with the formulation of retrospective and repetition questions. It is quite interesting to note that none of the retrospective questions and many of the repetition questions were formulated correctly by the lecturers. In fact, all the instances of retrospective questions were formulated with the verb *remember* embedded in an utterance with rising intonation, mirroring the Spanish syntactic pattern and a few interjections were used as repetition questions. Since the linguistic patterns used by the lecturers may serve as models that students incorporate into their own repertoire, it is essential that lecturers are reminded of the importance of formulating their questions correctly. In this regard, we believe that lecturers would benefit from the observations of their own classes and from training sessions dealing with the specific language issues detected during their viewing.

With regards to the second research question, our study revealed that each lecturer had specific traits in their production of questions. Firstly, we noted that, if comprehension checks are not taken into account, there were remarkable differences in the number of questions posed by the lecturers. T3 asked the most questions by far, 15.54 per 1000 words, then T4 and T2 with considerably fewer questions (7.16‰ and 7.04‰, respectively); and T1 only 6.16%. The implications of this difference should be researched, but our results seem to indicate that customised training sessions could be very productive as teachers' questioning practices vary considerably.

Secondly, all the lecturers used the same question categories: comprehension checks, display, self-answered and referential primarily, although to a different extent. Once again, if confirmation checks are disregarded, T1 and T2 used display and self-answered questions the most, whereas T3 and T4 used display and referential questions the most, although in different proportions. These differences have an impact on the teacher-student interaction as T3 and T4 obtained the highest student response rates, although it was still low. When compared with the results obtained in a different EMI discipline (e.g.,

Business administration), confirmation checks, display and referential were also the most frequently used questions by the lecturers [23]. Our results concur with Chang's [21], who also observed that the use of questions was similar irrespective of the disciplinary culture. Hence, it would be interesting to find the common ground among all the teachers and learn from the most successful strategies used by the lecturers. At the same time, there are also differences among the lecturers, which beg the question if these differences are the result of personal teaching style differences, specific teaching goals and/or are related to the EMI context. This is an issue that needs to be researched.

Finally, we believe that the extensive use of comprehension checks should be further examined. All the lecturers used comprehension check questions very frequently. This was particularly the case of T2. Whether this tendency is specific to the EMI context or a truly personal trait also needs to be addressed. It may be the case that underlying this tendency is a sense of insecurity related to the teacher's self-perceived level of competence in English [32] and therefore this is an issue that would need to be resolved.

## 7. Conclusions

Our study suffers from some limitations. First, we did not focus on students' answers because our study was aimed at analysing only the questioning practices of EMI teachers and, importantly, students' answers were not always understandable in the recordings (i.e., due to noise in the background or students speaking in a low voice), which prevented us from carrying out a detailed analysis of them. Further research should aim at examining the impact of EMI teachers' questions on students' answers. Second, we did not study teachers' beliefs, and this is an interesting avenue for future research, as their beliefs may influence their instructional practices and interactions with students. In addition, it would be worth considering whether different disciplines have any kind of impact on teachers' questioning practices, in other words, whether the questions posed by EMI lecturers from different disciplines differ from each other (for more on this, see [10]).

The systematic observation and analysis of teacher questioning behaviours is crucial as it can help us to reach a better understanding of gains in student achievement [13]. This is why it is important to analyse teacher–student interaction. An important issue that needs to be addressed is EMI teachers' need to improve their use of questions, as our findings indicate that the negotiation of meaning is not as rich as it should be. Hence, it may be useful to provide content teachers with some training and design teacher training courses that "include activities to develop interactional and multimodal competences" [33] (p. 320) in order to enhance student participation and promote learning. Similarly, teachers need to become aware of the fact that high-level questions facilitate deep comprehension [34], which is why display and referential questions should be used much more frequently in EMI classes. High-level questions are not yes-or-no questions, they never have an obvious answer, nor have they only a single answer. They cannot be answered only by simple recollection or by quoting directly from a written or oral text. High-level questions foster critical thinking because they expect students to analyse, synthesise and evaluate information, instead of just recalling a particular fact. For example, if the teacher asks "When was the French revolution?" (a display question), the answer would simply be a date, but if the same teacher asks "What were the causes behind the French revolution?" (also a display question, but much more cognitively demanding), students are forced to compare and contrast information, make judgements, explain reasons for their judgements, and develop reasoning. The latter question would be labelled as a high-level question, whereas this is not the case of the first question. Both types of questions fulfil an important role in class, but EMI teachers need to realise that high-level questions foster interaction and thinking skills to a much higher degree.

**Author Contributions:** Conceptualization, A.D. and D.L.; methodology, A.D. and D.L.; software, A.D. and D.L.; validation, A.D. and D.L.; formal analysis, A.D. and D.L.; investigation, A.D. and D.L.; resources, A.D. and D.L.; data curation, A.D. and D.L.; writing—original draft preparation, A.D. and D.L.; writing—review and editing, A.D. and D.L.; visualization, A.D. and D.L.; supervision, A.D. and

D.L.; project administration, D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is part of the following research projects: PID2020-117882GB-I00 (Spanish Ministry of Science and Innovation) and IT1426-22 (Basque Government).

**Institutional Review Board Statement:** Not applicable.

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

**Data Availability Statement:** Data is unavailable due to privacy restrictions.

**Acknowledgments:** We would like to express a special debt of gratitude to Kebir Colmenero for his help with validation.

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

## Nomenclature

| Transcription conventions | |
|---|---|
| # | pause/interruption |
| [ . . . ] | some part of the transcription is skipped for the sake of space |
| [dis] | display question |
| [ref] | referential question |
| [lang] | language question type |
| [con] | confirmation question type |
| [ret] | retrospective question type |
| [self] | self-answered question type |
| [rhet ] | rhetorical question type |
| [ind] | indirect question type |
| [proc] | procedural question type |
| [off] | off-task question type |
| Xxx | unintelligible speech |
| ( ) | translation of Spanish terms/words into English |
| **In bold** | our emphasis |
| S | Student (since the students were not the focus of this study, all the students were referred to as S) |

## References

1. Kawalkar, A.; Vijapurkar, J. Scaffolding science talk: The role of teachers' questions in the inquiry classroom. *Int. J. Sci. Educ.* **2013**, *35*, 2004–2027. [CrossRef]
2. Al-Adeimi, S.; O'Connor, C. Exploring the relationship between dialogic teacher talk and students' persuasive writing. *Learn. Instr.* **2021**, *71*, 101388. [CrossRef]
3. Chin, C. Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *J. Res. Sci. Teach.* **2007**, *44*, 815–843. [CrossRef]
4. Rose, H.; Mckinley, J. Japan's English-medium instruction initiatives and the globalization of higher education. *High. Educ.* **2018**, *75*, 111–129. [CrossRef]
5. Lasagabaster, D. *English-Medium Instruction in Higher Education*; Cambridge University Press: Cambridge, UK, 2022.
6. Howe, C.; Hennessy, S.; Mercer, N.; Vrikki, M.; Wheatley, L. Teacher–student dialogue during classroom teaching: Does it really impact on student outcomes? *J. Learn. Sci.* **2019**, *28*, 462–512. [CrossRef]
7. Lo, Y.Y.; Macaro, E. The medium of instruction and classroom interaction: Evidence from Hong Kong secondary schools. *Int. J. Biling. Educ. Biling.* **2012**, *15*, 29–52. [CrossRef]
8. Pun, J.; Macaro, E. The effect of first and second language use on question types in English medium instruction science classrooms in Hong Kong. *Int. J. Biling. Educ. Biling.* **2019**, *22*, 64–77. [CrossRef]
9. Van der Wilt, F.; Bouwer, R.; van der Veen, C. Dialogic classroom talk in early childhood education: The effect on language skills and social competence. *Learn. Instr.* **2022**, *77*, 101522. [CrossRef]
10. Lasagabaster, D.; Doiz, A. Classroom interaction in English-medium instruction: Are there differences between disciplines? *Lang. Cult. Curric.* **2022**, 1–17. [CrossRef]
11. Kyriacou, C.; Issitt, J. *What Characterizes Effective Teacher–Pupil Dialogue to Promote Conceptual Understanding in Mathematics Lessons in England in Key Stages 2 and 3? EPPI-Centre Report no. 1604R*; Social Science Research Unit, Institute of Education, University of London: London, UK, 2008. Available online: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2368 (accessed on 12 May 2022).
12. Rojas-Drummond, S.; Mercer, N. Scaffolding the development of effective collaboration and learning. *Int. J. Educ. Res.* **2004**, *39*, 99–111. [CrossRef]

13. Sánchez-García, M.D. Mapping lecturer questions and their pedagogical goals in Spanish- and English-medium instruction. *J. Immers. Content-Based Lang. Educ.* **2020**, *8*, 28–52. [CrossRef]
14. Airey, J. *Science, Language and Literacy. Case Studies of Learning in Swedish University Physics*; Acta Universitatis Upsaliensis; Uppsala Dissertations from the Faculty of Science and Technology: Uppsala, Sweden, 2009. Available online: http://publications.uu.se/theses/abstract.xsql?dbid=9547 (accessed on 8 May 2022).
15. Airey, J.; Linder, C. Language and the experience of learning university physics in Sweden. *Eur. J. Phys.* **2006**, *27*, 553–560. [CrossRef]
16. Mortimer, E.F.; Scott, P.H. *Meaning Making in Secondary Science Classrooms*; Open University Press: London, UK, 2003.
17. Dafouz, E.; Hüttner, J.; Smit, U. New context, new challenges: Oral disciplinary language development in English medium instruction (EMI) and its implications for teacher education. *TESOL Q.* **2018**, *52*, 540–563. [CrossRef]
18. Suviniitty, J. Lecturers' questions and student perception of lecture comprehension. *Hels. Engl. Stud.* **2010**, *6*, 44–57.
19. Engin, M. Contributions and silence in academic talk: Exploring learner experiences of dialogic interaction. *Learn. Cult. Soc. Interact.* **2017**, *12*, 78–86. [CrossRef]
20. Tsou, W. Interactional Skills in Engineering Education. In *English as a Medium of Instruction in Higher Education: Implementations and Classroom Practices in Taiwan*; Tsou, W., Kao, S.-M., Eds.; Springer: New York, NY, USA, 2017; pp. 79–93.
21. Chang, Y.-Y. The use of questions by professors in lectures given in English: Influences of disciplinary cultures. *Engl. Specif. Purp.* **2012**, *31*, 103–116. [CrossRef]
22. Lammers, W.J.; Murphy, J.J. A profile of teaching techniques used in the university classroom. *Act. Learn. High. Educ.* **2002**, *3*, 54–67. [CrossRef]
23. Sánchez-García, M.D. A Contrastive Analysis of Spanish and English-Medium Instruction in Tertiary Education: Teacher Discourse Strategies in a Spoken Corpus. Ph.D. Thesis, Universidad Complutense de Madrid, Madrid, Spain, 2016.
24. Lasagabaster, D.; Doiz, A. *Language Use in English-Medium Instruction at University: International Perspectives on Teacher Practice*; Routledge: New York, NY, USA, 2021.
25. Guarda, M.; Helm, F. A Survey of Lecturers' Needs and Feedback on EMI Training. In *Sharing Perspectives on English-Medium Instruction*; Ackerley, K., Guarda, M., Helm, F., Eds.; Peter Lang: Bern, Switzerland, 2017; pp. 167–194.
26. Council of Europe. *Common European Framework of Reference for Languages (CEFR): Learning, Teaching, Assessment*; Council of Europe: Strasbourg, France, 2020.
27. Tobin, K. The role of wait time in higher cognitive level learning. *Rev. Educ. Res.* **1987**, *57*, 69–95. [CrossRef]
28. Mujis, D.; Reynolds, D. *Effective Teaching: Evidence and Practice*, 3rd ed; Sage: London, UK, 2011.
29. Evans, S.; Morrison, B. The student experience in English-medium higher education in Hong Kong. *Lang. Educ.* **2011**, *25*, 147–162. [CrossRef]
30. Kim, E.G. English Medium Instruction in Korean Higher Education: Challenges and Future Directions. In *English Medium Instruction in Higher Education in Asia-Pacific*; Multilingual Education; Fenton-Smith, B., Humphreys, P., Walkinshaw, I., Eds.; Springer: New York, NY, USA, 2017; Volume 21, pp. 53–69.
31. Doiz, A.; Lasagabaster, D. An analysis of the use of cognitive discourse functions in English-medium history teaching at university. *Engl. Specif. Purp.* **2021**, *62*, 58–69. [CrossRef]
32. Doiz, A.; Lasagabaster, D. Teachers' and students' L2 motivational self-system in English-medium instruction: A qualitative approach. *TESOL Q.* **2018**, *52*, 657–679. [CrossRef]
33. Morell, T.; Norte, N.; Beltrán-Palanques, V. How do Trained English-Medium Instruction (EMI) Lecturers Combine Multimodal Ensembles to Engage their Students. In *La Docencia en la Enseñanza Superior Nuevas Aportaciones desde la Investigación e Innovación Educativas*; Roig-Via, R., Ed.; Octaedro: Madrid, Spain, 2020; pp. 308–321.
34. Cerdán, R.; Vidal-Abarca, E.; Martínez, T.; Gilabert, R.; Gil, L. Impact of question-answering tasks on search processes and reading comprehension. *Learn. Instr.* **2009**, *19*, 13–27. [CrossRef]

education sciences

MDPI

*Article*

# Video Observation of Kindergarten Teachers' Use of Questions in Picture-Book Reading with Quiet Multilingual Children: A Pilot Study

**Marit M. Bredesen [1,2] and Kari-Anne B. Næss [1,2,*]**

[1] Department of Special Needs Education, University of Oslo, Mail Box 1140 Blindern, 0318 Oslo, Norway; m.m.bredesen@isp.uio.no
[2] Department of Pedagogy, Religion and Social Studies, Western Norway University of Applied Sciences, Mail Box 7030, 5020 Bergen, Norway
* Correspondence: k.a.b.nass@uv.uio.no; Tel.: +47-92240741

**Abstract:** Teacher questions asked during picture-book reading may stimulate the child's practice of new vocabulary. However, there is great variation in children's amount of verbal expression, and little knowledge exists about what level of openness in the questions elicits a response. We use video observation and pilot a set of digital picture-book dialog materials that are under development. The analysis included 234 questions asked during picture-book reading in the Norwegian language between three quiet multilingual children and their kindergarten teachers. The analysis was partly qualitative evaluating the types of questions and subsequent responses and partly quantitative in summarizing the occurrence of the types of questions and responses. The results show that between 75% and 97% of the half-open questions, between 60% and 80% of the closed questions, and between 14% and 60% of the open-ended questions elicited a response from the children. Overall, the results indicated that the frequency of responses varied both within and between question types. The fact that open-ended questions generated a limited number of responses among multilingual children may challenge the use of such questions as the gold standard in adult–child dialogs, regardless of child factors and context.

**Keywords:** dialogs; question types; vocabulary; multilingual children

**Citation:** Bredesen, M.M.; Næss, K.-A.B. Video Observation of Kindergarten Teachers' Use of Questions in Picture-Book Reading with Quiet Multilingual Children: A Pilot Study. *Educ. Sci.* **2023**, *13*, 1066. https://doi.org/10.3390/educsci13101066

Academic Editor: Martin Howard

Received: 25 July 2023
Revised: 13 October 2023
Accepted: 16 October 2023
Published: 23 October 2023



## 1. Introduction

Vocabulary, or knowing the meanings of words, is associated with a child's academic [1–4], social, emotional, and behavioral outcomes [5–8]. One activity that is assumed to stimulate vocabulary development is picture-book reading [9–14]. Even in simple books, there is often a higher frequency of complex words than there are in natural dialogs not supported by any written material [12]. In addition, dialogs about content give children the opportunity to learn new words in a meaningful context [15]. Questions asked during picture-book reading dialogs invite the child to reflect on and practice new vocabulary [16]. However, there is in general a great variation in children's amount of verbal expression, and to what extent they respond to questions in dialogic interactions. One group with an especially high risk of having low vocabulary [17] and silent periods [18] consists of multilingual children.

Due to increased migration over recent decades, the number of multilingual children in kindergartens has increased considerably [19]. Multilingual children are a heterogeneous group according to cultural background, time, and type of language exposure in their second language, as well as their level of language achievement [17,20].

Despite this wide variation, on average, research shows that they score significantly lower than their monolingual peers on measures of vocabulary [17,21] and therefore are in need of effective vocabulary interventions. The aim of the present paper is to pilot a

digital picture-book reading session from a vocabulary intervention under development and examine the feasibility of the questions included in the book. The results are expected to be important to the further development of this specific vocabulary intervention and to other researchers planning for picture-book interventions, and they might also be of interest to kindergarten teachers and parents who are interested in how to stimulate (their) child's language.

This paper takes both a linguistic approach, according to which types of questions receive responses from one's interlocutor, and a sociological approach, according to which the response is influenced by the social interaction between the interlocutors [22]. The main issue of interest is the level of openness in each question asked and the subsequent response. However, when needed, the questions are categorized based on the sequential initial turns. Thus, we do not ignore the fact that questions constitute part of a larger dialog sequence and are neither produced nor answered in isolation. However, it would be impossible to consider every aspect of complex social interactions, including the setting and broader social and cultural context; the interlocutors' backgrounds, experiences, and interests; paralinguistic features and elements of speech such as speed, pitch, intonation, and nonverbal communication with gesture and gaze; and various linguistic features including the words used and how they are understood, in an analysis [22]. Due to the lack of research tools that would be needed to investigate the interplay among all the relevant aspects of such a complex phenomenon, this paper focuses on a part of it: the openness of questions and their responses.

### *1.1. Picture-Book Reading and the Use of Questions*

In general, questions are more frequently responded to than narrative statements such as non-question comments or prompts [23]. This is possibly a general quality of questions [23]. Questions may therefore be an effective way of eliciting verbal responses from children but may vary in both linguistic and cognitive requirements. According to Walsh and Hodge [24], a lack of common terminology across studies challenges comparisons across question types. Questions can, for example, be categorized into dichotomous categories such as closed or open-ended, eliciting or non-eliciting, contextualized or decontextualized, literal or inferential, topic initiating or topic continuing, and high- or low-cognitive-demand questions. The same question can thus have several labels and fit into different categories depending on the research focus of interest.

### *1.2. Closed and Open-Ended Questions*

The classification of questions as closed vs. open-ended is widely used, and clear definitions of these two categories enhance the reliability of coding [25]. Walsh and Hodge [24], however, revealed a lack of consensus concerning the definitions of these question types in their systematic review. Examples of different descriptions of closed and open-ended questions are presented in Table 1.

As shown in Table 1, the categorization of questions as closed or open-ended may depend upon linguistic requirements, the cognitive level of abstraction, and/or the number of potentially correct responses. Different definitions can make categorization challenging. For example, the question *What does the girl feel?* can be interpreted as either a closed or open-ended question depending on the definition, the information given in the conversation, the number of possible responses, and the child's actual response. If the child responds only with *sad*, then the question could be classified as closed according to Wasik et al.'s [26] definition. Moreover, according to the definition of Strasser et al. [27], the question can be classified as open-ended unless the previous text has said something about the girl being sad so that the response no longer involves interpretation. Different definitions may be suitable for different research purposes. Lee et al. [25], De Rivera et al. [28], and Hargreaves [29] did not focus on the child's response but rather on the question itself. This emphasis on the question may possibly be appropriate when the adult reads with children who are

unresponsive—for example, children who give mainly one-word responses regardless of the type of question.

**Table 1.** Overview of different descriptions of closed and open-ended questions.

| Authors | Closed Questions | Open-Ended Questions |
| --- | --- | --- |
| Wasik et al. [26] | Require one-word responses | Require more than one-word responses |
| Strasser et al. [27] | Not included in this study | Allow children to make predictions, interpret images and draw parallels to their own lives and experiences |
| Lee et al. [25] | One acceptable response exists, and the question constrains that response | Several different responses would be acceptable |
| De Rivera et al. [28] | Could be responded to with one or more words, the question constrained the child's response, and the answer is usually known to the adult | Could be responded to with one or more words, and the response is not predetermined by the question |
| Hargreaves [29] | Have one correct response and are often (but not always) factual | Often involve reasoning and judgment |

### *1.3. Half-Open Questions*

Hargreaves [29] introduced a third category called half-open questions, which are questions that can be answered with a *yes* or *no* response. Yes/no questions are often treated as closed, and this, Hargreaves [29] suggests, is a simple and reliable solution. At the same time, some children use this opportunity and follow up with extended responses, for example, as a justification of their answer. Presented with half-open questions, a child is always free to respond with a simple *yes* or *no* and thus treat it as closed or give a more comprehensive answer and thus treat it as open-ended. The above author noted that the introduction of a separate category for this type of question solved a troublesome coding problem and, perhaps just as important, elucidated the differences in the responsiveness of the children observed. Hargreaves [29] found that verbally active children, to a greater extent, treated half-open questions as open-ended, while more passive and nonresponsive children treated them as closed.

### *1.4. Present Pilot Study*

The purpose of the present pilot study is to examine kindergarten teachers' use of questions when reading picture-books with children and how different types of questions work in eliciting responses from multilingual children who use few words and few and short utterances in a one-on-one (teacher-child) classroom interaction when the Norwegian language is being used (hereafter referred to as "quiet children"). In general, multilingual children are at risk of weaker language development in terms of their second language due to less experience [17]. These quiet multilingual children, with limited verbal expression, can therefore be particularly vulnerable to weaker second-language development. Questions from adults during book reading may stimulate children's active verbal participation and offer them the opportunity to practice their second language. If certain types of questions are answered more often than others, then such knowledge can be valuable: first, to develop a general understanding of how to support these children's opportunities to participate in dialog and, second, to reveal knowledge that is useful to the development of vocabulary interventions for multilingual children.

There is no clear answer to which types of questions are best suited to stimulate dialog with multilingual children in kindergarten. In the present pilot study, we investigate the following research question:

To what types of questions do quiet children respond to a greater or lesser extent?

Here, a response refers to all forms of responses by the child, including nonverbal responses in the form of affirmative sounds or the nodding or shaking of the head.

## 2. Materials and Methods

The present pilot study is part of a larger Norwegian project approved by the Norwegian Centre for Research Data (reference number 983738). This larger project aims to develop and test the effects of a digital vocabulary intervention. The intervention consists of reading with digital picture-books and systematic activities and is per se in its piloting phase. As children do not start formal reading education in Norway before the children are six years of age, the picture-books used were wordless, but a written script on the bottom of the screen was available for the kindergarten teacher on days 1–4. The script included gradually more abstract questions across the days. The reason for choosing a digital book is the possibility of including animations that can explain the focus word in a more realistic way than is possible with a picture.

The present pilot study includes video observations of kindergarten teachers and children reading one of the books on day 4 (hereafter referred to as picture-book dialogs). Researchers have highlighted the need to pilot materials, strategies, and full interventions before eventually conducting large-scale randomized controlled trials [30,31]. The purpose of a pilot study "...should be to identify the necessity to modify questions or other procedures that do not elicit appropriate responses or enable the researchers to obtain rich data" [32] (p. 3).

### *2.1. Digital Picture-Book*

The digital picture-book and the story behind the drawings were specially developed for this project. The content of the book builds on the national curriculum for kindergartens in Norway and had been evaluated by teachers and revised accordingly before being used in the present pilot study. The book consists of 11 picture pages made available on an iPad. An example of a page from the picture-book is presented in Figure 1. The story is built around the keyword *grow,* and a script with suggested questions is available at the bottom of each page of the book (there were no story scripts included in the picture-book on day 4). In total, 57 possible scripted questions were available to the kindergarten teachers.

**Figure 1.** Example of a page in the picture-book (illustration Nina Skauge).

The suggested script questions regarding the picture in Figure 1 are as follows. *Can you tell me about the picture? (half-open question) What is happening here? (open-ended question) What kind of insects can you see? (closed question question) Do you know other things that grow? (half-open question) What do plants and animals need to grow? (open-ended question)*

### *2.2. Participants*

The Norwegian Centre for Research Data accepted this study (registration number 983738). After obtaining ethical approval, we recruited participants from specific areas around Oslo and in the western part of Norway. Participation was voluntary, and kindergarten teachers as well as the parents or legal guardians of all the participating children

provided informed written consent. The children gave oral consent in the way that they wanted to come with the kindergarten teacher to participate in the reading session. To be included in this study, the children had to be 4–5 years old, second-language learners attending kindergarten, and have parents with a non-Scandinavian first language. We did not set any criteria regarding language proficiency in any of the child's languages or how long the child had been in Norway or attended Norwegian kindergartens. Children with diagnoses that were expected to affect language skills, such as autism and disorders of intellectual functioning, were excluded. The kindergarten teachers had to be qualified teachers. No restrictions were set in terms of teachers' experience or years since graduation.

Four kindergartens, with a total of four kindergarten teachers and 11 children, accepted the invitation to participate in the pilot study and provided signed consent.

The participants in the present study were selected from the available pilot data material with the purpose of including quiet children. The sample in this study consisted of 2 kindergarten teachers and 3 children. Kindergarten teacher 1 had worked in kindergarten for 33 years (all years in the same kindergarten) and kindergarten teacher 2 had worked in kindergarten for 24 years (all years in the same kindergarten).

All participating children had Norwegian as a second language. The categorization of verbal expression (quiet child) was conducted independently by the first author, who has a master's degree in speech and language therapy. It was performed based on video observations of the child's participation in the adult–child dialogs in the present pilot project, and it was categorized based on a relative frequency above 30% of no response to questions asked and the expression of a limited number of unique words (has an average mean length of utterance of one unique word or less). See Table 2 for other key information about the participating children.

**Table 2.** Background information about participating children.

| | **Child 1** | **Child 2** | **Child 3** |
|---|---|---|---|
| Gender | Female | Female | Female |
| Age | 4 years, 10 months | 4 years, 6 months | 4 years, 6 months |
| Country of birth | X | X | Undisclosed |
| Language | Kurdish, X | Kurdish, X | Serbian, English, X |
| Kindergarten [a] | 3.5 years | 3.5 years | 2 years |
| Known difficulties | None | None | None |
| Words per response [b] | 1.60 | 1.02 | 0.44 |

[a] Number of years in kindergarten; [b] Based on video observations of the child's participation in a picture-book dialog.

Kindergarten teacher 1 carried out picture-book dialogs with Child 1 and Child 2, while Kindergarten teacher 2 carried out picture-book dialogs with Child 3. As the children and the kindergarten teachers had been in the same kindergarten for years, they had been familiar with each other for a long time.

### *2.3. Data Collection*

The picture-book dialog was carried out with each child separately and took place in a separate room at the kindergarten to prevent any disturbances. At the time of the data collection relevant for the present pilot study, the picture-book had already been read on three previous occasions (days). Beforehand, the kindergarten teachers had received training in how to carry out the book-reading sessions over the week. They were told that one goal of the intervention was to stimulate active verbal participation from the child and that questions (like the ones suggested in the script) may contribute to this, but at the same time, it was important to adapt the book reading session to the child in question depending on the child's initiative. The questions suggested in the script differed in regard to the level of difficulty, and the kindergarten teachers were instructed to help the children with easier (more concrete) questions if the difficult (more abstract or open) questions were hard for

them to respond to. They were also told that the goal of day 4 was for the child to be the teller of the story.

Because of the COVID-19 situation, the kindergarten teachers were given a video camera to record the picture-book dialogs. The video recordings were stored according to the General Data Protection Regulation (GDPR) on the Services for Sensitive Data (TSD) server.

### *2.4. Transcription*

The recordings of the picture-book dialogs were transcribed in line with conventions, as shown in Table 3.

**Table 3.** Transcription conventions.

| Marking | Explanation |
|---|---|
| [ | Indicates overlapping talk |
| = | Indicates that the utterances follow each other without pause |
| (2) | Indicates timed pauses in seconds |
| (.) | Indicates a micro pause |
| (xxx) | Indicates indecipherable talk |
| (ord) | Indicates that the transcriber is unsure about the correctness of the transcription |
| (( )) | Indicates nonverbal expressions |
| (. . .) | Indicates that words (e.g., names) are excluded from the transcription |
| . . . | Indicates that the kindergarten teacher pauses for the child to complete a sentence or a word |

Note. The conventions are modified from Engevik et al. [33].

For a more detailed description of the transcription procedure, see Supplementary Materials S1.

### *2.5. Coding of Teachers' Questions*

In this pilot study, we did not consider that the participants had read the book before. All questions were coded as if they were asked for the first time. All the questions that received responses were included in the coding process. The analysis also included unanswered questions if they were followed by a pause of three seconds or more. The cutoff of three seconds was based on Wasik and Hindman [4]. They emphasized that the child must be given enough time to respond and recommended a pause of at least three seconds before continuing the dialog. The cutoff of three seconds has also been used in a range of other studies [33]. During the transcription process, a stopwatch was used to decide whether three seconds had elapsed between the end of a question and the adult's next utterance.

The kindergarten teachers' questions were classified in terms of *type* as (1) half-open, (2) closed, or (3) open-ended. Each question was also classified by *subtype*.

Half-open questions can be answered with a 'yes', 'no', affirmative sound, nod, or shaking the head (hereafter yes-/no-type response), but they can also be answered with words other than this (there were no yes-/no-type of responses in the two other main categories). The subtypes of the half-open questions were (1) half-open closed (HO:C) and (2) half-open open (HO:O), based on the child's response.

Closed questions are defined as those wherein a particular response is expected. The subtypes of the closed questions were (1) completion, (2) localization, (3) labeling, (4) attribute, (5) questions with response options, (6) recall, or (7) closed decontextualized.

Open-ended questions are defined as those for which several different responses are acceptable. These questions usually require higher cognitive skills, such as reasoning and judgment. The subtypes of the open-ended questions were (1) summary, (2) descriptions, (3) assessment of emotions, (4) inferential, (5) open-ended decontextualized, or (6) predictions.

The main categories of all three question types were coded according to the questions themselves based on their possible responses or the types of responses they encouraged. For a more detailed description of the main categories and the different subtypes, see Supplementary Materials S1.

### *2.6. Coding of Children's Responses*

The coding of the responses focused mainly on the relevance to the content rather than the linguistic sophistication. Responses to the questions were classified as:

(1) do not know/no response
(2) yes-/no-type
(3) with words
(4) with inadequate words

Criteria for the different response types were:

1. When a child did not respond or answered, 'I don't know', the response was classified as 'do not know/no response.'
2. 'Yes-/no-type' denotes those responses with yes, no, affirmative sound, nodding, or shaking of the head.
3. 'With words' denotes adequate responses with words other than 'yes-/no-type' or 'do not know.' A response was considered adequate even if it was not fully correct based on the question asked.
4. 'With inadequate words' denotes responses with words other than 'yes/no' or 'do not know', where the response was directly wrong or completely irrelevant to the question asked.

### *2.7. Data Analysis*

First, after the dialog was transcribed, questions that were not related to the content of the book (for example, *Are you tired today?*) or undecipherable talk were excluded. Second, in further analysis, the questions and corresponding responses were classified. Third, frequency analysis was conducted in Excel (version 2019). Due to the sample size, no significance testing was performed in this pilot study. Therefore, any references to variation in incidence do not refer to statistically significant differences.

### *2.8. Double Coding*

A total of 234 questions were coded. A trained research assistant double-coded 100% of these questions in order to ensure that the code categories we used were well described and that multiple ratings for the same question were constant. For descriptions of the double coder, see Supplementary Material S2. Thus, there were 234 codings of type and 234 codings of subtype. With regard to type, there were 7 discrepancies between the coders, and regarding subtype, there were 10 discrepancies, giving an interrater agreement (Cohen's Kappa) of 0.967 and 0.943, respectively. Coding discrepancies were resolved via discussions between the two authors until an agreement was reached.

## 3. Results

The dialogs lasted from approximately 9 min to 13 min. Child 1 was asked 92 questions (7.03 questions per minute), Child 2 was asked 71 questions (6.15 questions per minute), and Child 3 was asked 71 questions (7.61 questions per minute), which meant that Child 2, on average, received one new question (that was followed by a pause of a minimum of 3 s) approximately every 10th second. For Children 1 and 3, this number was slightly higher.

Table 4 shows that the most frequently asked question type was half-open, followed by closed, and finally, open-ended. This pattern applied to all three dialogs. Table 5 shows the relative occurrence of each response type for the half-open-ended questions.

There was a pattern across all the recorded picture-book dialogs included in the present pilot study in that the half-open questions elicited a high proportion of yes-/no-type responses and, to a lesser extent, responses of other types. This was particularly prominent for Children 2 and 3. Child 1 had the lowest proportion of yes-/no-type responses and answered more often with other words compared with Children 2 and 3. Approximately half of Child 1's responses (that were not yes-/no-type) were inadequate.

**Table 4.** Frequency of questions by type reported as a percentage.

| Type | Child 1 % | Child 2 % | Child 3 % |
|---|---|---|---|
| Half-open | 40 | 45 | 45 |
| Closed | 33 | 31 | 35 |
| Open | 27 | 24 | 20 |
| Total | 100 | 100 | 100 |

**Table 5.** Half-open questions by response type reported as a percentage.

| Response Type | Child 1 % | Child 2 % | Child 3 % |
|---|---|---|---|
| No response | 3 | 9 | 25 |
| Yes-/no-type | 67 | 91 | 72 |
| With words | 14 | 0 | 3 |
| Inadequate | 16 | 0 | 0 |
| Total | 100 | 100 | 100 |

Table 6 shows the relative occurrence of each response type for the closed questions.

**Table 6.** Closed questions by response type reported as a percentage.

| Response Type | Child 1 % | Child 2 % | Child 3 % |
|---|---|---|---|
| Do not know/No response | 20 | 45 | 40 |
| With words | 40 | 50 | 52 |
| Inadequate | 40 | 5 | 8 |
| Total | 100 | 100 | 100 |

In all three dialogs, approximately half of the closed questions were answered adequately with words (other than yes/no/I do not know). Again, Child 1 had a relatively high proportion of inadequate responses and showed the same tendency as that with the half-open questions. At the same time, Child 1 had a lower proportion of do not know/no response answers compared with the two other children. Children 2 and 3 had a relatively low occurrence of inadequate responses and, at the same time, a higher proportion of do not know/no response. Table 7 shows the total number of closed questions by subtype and the occurrence of each response type for different subtypes.

As Table 7 shows, closed questions of the subtypes *completion* and *labeling* were used relatively often in all three picture-book dialogs. The other subtypes were used infrequently and to varying degrees between dialogs.

Table 8 shows the relative occurrence of each response type for the open-ended questions.

A relatively high proportion of the open-ended questions resulted in do not know/no response. This pattern emerged across all three dialogs. In Child 1′s case, 40% of these questions elicited a do not know response or no response at all. Children 2 and 3 had particularly high proportions of this response type, 65% and 86%, respectively. At the same time, Child 1 had a higher proportion of responses with words in total (adequate and inadequate) than do not know/no response, while the opposite was true for Children 2 and 3. Again, Child 1 stood out with a relatively high proportion of inadequate responses (24%), while Children 2 and 3 had no inadequate responses (0%).

Table 9 shows the total number of open-ended questions by subtype and the occurrence of each response type for different subtypes.

**Table 7.** Closed questions by subtype and response type reported as a raw score.

| Subtypes | Child 1 | | | | Child 2 | | | | Child 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | D/N [a] | WW [b] | IA [c] | Total | D/N [a] | WW [b] | IA [c] | Total | D/N [a] | WW [b] | IA [c] |
| Completion | 15 | 1 | 7 | 7 | 5 | 2 | 2 | 1 | 15 | 6 | 8 | 1 |
| Location | 3 | 2 | 1 | - | 3 | - | 3 | - | 3 | 2 | 1 | - |
| Labeling | 12 | 3 | 4 | 5 | 8 | 7 | 1 | - | 4 | 1 | 3 | - |
| Attribute | 0 | - | - | - | 0 | - | - | - | 0 | - | - | - |
| Question with response options | 0 | - | - | - | 5 | 1 | 4 | - | 2 | 1 | - | 1 |
| Recall | 0 | - | - | - | 1 | - | 1 | - | 0 | - | - | - |
| Closed decontextualized | 0 | - | - | - | 0 | - | - | - | 1 | - | 1 | - |

[a] Do not know/no response. [b] With words. [c] Inadequate.

**Table 8.** Open-ended questions by response type reported as a percentage.

| Response Type | Child 1 % | Child 2 % | Child 3 % |
|---|---|---|---|
| Do not know/No response | 40 | 65 | 86 |
| With words | 36 | 35 | 14 |
| Inadequate | 24 | 0 | 0 |
| Total | 100 | 100 | 100 |

**Table 9.** Open-ended questions by subtype and response type reported as a percentage.

| Subtypes | Child 1 | | | | Child 2 | | | | Child 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | D/N [a] | WW [b] | IA [c] | Total | D/N [a] | WW [b] | IA [c] | Total | D/N [a] | WW [b] | IA [c] |
| Summary | 2 | 1 | 1 | - | 0 | - | - | - | 0 | - | - | - |
| Descriptions | 12 | 5 | 6 | 1 | 10 | 8 | 2 | - | 11 | 9 | 2 | - |
| Assessment of emotions | 1 | - | 1 | - | 0 | - | - | - | 1 | 1 | - | - |
| Inferential | 7 | 4 | 1 | 2 | 1 | - | 1 | - | 1 | 1 | - | - |
| Decontextualized | 3 | - | - | 3 | 1 | - | 1 | - | 1 | 1 | - | - |
| Predictions | 0 | - | - | - | 1 | 1 | - | - | 0 | - | - | - |

[a] Do not know/no response. [b] With words. [c] Inadequate.

Of the six subtypes of open-ended questions, *descriptions* occurred most frequently in all three picture-book dialogs. The other subtypes were used infrequently and to varying degrees among dialogs. Descriptions are typically questions where the child is asked to talk about something with the support of the pictures in the book. An example of a question in this category is *What is happening here?*

Table 10 is based on the occurrences displayed in Table 9 and shows the proportion of do not know/no response vs. response for the subtype descriptions. Adequate and inadequate responses to these questions were combined to indicate their ability to elicit a response from the children.

**Table 10.** Subtype descriptions and response types reported as a percentage.

| Subtype | Child 1 | | Child 2 | | Child 3 | |
|---|---|---|---|---|---|---|
| | D/N [a] | WW/IA [b] | D/N [a] | WW/IA [b] | D/N [a] | WW/IA [b] |
| Descriptions | 42 | 58 | 80 | 20 | 82 | 18 |

[a] Do not know/no response. [b] With words/inadequate.

As Table 10 shows, a relatively high proportion of the subtype *descriptions* resulted in do not know/no response in all three dialogs, especially for Children 2 (80%) and 3 (82%).

Table 11 is a compilation of Tables 5, 6 and 8 and displays the occurrence of responses to the different types of questions. All forms of responses are included. The table indicates the ability of different question types to elicit some kind of response from the children.

**Table 11.** Compilation of response occurrences from Tables 5, 6 and 8 reported as a percentage.

| Question Types | Child 1 Responses [a] % | Child 2 Responses [a] % | Child 3 Responses [a] % |
|---|---|---|---|
| Half-open | 97 | 91 | 75 |
| Closed | 80 | 55 | 60 |
| Open | 60 | 35 | 14 |

[a] Includes all forms of responses (yes-/no-type, with words and inadequate for the half-open questions and with words and inadequate for the closed and open-ended questions).

As Table 11 shows, there is a pattern across the three dialogs in that the half-open questions had a high response frequency. The closed questions had a medium-high response frequency, and the open-ended questions had a relatively low response frequency. These tendencies applied to all children, even though the response rates for different questions were not the same. Child 1, for example, responded to 60% of the open-ended questions, while Child 2 responded to 35% of the open-ended questions, and Child 3 responded to only 14% of this question type. Notably, Child 1 had a mean of 1.00 words per response to the half-open questions, 1.63 words per response to the closed questions, and 2.40 words per response to the open-ended questions. Child 2 had a mean of 0.78 words per response to the half-open questions, 0.82 words per response to the closed questions, and 0.88 words per response to the open-ended questions. Child 3 had a mean of 0.22 words per response to the half-open questions, 0.64 words per response to the closed questions, and 0.57 words per response to the open-ended questions.

## 4. Discussion

The results from the present pilot study showed that the quiet children overall were asked more questions than suggested by the scripts, possibly due to their generally low responsivity. However, the results indicated that the children responded differently to different types of questions. Between 75% and 97% of the half-open questions elicited a response from the children (mainly yes-/no-type responses). Between 60% and 80% of the closed questions elicited a response (with words), and between 14% and 60% of the open-ended questions elicited a response (with words). We discuss these findings below.

### *4.1. Half-Open Questions—Easy to Answer?*

The half-open questions had a high response rate, but the responses were mainly the yes-/no-type. In line with Hargreaves [29], half-open questions were thus found to not be very suitable for getting the children to practice new vocabulary. Hargreaves [29] investigated how two groups of young school children (unresponsive and responsive) responded to different types of questions. The unresponsive group treated the half-open questions as closed, i.e., gave yes/no responses with no further elaborations or justifications, more often than the responsive group did. Yes-/no-type questions are the first type of questions that children understand [34], and they place relatively low linguistic demand on them [18]. Recommendations to minimize the use of yes-/no-type questions can be found in several studies because this question type may place the child in a passive role [35–37]. At the same time, removing the possibility of giving a yes-/no-type response by reducing the number of half-open questions may also present challenges for children, such as perceived pressure to speak in longer sentences [38] and/or higher demands in terms of prior language competence.

For some children, in certain developmental periods or in certain situations, it may therefore be possible that some well-designed half-open questions, rather than more de-

manding questions, could be appropriate. Well-designed half-open questions encourage more than a simple yes-/no-type answer without prohibiting this type of answer. In contrast, answering more demanding questions requires inferences and longer utterances. Not only is it conceivable that such questions exert less pressure because it is easier to respond to them, but situations in which a child does not meet the requirements set forth can also be avoided. The adult can further, through such questions, model language for children who are weak to express themselves at the same time as the question formulation invites children to take an active position in what is said. In addition, these questions represent an opportunity to expose children to target words by incorporating them into the question itself.

### *4.2. Closed and Open-Ended Questions—Suitable for Eliciting Responses with Words?*

While closed questions had a medium-high response rate and open-ended questions had a low response rate, both types require responses with words (other than yes/no). However, closed and open-ended questions can differ in several ways, and these differences may in turn impact the children's responses. One such difference is that closed questions generally facilitate short responses [28], while open-ended questions generally encourage extended and elaborated answers. Hindman et al. [16] noted that open-ended questions often evoke more child speech in the form of longer and more complex responses. In line with this, Deshmukh et al. [39] found that Wh-style questions (many of them open-ended) boosted children's speech and recommended that teachers ask more questions of this type to elicit verbal responses at a higher level. Perhaps here lies a presumption that the verbal activity and responsiveness of the child are a result of the questioning practices of adults. Hargreaves [29] agreed with the notion that there is a correlation between questioning style and responsiveness/response length. At the same time, he questioned whether such a correlation is causal. If the correlation is causal, then the verbal activity of the children should increase when the children are given the opportunity to talk more—for example, by the adult asking open-ended questions. As the present study shows, open-ended questions do not necessarily always have such an effect. On the one hand, this finding is not in line with that of Deshmukh et al. [39]. It is, on the other hand, well in line with Hargreaves' [29] findings that open-ended questions do not automatically have this effect on verbal activity for all children. The less-responsive group of children he observed simply left many of the open-ended questions unanswered, similar to the children in the present study.

It can be argued that giving elaborate answers is more linguistically challenging than responding with one or two words. If so, weak vocabulary may hinder children's responses to open-ended questions. At the same time, as demonstrated by the children in the present study and as De Rivera et al. [28] noted, it is also possible to respond to open-ended questions with only one or a few words.

Open-ended questions may also involve the use of higher cognitive skills, such as reasoning and judgment [29]. Questions that require decontextualized thinking or reasoning are often considered more demanding for a child to answer [24]. The cognitive demand inherent in some types of open-ended questions can, therefore, also be another limiting factor with respect to response rate. In contrast, open-ended questions of the subtype descriptions (the most frequently used subtype of open-ended questions in the present study) are, as Van Kleeck et al. [40] noted, at the concrete level. According to Walsh and Hodge [24], questions at this level are usually not very cognitively demanding on the child.

Although it is possible to respond to open-ended questions of the subtype descriptions with only one or a few words, and even though they should not be very cognitively challenging for the children, they were often left unanswered in this study. One feature of these questions is that they can be relatively "wide" in the sense that they provide little guidance for how to answer and offer little support, focus, and direction to the child. It is therefore up to the child to take the initiative and decide what the answer should be about, in contrast to the situation under closed questions. Thus, it is possible that the absence of restrictions, which are intended to lead to extended responses, reduces the product

ion of language for these quiet children. It is also possible that the directiveness inherent in the closed questions, i.e., that the response to a large degree is predetermined by the question [28], helps steer the child's focus toward something specific and thereby possibly reduces the need for the child to take his or her own initiative. One of the common features of these quiet children was their low degree of initiative, as indicated by their tendency to treat the half-open questions as closed [29,41]. It cannot be ruled out that the lower requirement for their own initiative embedded in the closed questions may contribute to explaining the relatively high response rate among the children for this question type.

### *4.3. Implications*

The high total number of questions, particularly the high number of unanswered questions, may indicate that the kindergarten teachers' strategy for interacting with quiet children is to fill the silence with new questions. In this regard, it is possible that the pretraining of these kindergarten teachers could offer some explanation. Even though the kindergarten teachers were instructed to adapt the book reading session to the child in question depending on the child's initiative, it may seem that asking questions, presumably with the goal of verbal activity among the children, became too dominant. It is thus possible that the pretraining program was insufficient in conveying to these teachers how to adjust the questioning strategies when the child was reluctant to talk and answer questions. Two direct implications for the further development of our interventions are that quiet multilingual children seem to be in need of other educational methods than questions alone and that the teachers are in need of better training before the intervention starts.

### *4.4. Limitations*

One obvious limitation of this study is the small number of participants. It is therefore uncertain to what extent the findings would be valid in other contexts and with other quiet multilingual children. Additionally, only one day, with one specific book, was examined. It is possible that the book itself was not engaging enough and that the children's responsivity could have been different with another picture-book. The video recording may also have affected both the children and adults, and it cannot be ruled out that reading in a more natural setting would have proceeded differently. Other potential limitations include treating the question-response sequence out of the broader context and the decision to treat the questions as if they were asked for the first time when classifying them. Similar questions may have been asked by the kindergarten teachers during the book reading on previous days, which may, in turn, have led to a "closing" of the open-ended questions if the child sensed that a certain response was expected. A possible consequence of this is that the proportion of responses to the open-ended questions becomes too high. It should further be acknowledged that the children's prior semantic knowledge and interests, as well as cultural background, might have played a role in their responses. Also, the relationship between the child and the kindergarten teacher as well as the kindergarten teachers' pedagogical skills may have influenced the results. Thus, it is a limitation that no comparison group was included. It may be that less-quiet children would have answered in the same way. Karlsen et al. [42], e.g., found that a significant part of the conversation contributions from the kindergarten teachers consisted of closed questions, while the children's contributions were mainly short verbal or non-verbal responses.

## 5. Conclusions

The kindergarten teachers asked a high number of questions during the picture-book sessions, the results of which indicated that the quiet multilingual children in this study responded differently to different types of questions. The half-open questions had a high proportion of responses and elicited mainly yes-/no-type responses. The closed questions had a medium-high proportion of adequate responses and elicited only answers with words other than the yes-/no-type. The open-ended questions had a relatively low response rate and elicited answers only with words other than the yes-/no-type if they were indeed

answered. This finding implies that the degree of verbal responses to questions may rely not only on question type but also on factors such as the children's initiative and willingness to speak and/or their verbal expression ability. Thus, the types and number of questions used should be carefully adapted to each child and should be accounted for when planning dialogic interactions involving quiet multilingual children. To fill the silence with new questions seems not to be the perfect pedagogic strategy to use in picture-book reading with quiet multilingual children.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13101066/s1, Supplementary Information S1 Classification of questions, Supplementary Information S2 introduction to double coding.

**Author Contributions:** Conceptualization, M.M.B. and K.-A.B.N.; methodology, M.M.B. and K.-A.B.N.; formal analysis, M.M.B.; data collection, M.M.B. and K.-A.B.N.; resources, K.-A.B.N.; data curation, M.M.B. and K.-A.B.N.; writing, M.M.B. and K.-A.B.N.; project administration, K.-A.B.N.; funding acquisition, K.-A.B.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Research Council of Norway, grant number 299197.

**Institutional Review Board Statement:** This study was approved by the Norwegian Center for Research Data reference number 983738.

**Informed Consent Statement:** Written informed consent has been obtained from the parents of the children to publish this paper.

**Data Availability Statement:** The data will be made available in a written format on Open Science Framework.

**Acknowledgments:** We would like to thank the participating children and parents, kindergartens and kindergarten teachers, the SL+ team and research assistants Susanne Haakonsen and Øyvind Sundstrøm. Finally, we would like to thank the funder of the study: Research Council of Norway.

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

## References

1. Sénéchal, M.; LeFevre, J.A. Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Dev.* **2002**, *73*, 445–460. [CrossRef] [PubMed]
2. Blewitt, P.; Rump, K.; Shealy, S.; Cook, S. Shared book reading: When and how questions affect young children's word learning. *J. Educ. Psychol.* **2009**, *101*, 294–304. [CrossRef]
3. Rogde, K.; Melby-Lervåg, M.; Lervåg, A. Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *J. Res. Educ. Eff.* **2016**, *9*, 150–170. [CrossRef]
4. Wasik, B.A.; Hindman, A.H. Why wait? The importance of wait time in developing young students' language and vocabulary skills. *Read. Teach.* **2018**, *72*, 369–378. [CrossRef]
5. Hart, B.; Risley, T.R. *Meaningful Differences in the Everyday Experience of Young American Children*; Paul H Brookes Publishing: Towson, MD, USA, 1995.
6. Mendez, J.L.; Livy, M.F. Parental reports of preschool children's social behavior: Relations among peer play, language competence, and problem behavior. *J. Psychoeduc. Assess.* **2002**, *20*, 370–385. [CrossRef]
7. Gremillion, M.L.; Martel, M.M. Merely misunderstood? Receptive, expressive, and pragmatic language in young children with disruptive behavior disorders. *J. Clin. Child Adolesc. Psychol.* **2014**, *43*, 765–776. [CrossRef]
8. Stangeland, E. The impact of language skills and social competence on play behaviour in toddlers. *Eur. Early Child. Educ. Res. J.* **2017**, *25*, 106–121. [CrossRef]
9. Robbins, C.; Ehri, L.C. Reading storybooks to kindergartners helps them learn new vocabulary words. *J. Educ. Psychol.* **1994**, *86*, 54–64. [CrossRef]
10. Bus, A.G.; Van Ijzendoorn, M.H.; Pellegrini, A.D. Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Rev. Educ. Res.* **1995**, *65*, 1–21. [CrossRef]
11. Sénéchal, M.; LeFevre, J.A.; Hudson, E.; Lawson, E.P. Knowledge of storybooks as a predictor of young children's vocabulary. *J. Educ. Psychol.* **1996**, *88*, 520–536. [CrossRef]

12. Mol, S.E.; Bus, A.G.; De Jong, M.T.; Smeets, D.J.H. Added value of dialogic parent–child book readings: A meta-analysis. *Early Educ. Dev.* **2008**, *19*, 7–26. [CrossRef]
13. Wasik, B.A.; Hindman, A.H.; Snell, E.K. Book reading and vocabulary development: A systematic review. *Early Child. Res. Q.* **2016**, *37*, 39–57. [CrossRef]
14. Hjetland, H.N.; Hofslundsengen, H.; Klem, M.; Karlsen, J.; Hagen, Å.M.; Engevik, L.I.; Geva, E.; Norbury, C.; Monsrud, M.-B.; Næss, K.-A.B. Vocabulary interventions for second language (L2) learners up to six years of age. *Cochrane Database Syst. Rev.* **2023**, CD014890. [CrossRef]
15. Wasik, B.A.; Bond, M.A. Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *J. Educ. Psychol.* **2001**, *93*, 243–250. [CrossRef]
16. Hindman, A.; Wasik, B.; Bradley, D. How classroom conversations unfold: Exploring teacher–child exchanges during shared book reading. *Early Educ. Dev.* **2019**, *30*, 1–18. [CrossRef]
17. Karlsen, J.; Lyster, S.H.; Lervåg, A. Vocabulary development in Norwegian L1 and L2 learners in the kindergarten-school transition. *J. Child Lang.* **2017**, *44*, 402–426. [CrossRef] [PubMed]
18. Yang, H. On teaching strategies in second language acquisition. *US-China Educ. Rev.* **2008**, *5*, 61–67.
19. Statistics Norway. *Minority Children in Kindergarten 1–5 Years 2015–2022. [In Norwegian: Minoritetsspråklige Barn i Barn Ehager 1–5 år (K) 2015–2022]*; Statistics Norway: Oslo, Norway, 2023.
20. Garcia, O.; Wei, L. *Translanguaging: Language, Bilingualism, and Education*; Palgrave MacMillan: New York, NY, USA, 2014. [CrossRef]
21. Lervåg, A.; Aukrust, V.G. Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *J. Child Psychol. Psychiatry* **2010**, *51*, 612–620. [CrossRef] [PubMed]
22. Stivers, T.; Rossano, F. Mobilizing response. *Res. Lang. Soc. Interact.* **2010**, *43*, 3–31. [CrossRef]
23. Hayano, K. Question design in conversation. In *Conversational Repair and Human Understanding*; Hayashi, M., Raymond, G., Sidnell, J., Eds.; Cambridge University Press: Cambridge, UK, 2013; pp. 395–414.
24. Walsh, R.L.; Hodge, K.A. Are we asking the right questions? An analysis of research on the effect of teachers' questioning on children's language during shared book reading with young children. *J. Early Child. Lit.* **2018**, *18*, 264–294. [CrossRef]
25. Lee, Y.; Kinzie, M.B.; Whittaker, J.V. Impact of online support for teachers' open-ended questioning in pre-k science activities. *Teach. Teach. Educ.* **2012**, *28*, 568–577. [CrossRef]
26. Wasik, B.A.; Bond, M.A.; Hindman, A. The effects of a language and literacy intervention on Head Start children and teachers. *J. Educ. Psychol.* **2006**, *98*, 63–74. [CrossRef]
27. Strasser, K.; Larraín, A.; Lissi, M.R. Effects of storybook reading style on comprehension: The role of word elaboration and coherence questions. *Early Educ. Dev.* **2013**, *24*, 616–639. [CrossRef]
28. De Rivera, C.; Girolametto, L.; Greenberg, J.; Weitzman, E. Children's responses to educators' questions in day care play groups. *Am. J. Speech Lang. Pathol.* **2005**, *14*, 14–26. [CrossRef]
29. Hargreaves, D.H. Teachers' questions: Open, closed and half-open. *Educ. Res.* **1984**, *26*, 46–51. [CrossRef]
30. Campbell, M.; Fitzpatrick, R.; Haines, A.; Kinmonth, A.L.; Sandercock, P.; Spiegelhalter, D.; Tyrer, P. Framework for design and evaluation of complex interventions to improve health. *BMJ* **2000**, *321*, 694–696. [CrossRef] [PubMed]
31. Möller, H.J. Effectiveness studies: Advantages and disadvantages. *Dialogues Clin. Neurosci.* **2011**, *13*, 199–207. [CrossRef] [PubMed]
32. Malmqvist, J.; Hellberg, K.; Möllås, G.; Rose, R.; Shevlin, M. Conducting the pilot study: A neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. *Int. J. Qual. Methods* **2019**, *18*, 1609406919878341. [CrossRef]
33. Engevik, L.I.; Næss, K.A.; Hagtvet, B.E. Cognitive stimulation of pupils with Down syndrome: A study of inferential talk during book-sharing. *Res. Dev. Disabil.* **2016**, *55*, 287–300. [CrossRef] [PubMed]
34. Kristoffersen, K.E.; Simonsen, H.G. *Tidlig Språkutvikling Hos Norske Barn: Macarthur-Bates Foreldrerapport for Kommunikativ Utvikling*; Novus Forlag: Sofiemyr, Norway, 2012.
35. Hargrave, A.C.; Sénéchal, M. A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Child. Res. Q.* **2000**, *15*, 75–90. [CrossRef]
36. Huebner, C.E.; Meltzoff, A.N. Intervention to change parent–child reading style: A comparison of instructional methods. *J. Appl. Dev. Psychol.* **2005**, *26*, 296–313. [CrossRef]
37. Huebner, C.E.; Payne, K. Home support for emergent literacy: Follow-up of a community-based implementation of dialogic reading. *J. Appl. Dev. Psychol.* **2010**, *31*, 195–201. [CrossRef]
38. Haynes, J. *Getting Started with English Language Learners: How Educators Can Meet the Challenge*; ASCD: Alexandria, VA, USA, 2007.
39. Deshmukh, R.S.; Zucker, T.A.; Tambyraja, S.R.; Pentimonti, J.M.; Bowles, R.P.; Justice, L.M. Teachers' use of questions during shared book reading: Relations to child responses. *Early Child. Res. Q.* **2019**, *49*, 59–68. [CrossRef]
40. Van Kleeck, A.; Gillam, R.B.; Hamilton, L.; McGrath, C. The relationship between middle-class parents' book-sharing discussion and their preschoolers' abstract language development. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1261–1271. [CrossRef]

41. Wood, H.; Wood, D. Questioning the pre-school child. *Educ. Rev.* **1983**, *35*, 149–162. [CrossRef]
42. Karlsen, J.; Røe-Indregård, H.; Wold, A.H.; Lykkenborg, M.; Hagtvet, B.E. Initiative and Response Patterns in Book Reading Dialogues between Kindergarten Teachers and Children with Minority Language Backgrounds. NOA—Norwegian as a Second Language. [In Norwegian: Initiativ og Responsmønstre i Boklesingsdialoger Mellom Barnehagelærer og Barn Med Minoritetsspråklig Bakgrunn. NOA—Norsk som andrespråk]. 2018, pp. 1–2. Available online: http://hdl.handle.net/10852/71298 (accessed on 24 July 2023).

education sciences

MDPI

*Article*

# Interactive Alignment in L2 Learning: The Link between Social Interaction and Psycholinguistic Phenomena

**Nadia Mifka-Profozic**

Department of Education, University of York, York YO10 5DD, UK; nadia.mifka-profozic@york.ac.uk; Tel.: +44-7572910560

**Abstract:** To engage successfully in conversational activities, participants need to coordinate and synchronise their talk with the talk of their interlocutors. Apart from a set of social strategies and natural routines involved in sequence organisation, a significant contributor to this goal is a psycholinguistic mechanism identified as interactive alignment. The present study set out to examine whether interactive alignment occurs in L2 speech of upper intermediate second language users who have been learning English at school for around 11 years. The participants were a group of twenty Croatian students in their second year of university study, majoring in English. They worked on two collaborative tasks: one carried out in dialogues and the other one in groups of four. Their interactions were analysed both quantitatively and qualitatively, to closely examine how interaction evolves in unscripted task-based L2 production. With a significantly larger number of alignment occurrences recorded in dialogues than in groups of four, both between speaker and within speaker, this study demonstrates that processes in L2 learning and use are interconnected and interdependent at all levels, involving cognitive, psychological, psycholinguistic, and social dimensions.

**Keywords:** alignment; interaction; priming; turn-taking; dialogue; conversation; unscripted tasks; foreign language; second language

**Citation:** Mifka-Profozic, N. Interactive Alignment in L2 Learning: The Link between Social Interaction and Psycholinguistic Phenomena. *Educ. Sci.* **2023**, *13*, 792. https://doi.org/10.3390/educsci13080792

Academic Editor: Martin Howard

Received: 1 May 2023
Revised: 12 July 2023
Accepted: 14 July 2023
Published: 3 August 2023



## 1. Introduction

Levinson [1] contends that all language usage is predominantly interactive, which has important implications for language processing and language acquisition. Interaction in any spoken language is realised through conversation or dialogue as a basic form of interaction involving conversational exchanges between interlocutors. These exchanges constitute a well-established and widely researched system of turn-taking, a fundamental organisation of talk-in-interaction [2,3]. To engage successfully in such conversational activities, participants need to coordinate and synchronise their talk with the talk of their interlocutors, since a natural conversation is guided by the aim of achieving mutual understanding between the interlocutors. By adjusting individual utterances, fine-tuning, anticipating, clarifying and asking for clarification, confirming and repeating others' utterances, speakers make constant attempts to co-construct the meaning in their conversation. In this process towards a shared understanding, the projection or anticipation of what the other will say plays an important role. Achieving the goal of mutual understanding and synchronisation is considered one of the central issues in interaction analysis from social and pragmatic positions [4].

Apart from a set of social strategies and natural routines involved in a conversation sequence organisation [3], a significant contributor to this goal is a psycholinguistic mechanism identified as alignment, which is thought to be driven by an implicit psycholinguistic phenomenon known as structural or syntactic priming. Structural priming is a natural tendency in speakers to use the same syntactic structure repeatedly, either by reproducing one's own or the interlocutor's structures that were recently heard.

The interactive alignment model, according to Pickering and Garrod [5], assumes that in dialogue, production and comprehension cannot be separated: they become intertwined,

which leads to the alignment at the level of linguistic representations, but also includes the context in which dialogue takes place. This process of convergence, in which mental representations are shared between the interlocutors, may take place at the level of syntactic structure, lexis, and phonology simultaneously, which makes linguistic processing in dialogue considerably easier by reducing the processing load on interactants.

The following is an example from Dao et al. [6] where two participants describe a picture, and participant B aligns with the relative clause prime heard in the utterance of participant A.

(1)

A: . . .uh the guy who wants to steal the money (prime)

B: ok I think the first one is **the man who. . .wear. . .wear glasses** (target)

The terminological distinction (i.e., "structural priming" and "interactive alignment") reflects the fine line of differentiation in the approach, namely that structural priming refers to the sharing of linguistic representations between the interlocutors, whereas alignment reflects the same, but it is based on an awareness of the social context including space, time, reasons, and intentions of interaction [5].

The field of second language (L2) acquisition (In this paper, the term *second language* (L2) is used in a broad sense, to cover both *foreign language* and *second language* in a narrow sense (where an L2 is the language of the community)) has also recognised interaction as a driving force in language development [7–11]. This line of research has resulted in a wealth of studies within the framework of the interaction hypothesis [12]. So far, research into L2 acquisition has used syntactic priming in collaborative tasks as a method to benefit L2 grammatical learning (e.g., [10,13–18]). Such tasks have usually been highly structured and scripted. Only a few studies to date have utilised unstructured tasks in which spoken interaction took place in a more spontaneous fashion [6,19,20].

The present research is also interested in interactive alignment among L2 learners in more naturalistic conditions and specifically in the manifestation of alignment as it happens, moment by moment. The aim of the present study was to examine whether interactive alignment occurs in L2 speech of upper intermediate L2 learners who have been exposed to communicative language teaching methods during their schooling. The participants were twenty Croatian students majoring in English, who worked on two collaborative tasks. Since spoken interaction can be realised in a pair and in a multi-party group, the study also looked into the possible differences between the interactive behaviour in pairs and in groups of four.

In what follows, I first refer to interactive alignment in L1 and L2 and the mechanisms involved in its realisation. This is followed by a discussion referring to interaction in L2 and interactional competence, which is the ultimate goal of becoming a fully proficient L2 speaker. Since this goal is linked to the pragmatic and socio-cultural aspects of conversation, the paper also considers the conversation analytic tradition. The study employed conversation analysis (CA) as an analytical tool which enables a close examination of spontaneous talk in collaborative tasks. In doing this, the paper makes an attempt at approaching spoken interaction in L2 from psycholinguistic, interactional, and conversation analytic perspectives.

### *1.1. Interactive Alignment in L1 and L2*

As Garrod and Pickering [21] put it, dialogue encompasses production and comprehension happening almost simultaneously, whereby it is sometimes difficult to disentangle the two processes. At the linguistic level, this is manifested as a tendency to repeat the structures of the previously heard utterances produced by interlocutors. Costa et al. [22] contend that interactive alignment involving L2 speakers differs from alignment in L1 speakers in terms of the level of automaticity, i.e., in L2, it is not an automatic process since it involves the activation of explicit memory mechanisms. In addition, Markman et al. [23] suggested that different mechanisms, even in L1 speakers, may trigger alignment at different linguistic levels. Specifically at the level of discourse, the synchronisation

between the information shared by two partners in interaction will be driven in a different manner than the synchronisation at the level of syntax, semantics, or phonology. While for alignment to occur at these levels, neither awareness nor specific knowledge about the interlocutor's competence is needed, for alignment at the discourse level listeners must be aware of the speaker's knowledge and they must project their utterances containing the appropriate knowledge, at the right moment and in an appropriate way. Therefore, as explained, interactive alignment occurring at the level of syntax, semantics, and phonology is an implicit process involving implicit memories, but alignment occurring at the discourse level must be accessible to consciousness and explicit memories.

Explicit memory of the prime sentence, particularly in the presence of lexical repetition (i.e., lexical boost), can explain short-term priming in both L1 and L2 speakers [24]. In terms of the mechanisms involved in priming, short-term or immediate, and long-term or delayed syntactic priming differ. Long-term priming is explained as an implicit process based on error in expectation or the effect of surprise [25,26] where stronger priming is linked to inverse frequency effects. In other words, this means that priming has a larger magnitude where less frequent syntactic options are primed. On the other hand, short-term priming, particularly when it is enhanced by lexical overlap, involves a component of explicit memory [14,27,28]. In L2 learners, the role of explicit memory is also built into the developmental model of shared syntax [27] where explicit memory (of just heard sentence), rather than the activation of abstract syntactic representations, is involved in the repeated use of syntactic structures. However, changes may occur, depending on the level of proficiency [24,27,29]. At advanced levels, which imply longer experience with the target language that enables the formation of some L2 syntactic representations, the abstract combinatorial nodes of syntactic structures may be activated and strengthened due to the residual activation of recently encountered structures. Residual activation of recently heard sentences has been found in numerous priming studies, showing that what was heard does not decay immediately but stays active for several, even up to 20 s after being attended to [7]. This process in long-term (delayed) priming seems to be modulated by the inverse frequency effect on the one hand, and the closeness of L1 and L2 on the other hand.

### *1.2. Conversation as Social Interaction*

The interactive alignment model, as suggested by Pickering and Garrod [5], does not depend only on shared psycholinguistic representations, but it also involves coordinated and aligned situational representations by interlocutors. Explicit recognition of social context is important because it defines interaction as a psychological and a social construct [30]. It is indicative that almost in parallel with the notion of shared mental representations put forward by psycholinguists in their exploration of dialogue, scholars investigating interaction from the social position introduced the notion of socially shared knowledge, or shared understanding, or intersubjectivity [31]. Coordination in conversation is a fundamental question for conversation analysts as much as it is for psycholinguists. In the conversation analysis paradigm, this coordination is supported by the tight organisation consisting of turn adjacency pairs which are complementary to each other [3] and where the production of the first part creates an expectation for the second part to take place.

In the realm of L2 interaction, which is considered to be a necessary component of L2 development from the earliest to the advanced stages [9,12,16,17] the crucial question concerns evidence of interactional competence. For example, interactional competence can be seen even in speakers of low linguistic ability who may use some formulaic features of conversation such as backchannelling [30]. On the other hand, interactionally competent proficient speakers will participate in conversation by providing responses that are contingent with the interlocutor's interactional behaviour, showing an awareness of the social context [32]. Turn-taking management and topic negotiation which includes topic development and topic shifts, have thus been identified as the conspicuous features of interactional competence at the micro level, with the addition of interactive listening,

breakdown repair, and non-verbal behaviour [33]. The role of non-verbal behaviour is one of the areas that still need to be addressed in discussions on interactional competence, as is the role of task design and its relationship with the elicited interactional behaviour [32].

### *1.3. Interactive Alignment and the Role of Task Design*

Following calls for more research in richer, more natural contexts of L2 use [24,34] where alignment could be investigated in speaker and context dependent discourse, taking a task-based approach might offer a possibility to elicit language in a communicative context. However, tightly scripted tasks that have so far dominated investigation of interactional alignment in L2 may be too restrictive in the sense of narrowing the opportunities for co-construction of meaning in interaction. On the other hand, unscripted and less controlled tasks, which give participants more opportunities to co-construct the meaning in their interaction, may be criticised for making the aligned structures "task necessary" or driven by the task rather than alignment. For example, in one of the rare unscripted studies, such as Dao et al. [6], it was found that the types of aligned structures were in fact related to the type of task. It might be the case, indeed, that it is very difficult to completely separate alignment from collaborative interaction in unscripted tasks and to precisely assign the source of evidence to one or the other.

The present study also used unscripted, unstructured tasks with L2 speakers to elicit interactions that might be similar to naturally occurring conversations. Methodologically, this study made an attempt to connect two fields in linguistics that have historically been separated and separately researched, namely, psycholinguistics and CA, to show the nature of interactional alignment and how it occurs in spontaneous L2 production. Calls for closer collaboration of cognitive and sociocultural traditions in applied linguistics have been put forward by researchers on both sides over the years [35–38], to use the best of both approaches for the sake of research thoroughness, truthfulness, and comprehensiveness in applied linguistics. It is believed that linking the two approaches can assist and move forward the field of both L2 acquisition and L2 education as both approaches used simultaneously help us better understand the processes in L2 learning and consequently, build the strategies for L2 teaching.

## 2. Methods

The present study set out to address the following research questions:

1. Do upper intermediate L2 learners align their interaction in an unscripted task to successfully complete the task as (a) a pair and as (b) a group of four?
2. Are there any differences in participant behaviour while working in pairs and in groups of four?

The aim of the study was to find out whether alignment is present in L2 speakers' task-based interactions that might be similar to spontaneous conversation, and whether the student behaviour demonstrates any differences between interactions in pairs and in groups. The study was designed as an exploratory, descriptive study, bringing in a novel methodological approach, in line with the calls for linking the advantages of qualitatively described processes in L2 interaction and the need to generalise and categorise [36]. Therefore, the present study employed both CA and quantitative analysis to show how interactional features of coordination and synchronisation described in CA translate at the level of language use, as categorised in psycholinguistics.

### *2.1. Participants and Setting*

Participants in the study were 20 Croatian students, age 19–20 (17 females) majoring in English at a university in Croatia. At the time of data collection they were in the second year of their study. They had been learning English for 11 years on average (8–12) as most of them started with English lessons in school at the age of 10, but some started earlier in private language schools and some later if English was their second foreign language. Most of them were exposed to communicative language teaching methods during their entire

education. None of them had spent a period of more than two weeks in an English-speaking country, but they reported regular watching of films, reading, and use of social media in English. Their proficiency at the time when this study was conducted was estimated as B2 in the Common European Framework of Reference for Languages (equivalent to IELTS score of 6.0–6.5 or TOEFL iBT score of 72–94). They had been together as a group for two years, since the beginning of their undergraduate study. Some even went to the same high school or have known each other from primary school. Thus, the level of familiarity among the members in this group was rather high.

### *2.2. Procedure*

The students completed two communicative tasks in which they had to solve problems and make decisions. Prior to taking part in the study, they signed consent forms, including the consent to being audio-recorded. In the first task they worked in pairs, and in the second task, in groups of four. In the first task, they were asked to make a proposal for a renovation project in a community that experienced massive damage to their homes. The students had to take on the roles of councillors, whose task was to discuss new buildings planned for that site. To complete the task, they had to present the main ideas that resulted from their discussion. In the second task they had to discuss and decide which roles to take in a magazine editorial team. The outcome of this task was a presentation of a new magazine, including the description of the four roles in the editorial team. The students were advised to use their own interests when selecting their roles. The purpose of these two tasks was to encourage the use of L2 spontaneous speech as much as possible and to elicit interactions that might be comparable to talk occurring in naturalistic settings. These tasks can be described as holistic tasks in which language is used in the same way as it is used in everyday talk [39]. Holistic tasks involve the learner's knowledge at different levels, from phonology and grammar to discourse, and in that sense they contrast with analytical tasks. Problem solving is considered to be a typical model of a holistic task, and in some educational views [40], it is highly valued for its contribution to the students' ability of meaning-making. These tasks were implemented in the present study exclusively for research purposes, as data collection for this study was not part of the students' curriculum. However, the students had had much experience with working on communicative tasks on different topics. It is important to note, though, that such unstructured tasks are more appropriate for L2 learners who have higher levels of communicative ability, vocabulary, and overall proficiency [40]. Task instructions are available in Supplementary file S1.

Each interaction in pairs lasted between 10 and 15 min (the minimum was 10 min, but students were allowed to use a few more minutes if needed). The interaction in the group of four lasted between 15 and 20 min (the minimum was 15 min). The tasks were carried out in a large lecture room where it was possible to hold four dialogues or two to three conversations in a group of four at a time, so that the voices of other people could not be heard in the other corners of the room. Voice recorders were placed on each desk between the interlocutors and each participant had a lapel microphone tied to their clothes. Only the interaction part of each task was recorded, while the presentations of the outcomes took place later. The performance of the two tasks (ten dialogues and five group conversations) produced recordings in duration of 215 min (3 h 35 min). The recordings were transcribed by two trained students and the transcription was checked by the researcher. Table 1 presents the distribution of words and turns in each task, as well as the time needed to complete these interactions:

**Table 1.** The number of words and turns generated in two tasks, and the length of tasks.

| Interaction Type | Words | Turns | Minutes per Task |
|---|---|---|---|
| Pair interaction | 5603<br>M = 280<br>SD = 128.55 | 944<br>M = 47<br>SD = 14.04 | 132<br>M = 13.2<br>SD = 1.03 |
| Group Interaction | 4541<br>M = 227<br>SD = 123.73 | 758<br>M = 38<br>SD = 19.34 | 83<br>M = 16.6<br>SD = 1.14 |
| Total | 10,144 | 1702 | 215 |

*2.3. Coding*

Once transcribed and initially checked for accuracy of transcription, the recordings were listened to again and relistened to multiple times. In the first phase the focus was on instances of all repetition, both syntactic and lexical. Turns were operationalised as stretches of speech from the point when one started to speak to the point when they stopped speaking. Once all the repetitions were identified, they were coded for alignment following the practice suggested in [6,20]. The constructions [41] were coded if in each of at least five conversations they appeared five times or more. The coding of all structures was bottom-up, i.e., from those constructions identified in the recorded talk. The first mention of a structure that was later repeated, either by the same speaker or by the interlocutor, was identified as a prime. Alignment was coded for target if the repetition occurred within ten turns following the prime [20,42,43]. If the same structure occurred in one of the turns following the tenth turn, it was coded as a new prime. If there was no repetition following the prime within the next ten turns, it was coded as a prime with no alignment. Each occurrence of a structure was coded only once, either as a prime or a target. Targets were also coded for source, depending on whether they occurred within the speaker who produced the prime or between the speakers, i.e., when the interlocutor repeated the structure. If two types of structure could be identified in a construction, it was coded only as one structure (for example, the utterance "...(have) another staircase to go to the next floor" was coded as a non-finite relative clause even though it also contained the construction "have + NP"). Such constructions were always coded for the less frequently present structure in the whole interaction. Another researcher coded two dialogues and one group conversation (around 20% of all data), and the agreement was 88%. For interpretation please refer to Plonsky and Derrick [44] (In Plonsky and Derrick's [41] meta-analysis of reliability coefficients in L2 research, the largest number of studies (369) used percentage agreement for interrater reliability, where the median score was 0.93 (the lowest was 0.81 for pronunciation, the highest was 0.96 for grammar)). All discrepancies were resolved in discussion before reaching the final full agreement.

Table 2 presents the identified structures in the two tasks and an example of each.

Once the coding was completed, a conversation analysis was carried out on those parts of the interactions which were deemed interesting for the persistence of interactive alignment. In doing this, Schegloff's [31] advice was followed, suggesting the introduction of analytic resources only by reference to the details of the interaction which require them for analysis. The aim of conducting a conversation analysis was to examine how individuals make progress in aligning with their interlocutors in the attempts to complete the task successfully. In analysing the transcripts, the transcription conventions were used following Markee [38], based on the Jeffersonian transcription tradition (Supplementary file S3).

**Table 2.** Most frequent constructions identified in the two tasks.

| Structure | Example |
|---|---|
| *Have* + NP | *We have this building and who do we want to benefit from it?* |
| *Have* + VP (infinitive) | *You have to pay more for heating* |
| Complementiser *that* (omitted)+ nominal clause | *I think* (that) *this is good* |
| *Going to* + VP | *There's probably going to be a park, with trees and flowers* |
| Imperative (*let's* + VP) | *Let's move on from here* |
| Non-finite relative clause | *A decent place to live in* |
| Finite relative clause | *This will be for people who are alone* |
| Stranded preposition | *Who would we give the apartments to?* |
| Comparative | *That's a more expensive thing* |
| *What kind of* + NP | *What kind of community do you mean?* |
| NP + *like* + NP | *Something like that* |

## 3. Results

### 3.1. Research Question 1

To answer the first research question, the following interaction demonstrates how the initial sequences of talk between the two L2 speakers are created and how collaboration between them develops gradually. In this analysis the focus is on instances of interactional alignment between participants 1 and 14 (for ease of reading, they are labelled as P1 and P2). Primes are operationalised as the first occurrences of structures that are later repeated either by the same speaker or the interlocutor. Repetitions of primes are identified as instances of alignment or targets. Syntactic repetitions are highlighted in bold:

(2)

001 P1: so::: (.) we need=
002 P2: =yeah
003 P1: um:: (.) we need to think of something (0.2) for **people** um (.) **who will be =**
[=living here]
004 P2: [that are er]
005 P1: ↓yeah (.) a decent place to live in ↓=
= maybe (.) **the people who will be living there** (.) are those people ↑that (.) ↓well (0.3) maybe they don't have that **big pay to ↑support** themselves
006 P2: um yeah (.) **a nice place to live in**
007 P1: ↑so:: <u>these</u> flats will be (.) kind of local council estate um or:: **something like that =**
**=** so, ↑**let's start** with the building
008 P2: ↓okay. (.) if we are aiming at people in need (0.3) um maybe um we should divide the building (.) like (0.3) into ↓flats (.) **for people who will be living there alone**
009 P1: yeah (1.0) um and those with their families
010 P2: yeah! **Let's see** (.) so, those who are alone (.) could be **smaller ones**
011 P1: yeah! and (.) for **smaller families**
012 P2: so, (.) we will provide **more apartments** for ↑**smaller families**
013 P1: yes, I believe there are **more families** in need (.) than (.) **people who will be =**
= living alone
014 P2: um yeah (0.3) if um they have ↑kids =
015 P1: =they usually have ↓kids↓
016 P2: okay
017 P1: ↑ so:::
018 P2: so, those are **going to be** small apartments but (.) functional (.) like small kitchen

019 P1: yeah (.) small

In the above excerpt, the two participants start their interaction with an agreement on what they have to do: P1 (F) seems to be more proactive and initiates the conversation in the pre-expansion, while P2 (M) is slowly catching up and uses backchannelling ("yeah") to agree with his interlocutor. P1 initiates the topic in turn 003 and uses the relative clause "people who will be living here", which she repeats in turn 005 (within speaker alignment). In turn 004, P2 makes an attempt at expanding the conversation, but realises that P1 may not have completed her turn yet, so he quickly stops and applies a repair mechanism, which allows speakers to deal with turn-taking errors such as overlaps. At the same time, just before repeating the relative clause in turn 005, P1 starts developing the topic by saying "a decent place to live in". This non-finite relative clause primes her own structural repetition ("that big pay to support themselves") at the closure of turn 005. This is an expansion of the turn which introduces a new topic, related to the financial ability of potential residents in the planned building. In the following turn, 006, P2 aligns with the prime heard in turn 005 ("a nice place to live in"), which indicates an agreement with the interlocutor. At this point P2 seems to have been fully involved in the conversation. P1 can now complete the topic initiation and expand her turn 007 by suggesting where to start from. In turn 008, P2 explicitly agrees ("okay") with the suggestion and adds his own thoughts while aligning with the relative clause structure heard in turn 005. In turn 009, P1 uses backchannelling, showing overt agreement, but then pauses, and the rest of the turn may be understood as polite disagreement with the idea related to people living alone. P2 has now taken over the management of the topic, and in turn 010 he reuses the interlocutor's construction "let's see" to mitigate the disagreement. He expands the turn by offering the idea of building smaller flats for people who live alone, but in turn 011, along with the backchanneled "pro forma" agreement, P1 again negotiates the idea of "smaller families", where the comparative structure is repeated, suggesting hedging rather than comparison. In turn 012, after a short delay ("so. . ."), P2 accepts the idea of providing "more apartments" for "smaller families". As can be seen, alignment with the interlocutor's idea is accompanied by structural alignment. In turn 013 P1 elaborates on and justifies her idea of building flats for families rather than for people who live alone. The elaboration is inserted between the explicit agreement ("yes") and the aligned relative clause "people who will be living alone." The interaction from turn 014 to turn 019 can be seen as a post-expansion that leads to the sequence-closing sequence, as suggested by Schegloff [3], where "okay", "so", and "yeah", used by two interlocutors, show their explicitly expressed collaborative intentions.

As researchers in pragmatics have proposed [1,2,4,38,45], a conversation in the form of a dialogue is regularly instantiated as an attempt of collaboration to complete the task. To do that, the two interactants in a dialogue must work close to each other, listen to each other, and coordinate their utterances. All this is accomplished in a natural, spontaneous manner.

In terms of interactive alignment, a question can be asked why comparative structures (e.g., "smaller families", "more families", or "more apartments") were identified as alignment, but lexically similar units, such as "small apartment" or "small kitchen", were not coded as primes nor as alignment. The reason is in the explanation provided by Reitter and Moore [43] who argue that immediate lexical repetition can easily lead to syntactic repetition and may inflate results. Therefore, they recommend excluding the cases of immediate verbatim lexical repetition. Comparative units, on the other hand, were counted as alignment because they in fact were not employed to suggest comparison but had other functions, most often hedging, and in many other instances compound comparatives were used.

The following extract is from the second task in which the same students were in groups of four, discussing their ideas about an imaginary magazine and their roles in the editorial team. In this group, participants P1, P10, P12, and P14 took part. For ease of reading, they are labelled as P1, P2, P3, and P4.

(3)

001 P1: o-kay. (.) shall we start?
002 P3: um (.) what type of magazine should we do? =
003 P2: =what type of magazine should we start? =
004 P4: =what do people read today?
005 P1: um (0.3) fashion?
006 P4: no. (.) [tabloids]
007 P2: [tabloids]
008 P4: yeah!
009 P2: we should [make -]
010 P4: [yes, we should make] a tabloid ↓magazine
011 P2: yellow press
012 P3: yep.
013 P2: yes, maybe.
014 P4: um (.) maybe music magazine (.) which would cover everything (.) from pop =
= I don't know (.) rock or -
015 P1: does anyone wanna read ↑that?
016 P3: yes, **I think they ↓do** (.) like Rolling Stone or (.) **something like that** I don't know
017 P2: um ↓maybe.
018 P3: or:: something like political [magazine-
019 P2: oh ↑no:::↑]
020 P3: come ↓on↓ =
021 P4: =or maybe something about nature?
022 P1: um::: who is **going to read** that?
023 P3: oh well, (0.3) =
= how much time do we have?
024 P1: [fifteen:: minutes]
025 P2: [fifteen:: minutes]
026 P4: ↓nice↓ (.) we [wasted two]
027 P2: [we wasted two]
028 P4: ok. (.) we have thirteen more minutes =
= I'm for tabloids.
029 P1: yeah. (.) that's probably best

In excerpt (3), four students start their conversation about the magazine ideas they should develop. P1 (F) initiates the talk, while P3 (M) and P2 (F) follow almost at the same time, with an overlap starting from "should we do?" in turn 002. This can be considered a pre-expansion, where P1 asks a question to invite the others to talk. The two overlapping instances of the question using a modal ("should we. . .") could have been coded as targets primed by the modal "shall we. . ." in turn 001. However, it may be argued that the use of modal verbs in this task is more task-driven than alignment-driven, and therefore, modals were excluded from coding in these tasks. The talk is expanded with very short turns that are inserted in this sequence as ellipses with additional overlapping. It seems that when overlapping occurs, the participants immediately become aware of the violations in their turn-taking and they stop immediately, offering a repair mostly by backchannelling ("yeah" or "yep"). This interaction can also be understood as brainstorming, where each participant offers their ideas in very brief verbal exchanges. Brainstorming is in the function of negotiation about the type of the magazine. In turn 014, P4 elaborates his idea by using a finite relative clause which is coded as a prime but is not aligned within the next ten turns. The question P1 asks in turn 015 indicates disagreement with P4, but P1 also faces disagreement by P3 in turn 016. Turn 016 contains two more primes: one is a subordinate WH clause with the omission of the complementiser "that", and another prime is the

construction "something like that", which is repeated in turn 018 ("something like political magazine") by the same speaker. In turn 022, there is one more prime ("going to read"), but again, no target is being recorded in the following ten turns. In turns 024 and 025, there are two instances of lexical immediate verbatim repetition ("fifteen minutes"), and in turns 026 and 027, there is also a verbatim repetition of "we wasted two" occurring as an overlap. These are not coded for alignment due to the reasons previously mentioned in [26]. As can be seen from this transcript, the interaction in this group of four participants differs from pair interaction as there is more overlapping and more short elliptic turns which do not appear to allow for more significant alignment of syntactic patterns. More evidence of these issues in group conversation can be seen in the transcript in Supplementary file S2.

### 3.2. Research Question 2

The following two tables present the descriptive data from the transcripts of the two tasks: the number of primes for each structure, within-interlocutor and between-interlocutor alignment, the number of non-aligned primes, and the number of total occurrences. Table 3 refers to interactive alignment in the dialogue task, and Table 4 refers to the group task.

**Table 3.** Aligned structures in pair interactions.

| STRUCTURE | Primes | Alignment within speaker | Alignment between speakers | Total alignment | Total occurrences |
|---|---|---|---|---|---|
| *have* + NP | 24 | 13 | 8 | 21 | 45 |
| *have* + VP (inf) | 16 | 11 | 6 | 17 | 33 |
| complementiser *that* omitted | 31 | 12 | 11 | 23 | 54 |
| *going to* + VP | 24 | 8 | 8 | 16 | 40 |
| Imperative (*let's*) | 22 | 9 | 6 | 15 | 37 |
| non-finite relative clause | 18 | 12 | 13 | 25 | 43 |
| finite relative clause | 47 | 17 | 15 | 32 | 79 |
| stranded preposition | 18 | 9 | 5 | 14 | 32 |
| comparative | 22 | 7 | 6 | 13 | 35 |
| *what kind of* + NP | 21 | 9 | 6 | 15 | 36 |
| NP + *like* + NP | 27 | 11 | 9 | 20 | 47 |
| TOTAL | 270 | 118 | 93 | 211 | 481 |
| Mean | 25 | 11 | 8 | 19 | 44 |
| SD | 8.58 | 2.80 | 3.27 | 5.76 | 13.45 |

**Table 4.** Aligned structures in group interactions.

| STRUCTURE | Primes | Alignment within speaker | Alignment between speakers | Total alignment | Total occurrences |
|---|---|---|---|---|---|
| *have* + NP | 15 | 8 | 4 | 12 | 27 |
| *have* + VP (inf) | 13 | 6 | 3 | 9 | 22 |
| complementiser *that* omitted | 17 | 8 | 6 | 14 | 31 |
| *going to* + VP | 22 | 10 | 5 | 15 | 37 |
| Imperative (*let's*) | 19 | 7 | 4 | 11 | 30 |
| non-finite relative clause | 11 | 4 | 3 | 7 | 18 |
| finite relative clause | 24 | 11 | 7 | 18 | 42 |
| stranded preposition | 8 | 5 | 3 | 8 | 16 |
| Comparative | 9 | 3 | 3 | 6 | 15 |
| *what kind of* + NP | 19 | 10 | 6 | 16 | 35 |
| NP + *like* + NP | 27 | 11 | 8 | 19 | 46 |
| TOTAL | 184 | 83 | 52 | 135 | 319 |
| Mean | 17 | 8 | 5 | 12 | 29 |
| SD | 6.18 | 2.81 | 1.79 | 4.47 | 10.50 |

To answer research question 2, a non-parametric Wilcoxon signed-rank test was carried out to examine whether there were significant differences between the alignment occurrences in pair and group interaction. The following report includes the test results with probability values and effect sizes *r*. The pair and group conditions were compared for the number of words, number of turns, number of primes, alignment between speakers, alignment within speakers, and total alignment. The results indicated that there was no significant difference between the number of words produced in pairs (*Md* = 276) and in groups (*Md* = 203), $z = -1.85$, $p = 0.064$, $r = 0.29$), but there was a significant difference between the number of turns in pair (*Md* = 47.5) and group (*Md* = 38.0) interaction, $z = -2.19$, $p = 0.029$, $r = 0.35$. For primes, the numbers were significantly higher in pair (*Md* = 14.0) than in group (*Md* = 9.5) interaction, $z = -3.14$, $p = 0.002$, $r = 0.50$. For total alignment, the numbers were also significantly higher in pairs (*Md* = 11.0) than in groups (*Md* = 7.0), $z = -3.73$, $p < 0.001$, $r = 0.59$. For between speaker alignment, there were significantly more occurrences in pairs (*Md* = 4.5) than in groups (*Md* = 2.5), $z = -3.78$, $p < 0.001$, $r = 0.60$. Similarly, for within speaker alignment, the number of occurrences was significantly higher in pairs (*Md* = 6.0) than in group (*Md* = 4.0) interaction, $z = -3.23$, $p = 0.001$, $r = 0.51$.

## 4. Discussion

The first research question asked if L2 speakers align their interaction when working on unscripted tasks in (a) a pair and (b) a group. The answer to this question is positive. A conversation analysis carried out on recorded conversations showed that interactants gradually developed their coordination, with priming and both within speaker and between speaker alignment occurring relatively frequently, particularly in pair interactions. However, this result needs to be considered with caution for at least two reasons. First, the sample was very small, with only 20 participants, so it is not possible to generalise these findings beyond this group to the entire L2 learner population at an upper intermediate level. The second reason is that this group of students have been studying and working together for two years, some of them even longer, which gives them a clear advantage in terms of developing the skills and competences associated with interactional alignment. In this period, they have had plenty of opportunities to interact among themselves, and this

must have contributed to their ease of communication and collaboration which is considered to be the basic precondition for interactive alignment to occur. Previous studies [19] have shown that peer interaction in the classroom where L2 learners know each other can boost structural priming and alignment.

The present findings are comparable with similar studies which used unstructured tasks [6,19,20]. In corpus-based research, Reitter et al. [26] found more priming and stronger between speaker alignment in dialogues that were task-oriented than in natural everyday conversation. In Dao et al.'s [6] study it was also found that the aligned structures were task driven. A question may be raised here related to the source of the repeated structures: is their use driven by alignment, or are they "task necessary" features if the task is to take part in a spontaneous interaction with minimal or no structure? One may argue that the latter is the case since in unscripted and unstructured tasks it may be more difficult than in tightly structured tasks to demonstrate that an optional syntactic structure could be used (such as, for example, the alternation between active and passive voice, or prepositional and double object dative in scripted tasks). In the present study it appeared that, specifically, modal verbs were extensively used as a function of the task type. For that reason, modals were excluded from coding. If they had been coded, they could have easily inflated the results since there were, in total, 109 modals used in pair interaction and 93 modals in group interaction. For the same reason, constructions using existential "there" and "need" + NP were also excluded from coding. These constructions were deemed to be directly related to the task necessary features, and therefore, less influenced by interactional alignment. For other constructions, those that were coded as primes and alignment, it may be possible to demonstrate that certain meanings could have been expressed in a different way. For example, in excerpt (2), instead of saying "people who will be living alone", following the relative clause prime "people who will be living there", one could have said "people/residents without families". Or, instead of saying "(they don't have) that big pay to support themselves", following the non-finite relative clause prime "a decent place to live in", one could have said "financially less able people", or "poorer people". However, conceptually, it may be impossible to isolate alignment and separate it from task essential features in natural interaction (such as speaker coordination, synchronisation, and convergence) because these features are also the preconditions for alignment to occur. In short, it appears that alignment and essential features of spontaneous interaction are the two sides of one coin: one cannot take place without the other, and their co-existence is fully acknowledged in Pickering and Garrod's [5] interactive alignment model.

### *4.1. Differences between Pair and Group Interaction*

The second research question asked if there was a difference in student behaviour associated with different grouping (two versus four participants). The results of the statistical analysis indicate that there are significant differences between the interaction occurring between two people and in a group of four. Even though in some parts of group interaction participation was seemingly distributed equally, most of the time and in most groups the distribution of talk was unbalanced. Some participants were proactive most of the time, but some were not; for example, one participant contributed only five words in two turns during the whole group discussion, and another one contributed eighteen words in three turns. Both were reasonably active in pair interactions. There may be three reasons which cannot be excluded when explaining the different behaviour in different groupings. Firstly, although the two tasks had the same global characteristics, there may have been some slight differences between them, at least in the topic, i.e., in the interest each of the two tasks generated in students.

The second explanation might be related to individual variables [24] whereby certain students might be too shy or feel intimidated if asked to participate in a group conversation. For example, Van Moere's [46] research identified personality and talkativeness as social factors influencing group dynamics in spoken exam tasks. Nakatsuhara [47], who examined group dynamics in spoken exam tasks with three and four speakers, found that in groups

of four, there was more avoidance behaviour and less collaborative atmosphere, as well as more mechanical turn-taking. This study also found that the students' extraversion had more impact on topic initiation in groups of four, and the level of extraversion impacted the amount of talk produced by each member. This suggests that individual factors such as extraversion may have significant implications for group dynamics, but also, that a group of four participants may not be an ideal structure for work on collaborative tasks. This might be a plausible answer for the differences found in the present study too, but it should be pointed out that the participants in the present study knew each other well and their participation in research was not part of a high stakes exam as it was in Nakatsuhara's study.

The third explanation also seems to be tenable, i.e., that the differences in the performance and the occurrence of interactive alignment stemmed from different group dynamics in pairs and in groups, irrespective of the individual factors. If interactive alignment is partly due to the interactive nature of dialogue, as Pickering and Garrod [48] suggested, then the degree of alignment should also mirror the nature of the interaction between the speakers and the listeners because the interactive alignment model assumes that successful interaction involves the development of aligned representations on different levels, including people, time, space, etc. Pickering and Garrod [5] suggest that there are fundamental differences between addressees and other listeners in group conversation, whereby stronger alignment is predicted for addressees than other listeners. Branigan et al.'s [49] study found that syntactic alignment can occur in multi-party interaction, but it is modulated by participant role of the speaker relative to the source utterance. Stronger alignment was found when the participant was the addressee of the source utterance. A subsequent study conducted by Branigan et al. [50] confirmed these results, demonstrating again that participants were more influenced by the prime if it had been directed at them rather than at a third participant in conversation.

These findings confirm that by its nature, conversation or interaction is both a social and a cognitive activity [49,51], and that structural priming in dialogue reveals its social communicative dimension. From the social psychological perspective, the balance ensured by one-speaker rule in a conversation [30] is not guaranteed in conversations involving three or more people. The reason is that in a group, when one person stops speaking, any of the others may start, and this makes it possible for some people to choose not to speak. Sacks et al. [2] even suggested that a "schism" may take place in situations where a group of four may split into two pairs and continue their interaction in parallel. Gibson's [52] research shows that the interaction of numbers and temporality creates a conflict between responsibility and opportunity for producing content. This "conversational latency" is seen as a consequence of speaker linearisation in group talk.

### *4.2. Implications for Teaching*

Considering the ubiquitous nature of priming and alignment, it is surprising that such naturally occurring processes are not used in language teaching more systematically. Specifically, structured and scripted tasks could be widely used to practice specific syntactic structures, such as passive, dative constructions, relative clauses, non-finite clauses, etc., as evidenced in several research studies [13,14,16,53]. Less structured tasks are more appropriate for more advanced language learners where they could employ their language skills to express more personal and creative ideas. Less structured tasks, such as problem solving and decision making, can be particularly beneficial for developing interactional competence, which is characterised by topic and turn management, negotiation of meaning, and repair strategies [33]. The present study has also touched upon the issue of interactive alignment in multi-party conversations. Even though this would require further and more comprehensive exploration, group structure might become more pronounced an issue when priming and alignment are viewed in light of their utility for language learning. In language teaching, pair work and group work are frequently used classroom practices aimed at developing communicative competence: consideration of priming techniques for teaching purposes should consider the present findings and balance pair work with group

work. However, to achieve equal and balanced participation in a group, clear roles should be given to each participant in the group. Otherwise, the group members might not be able to have equal chances for developing their interactional competence.

## 5. Conclusions

The tasks in the present study were of relatively short duration; thus, more tasks and more data with a larger number of participants are needed to make firm conclusions. Furthermore, some differently framed, naturally occurring conversations could be included, such as discussions about movies or books, or narratives concerning past experiences. An investigation of individual learner variables in relation to interactant behaviour and subsequent learning [24] is also a topic that requires further exploration, particularly in relation to participation dynamics in communicative tasks. Future investigations in L2 learning and use will certainly have to build upon the fact that the processes in L2 are interconnected, involving cognitive, psychological, psycholinguistic, and social dimensions. This leaves little space for isolated views unrelated to the context in which language is learnt or used.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13080792/s1, Supplementary file S1: Task Instructions; Supplementary file S2: Transcripts; Supplementary file S3: Transcription Conventions.

**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 Education Ethics Committee, University of York, UK (Ref. 15/006; Data protection reg. no.: z4855807; 16/03/2015).

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

**Data Availability Statement:** The data presented in this study (for which participants' consent was obtained) are available in supplementary material (Supplementary file S2).

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Levinson, S.C. Turn-taking in human communication—Origins and implications for language processing. *Trends Cogn. Sci.* **2016**, *20*, 6–14. [CrossRef]
2. Sacks, H.; Schegloff, E.A.; Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language* **1974**, *50*, 696–735. [CrossRef]
3. Schegloff, E.A. Introduction to sequence organization. In *Sequence Organization in Interaction: A Primer in Conversation Analysis*; Schegloff, E.A., Ed.; Cambridge University Press: Cambridge, UK, 2007; Volume 1, pp. 1–12. [CrossRef]
4. Jordan, B.; Henderson, A. Interaction analysis: Foundations and practice. *J. Learn. Sci.* **1995**, *4*, 39–103. [CrossRef]
5. Pickering, M.J.; Garrod, S. Interactive alignment and language use. In *The Oxford Handbook of Language and Social Psychology*; Holtgraves, T.M., Ed.; Oxford University Press: Oxford, UK, 2014. [CrossRef]
6. Dao, P.; Trofimovich, P.; Kennedy, S. Structural alignment in L2 task-based interaction. *ITL—Int. J. Appl. Linguist.* **2018**, *169*, 293–320. [CrossRef]
7. Doughty, C. Cognitive underpinnings of focus on form. In *Cognition and Second Language Instruction*; Robinson, P., Ed.; Cambridge University Press: Cambridge, UK, 2001; pp. 206–257.
8. Ellis, R. *Learning a Second Language through Interaction*; John Benjamins: Amsterdam, The Netherlands; Philadelphia, PA, USA, 1999.
9. Gass, S.M. *Input, Interaction, and the Second Language Learner*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1997.
10. Mackey, A. *Conversational Interaction in Second Language Acquisition: A Collection of Empirical Studies*; Oxford University Press: Oxford, UK, 2007.
11. Mackey, A. *Input, Interaction, and Corrective Feedback in L2 Learning*; Oxford University Press: Oxford, UK, 2012.
12. Long, M.H. The role of the linguistic environment in second language acquisition. In *Handbook of Second Language Acquisition*; Ritchie, W.C., Bhatia, K., Eds.; Academic Press: San Diego, CA, USA, 1996; pp. 413–468.
13. K McDonough, K.; Kim, Y. Syntactic priming, type frequency, and EFL learners' production of wh-questions. *Mod. Lang. J.* **2009**, *93*, 386–398. [CrossRef]

14. McDonough, K.; Mackey, A. Responses to recasts: Repetitions, primed production, and linguistic development. *Lang. Learn.* **2006**, *56*, 693–720. [CrossRef]
15. McDonough, K.; Mackey, A. Syntactic priming and question development. *Stud. Second Lang. Acquis.* **2008**, *30*, 31–47. [CrossRef]
16. Kim, Y.; Jung, Y.; Skalicky, S. Linguistic alignment, learner characteristics, and the production of stranded prepositions in relative clauses: Comparing FTF and SCMC contexts. *Stud. Second Lang. Acquis.* **2019**, *41*, 937–969. [CrossRef]
17. Kim, Y.; McDonough, K. Prime repetition and Korean EFL learners' comprehension and production of passives. *Int. Rev. Appl. Linguist. Lang. Teach.* **2016**, *54*, 319–346. [CrossRef]
18. Trofimovich, P.; McDonough, K.; Neumann, H. Using collaborative tasks to elicit auditory and structural priming. *TESOL Q.* **2013**, *47*, 177–186. [CrossRef]
19. Collentine, J.; Collentine, K. A corpus approach to studying structural convergence in task-based Spanish L2 interactions. In *Second Language Interaction in Diverse Educational Contexts*; McDonough, K., Mackey, A., Eds.; John Benjamins: Amsterdam, NL, USA, 2013; pp. 168–187.
20. Michel, M.; Cappellini, M. Alignment during synchronous video versus written chat L2 interactions: A methodological exploration. *Annu. Rev. Appl. Linguist.* **2019**, *39*, 189–216. [CrossRef]
21. Garrod, S.; Pickering, M.J. Alignment in dialogue. In *The Oxford Handbook of Psycholinguistics*; Gaskell, M.G., Ed.; Oxford University Press: Oxford, UK, 2007; pp. 442–452. [CrossRef]
22. Costa, A.; Pickering, M.J.; Sorace, A. Alignment in second language dialogue. *Lang. Cogn. Process.* **2008**, *23*, 528–556. [CrossRef]
23. Markman, A.B.; Kim, K.; Larkey, L.B.; Narvaez, L.; Stilwell, C.H. One alignment mechanism or many? *Behav. Brain Sci.* **2004**, *27*, 204–205. [CrossRef]
24. Jackson, C.N. Second language structural priming: A critical review and directions for future research. *Second Lang. Res.* **2017**, *34*, 539–552. [CrossRef]
25. Bernolet, S.; Collina, S.; Hartsuiker, R.J. The persistence of syntactic priming revisited. *J. Mem. Lang.* **2016**, *91*, 99–116. [CrossRef]
26. Reitter, D.; Moore, J.D.; Keller. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)*; Cognitive Science Society: Vancouver, BC, Canada, 2010; pp. 685–690.
27. Hartsuiker, R.J.; Bernolet, S. The development of shared syntax in second language learning. *Biling. Lang. Cogn.* **2017**, *20*, 219–234. [CrossRef]
28. Jackson, C.N.; Ruf, H.T. The importance of prime repetition among intermediate-level second language learners. *Stud. Second Lang. Acquis.* **2018**, *40*, 677–692. [CrossRef]
29. Bernolet, S.; Hartsuiker, R.J.; Pickering, M.J. From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition* **2013**, *127*, 287–306. [CrossRef]
30. Rœver, C.; Kasper, G. Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Lang. Test.* **2018**, *35*, 331–355. [CrossRef]
31. Schegloff, E.A. Turn organization: One intersection of grammar and interaction. In *Interaction and Grammar*; Ochs, E.A., Schegloff, E.A., Thompson, S.A., Eds.; Cambridge University Press: Cambridge, UK, 1996; pp. 52–133. [CrossRef]
32. Plough, I.; Banerjee, J.; Iwashita, N. Interactional competence: Genie out of the bottle. *Lang. Test.* **2018**, *35*, 427–445. [CrossRef]
33. Galaczi, E.; Taylor, L. Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Lang. Assess. Q.* **2018**, *15*, 219–236. [CrossRef]
34. Gries, S.T.; Kootstra, G.J. Structural priming within and across languages: A corpus-based perspective. *Biling. Lang. Cogn.* **2017**, *20*, 23–250. [CrossRef]
35. Hulstijn, J.H.; Young, R.F.; Ortega, L.; Bigelow, M.; DeKeyser, R.; Ellis, N.C.; Lantolf, J.P.; Mackey, A.; Talmy, S. Bridging the gap: Cognitive and social approaches to research in second language learning and teaching. *Stud. Second Lang. Acquis.* **2014**, *36*, 361–421. [CrossRef]
36. King, K.A.; Mackey, A. Research methodology in second language studies: Trends, concerns, and new directions. *Mod. Lang. J.* **2016**, *100* (Suppl. 1), 209–227. [CrossRef]
37. Lazaraton, A. Current trends in research methodology and statistics in applied linguistics. *TESOL Q.* **2000**, *34*, 175. [CrossRef]
38. Markee, N. Doing, and justifying doing, avoidance. *J. Pragmat.* **2011**, *43*, 602–615. [CrossRef]
39. Samuda, V.; Bygate, M. *Tasks in Second Language Learning*; Palgrave Macmillan: London, UK, 2008.
40. Barnes, D.; Todd, F. *Communication and Learning Revisited: Making Meaning through Talk*; Heinemann: Portsmouth, NH, USA, 1995.
41. Goldberg, A.E. *Constructions at Work: The Nature of Generalization in Language*; Oxford University Press: Oxford, UK, 2005.
42. Bock, K.; Griffin, Z.M. The persistence of structural priming: Transient activation or implicit learning? *J. Exp. Psychol. Gen.* **2000**, *129*, 177–192. [CrossRef]
43. Reitter, D.; Moore, J.D. Alignment and task success in spoken dialogue. *J. Mem. Lang.* **2014**, *76*, 29–46. [CrossRef]
44. Plonsky, L.; Derrick, D. A meta-analysis of reliability coefficients in second language research. *Mod. Lang. J.* **2016**, *100*, 538–553. [CrossRef]
45. Schegloff, E.A. Analysing single episodes of interaction: An exercise in conversation analysis. *Soc. Psychol. Q.* **1987**, *50*, 101–114. [CrossRef]
46. Van Moere, A. Validity evidence in a university group oral test. *Lang. Test.* **2006**, *23*, 411–440. [CrossRef]

47. Nakatsuhara, F. Effects of test-taker characteristics and the number of participants in group oral tests. *Lang. Test.* **2011**, *28*, 483–508. [CrossRef]
48. Pickering, M.J.; Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* **2004**, *27*, 169–190. [CrossRef]
49. Branigan, H.P.; Pickering, M.J.; McLean, J.F. Priming prepositional-phrase attachment during comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31*, 468–481. [CrossRef]
50. Branigan, H.P.; Pickering, M.J.; McLean, J.F.; Cleland, A.A. Syntactic alignment and participant role in dialogue. *Cognition* **2007**, *104*, 163–197. [CrossRef]
51. Rœver, C.; McNamara, T. Language testing: The social dimension. *Int. J. Appl. Linguist.* **2006**, *16*, 242–258. [CrossRef]
52. Gibson, D.R. Marking the turn: Obligation, engagement, and alienation in group discussions. *Soc. Psychol. Q.* **2010**, *73*, 132–151. [CrossRef]
53. Trofimovich, P.; McDonough, K. (Eds.) *Applying Priming Methods to L2 Learning, Teaching and Research: Insights from Psycholinguistics*; John Benjamins Publications: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2011.

education sciences

MDPI

*Article*

# Perspectives on the Effectiveness of Madrid's Regional Bilingual Programme: Exploring the Correlation between English Proficiency Level and Pre-Service Teachers' Beliefs

**Lidia Mañoso-Pacheco [1,*] and Roberto Sánchez-Cabrero [2]**

[1] Department of Philology and Its Didactics, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[2] Department of Evolutive and Education Psychology, Universidad Autónoma de Madrid, 28049 Madrid, Spain; roberto.sanchez@uam.es
* Correspondence: lidia.mannoso@uam.es

**Abstract:** This study aims to examine the opinions of prospective pre-primary and primary teachers about Madrid's regional Bilingual Programme in Spain, assessing the correlations between their self-perceived level of English and their positioning concerning the effectiveness of the regional programme. Although there is a growing body of research in the field of education on how CLIL (Content and Language Integrated Learning) provisions impact in-service teachers' attitudes concerning bilingual education, there is a dearth of literature on the way student teachers depict the teaching reality. Thus, this paper explores pre-service teachers' beliefs towards the Bilingual Programme via an ad hoc questionnaire, administered to a non-probabilistic sample of 170 undergraduate students at the Universidad Autónoma de Madrid. The data collected were explored using *Chi-square* and *Somers' D* tests. The results show that the self-perceived English level, greatly determined by prior bilingual schooling, has a strong influence on their perceptions about the Bilingual Programme. The findings also indicate that, although the learning experience at the pre-primary stage is valued positively by students, the acquisition of curricular contents in primary education is seen as negatively affected due to English medium instruction.

**Keywords:** bilingual education; CLIL; English level; pre-service teachers

**Citation:** Mañoso-Pacheco, L.; Sánchez-Cabrero, R. Perspectives on the Effectiveness of Madrid's Regional Bilingual Programme: Exploring the Correlation between English Proficiency Level and Pre-Service Teachers' Beliefs. *Educ. Sci.* **2022**, *12*, 522. https://doi.org/10.3390/educsci12080522

Academic Editor: Martin Howard

Received: 13 July 2022
Accepted: 28 July 2022
Published: 31 July 2022



## 1. Introduction

The Region of Madrid Bilingual Programme (henceforth, BP) or "Programa Bilingüe de la Comunidad de Madrid" is a large education plan working in state schools that began in 2004 in primary education, which has progressively been extended to the secondary level. Although BP training is not mandatory, teachers need to be accredited in foreign language proficiency to work in bilingual schools in this region. This compulsory requirement can be met through a test that evaluates the knowledge to teach the Advanced English curriculum, comprising the assessment of the candidate's methodological skills, or by holding a university degree or an official language certification equivalent to CEFR (Common European Framework of Reference for Languages) level C1 or above [1]. Until the publication of Order 1672/2009, the accreditation process was solely focused on the criterion of linguistic competence in English; however, due to the growth of the BP in the region, it now includes a methodological training evaluation. Although it is not a requirement as such, bilingual schools value positively that teachers are specialised in English as a Foreign Language (EFL), as this provides them with deeper knowledge and learning skills related to this area.

The BP implemented in the region rests on a CLIL (Content and Language Integrated Learning) approach, whereby certain content subjects are taught in a foreign language—principally English. CLIL may be then regarded as "a dual-focused education approach in which an additional language is used for the learning and teaching of both content and language" [2] (p. 1). Currently, 50.4% of primary schools and 63.6% of secondary schools

in Madrid are bilingual in their entirety [3]. In the case of pre-primary education, the BP started in the academic year 2016–2017, and nowadays, 126 state schools participate in the programme from 3 to 6 years old [4].

CLIL provisions are ample in Spain, and empirical research findings indisputably report the superiority of CLIL instruction as opposed to language-driven tuition, particularly over the long run. CLIL is supported by well-grounded research in Europe, which affirms that CLIL students generally outperform their non-CLIL counterparts concerning English proficiency [5–7] and that this teaching approach also offers cognitive advantages to students [8]. The literature has provided findings that the implementation of CLIL helps develop among students a positive attitude towards multilingual instruction and the CLIL approach itself, alongside higher motivation towards foreign language learning and the promotion of intercultural competence [9–12]. Hence, CLIL has proven to be effective in the overall learning experience of students, comprising "content (i.e., subject knowledge), linguistic and intercultural competence" [13] (p. 245). The learning experience in CLIL settings is cross-curricular or transversal, in the sense that this approach allows the acquisition of inter-related content exposing learners to real-life communicative situations in which prior knowledge scaffolds the attainment of new information and the reinforcement of target language proficiency. According to Castey and Paz-Albo [14], exposition to two languages from early childhood (0–6 years), i.e., during the critical period, is beneficial for learners' cognitive development since this is the optimal period in which the brain is configurated and the neural networks are shaped. It is at this moment when mirror neurons are at work to link words (input) to their implicit relational experience (world). Hence, those infants who are exposed to bilingual language input often show higher cognitive flexibility and the enhanced performance of executive functions, resulting in later success in problem-solving abilities [15].

Despite a considerable number of investigations centred on the benefits of CLIL in both pre-primary and primary education [7,8,16–18], its pedagogical implementation in Spain is a matter of controversy. Teachers frequently point to structural difficulties in implementing this approach and the need for further methodological training, enhanced coordination, and more resources to improve the quality of education under this curricular paradigm [19]. Moreover, in a study by Alonso-Belmonte and Fernández-Agüero [20], it was found that in-service teachers working in Madrid's BP consider that they lack linguistic expertise and/or content knowledge, which makes them feel insufficiently prepared to master the various challenges that CLIL entails. The limited linguistic performance of certain teachers and/or their own perception of the students' limited linguistic proficiency may lead to content simplification [21], that is, focusing on conceptual comprehension rather than the practical application of those notions, thus lowering learners' cognitive development. Students tend to learn a new concept by receiving an insufficiently detailed explanation or by just seeing its correspondence with their mother tongue, without deeply reasoning about its implications and/or knowing in depth its application to a real-life situation.

Although this approach has been implemented in Spain since 1996, an English First English Proficiency Index (EF EPI) report published in 2021 [22], which evaluated data from non-English speakers over the world, placed Spain in the 33rd position in the global ranking and concluded that, unlike in other parts of Europe, the improvement in English is stagnated in this country. Among the four strategic objectives that are mentioned within the framework of ET (Education and Training) Monitor elaborated by the European Commission in 2020 [23], enhancing "the quality and efficiency of education and training" is mentioned. To that end, a number of national priorities are set for each Member State; one of them is spending some time studying or training abroad in the case of higher education graduates and professionals with an initial vocational qualification. Concerning Spain, learning mobility figures provided by ET Monitor show that the country is below the EU (European Union) average, which may have a negative effect on students' development of foreign language proficiency and intercultural communicative competence.

The debate on the effectiveness of the BP goes beyond the academic field and is present in Spanish society at all levels, thus affecting family schooling decisions. Those parents against CLIL complain that students in BP courses learn neither the language nor the subject matter adequately and that this approach is also intensifying school segregation [24,25]. According to Ferrer and Gortazar's [26] report, based on PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) measurements, Spain is the EU country that displays the highest level of school segregation at the primary school level, with Madrid being the region with the highest rate. In this region, there are high proportions of similar students segregated by socioeconomic status. However, there is no clear evidence that this fact is directly linked to the BP. This subject is not only questioned among stakeholders, but the debate over the effectiveness of BPs is also registered in the Spanish national press. News that bilingual schools are dropping out of the BP is reaching the media, contributing to increases in the resistance of public opinion towards the implementation of this dual-focused programme.

To date, a considerable number of studies have been published on in-service teachers' understanding, knowledge, perceptions, and attitudes concerning BPs and CLIL education in Spain; see [9,20,27,28] and elsewhere [29–32]. However, there is scarce knowledge on prospective pre-primary and primary teachers' opinions in this regard. Hence, exploring student teachers' perceptions is critical as their professional career in Madrid will connect with CLIL in one way or another. Only a deep understanding of student teachers' beliefs and opinions can inform teacher education programmes to improve training and practice. The concept of 'belief' in the field of education entails an examination of the teaching–learning process from a particular angle [33]. Beliefs are difficult to analyse because they are experiential, mediated, and sometimes paradoxical [34]. They have also a strong affective–evaluative component [35], which, in the case of CLIL effectiveness, is likely to be influenced by the existing polarised social debate in Spain around this issue.

At this point, the questions that guide our study are: what do future pre-primary and primary teachers think about BPs? Particularly, are their perspectives influenced by their linguistic background? Two research objectives are proposed in this paper. First, (1) we aim to describe the opinions of prospective pre-primary and primary teachers in Spain in relation to bilingual education considering their self-perceived level of English and education background. Secondly, (2) the purpose is to show the correlations between their self-perceived level of English and different factors related to bilingual education.

## 2. Materials and Methods

### *2.1. Experimental Design*

A descriptive, cross-sectional, and correlational study of the profile and beliefs of students enrolled in education degrees in Spain regarding the BP was designed. To this end, a descriptive statistical study was carried out using the frequency distributions and percentages of the nominal and ordinal variables evaluated. For the correlational analyses, contingency tables via a *Chi-square* non-parametric test were employed, including *Somers' D* to determine the orientation of the relationship between the variables analysed and the effect size.

Results were considered significant when the significance level exceeded 95% ($\alpha \geq 0.05$ in dichotomous variables), and Bonferroni correction was applied when there were more than two levels in the variable.

### *2.2. Participants*

This study collected and analysed the opinions of 170 students from different degrees in education (pre-primary education, primary education, and the joint degree in pre-primary and primary) enrolled in the subjects "English as a Foreign Language I and II" at a university in Madrid (Spain) in the 2020–2021 academic year. The study participants in this research were viewed as key elements that may contribute to fostering educational effectiveness [36]. In addition, they were seen as representative of the whole trainee teaching

population since, as Skinnari and Bovellan claim, "although teacher's attitudes, assumptions and expectations [ . . . ] vary individually, they are situationally constructed and reflect the values of the communities they belong to" [37] (p. 148). The sampling method was based on clusters. Table 1 below presents the sample according to the attributive variables selected in this study.

**Table 1.** Sample description.

| **University Degree** | **Frequency** | **%** |
|---|---|---|
| Pre-primary education | 63 | 37.1 |
| Primary education | 56 | 32.9 |
| Joint degree in pre-primary and primary education | 51 | 30.0 |
| **Academic year** | **Frequency** | **%** |
| 1st | 5 | 2.9 |
| 2nd | 58 | 34.1 |
| 3rd | 106 | 62.4 |
| 4th | 1 | 0.6 |
| **Age category** | **Frequency** | **%** |
| 18–20 years | 79 | 46.5 |
| 21–23 years | 75 | 44.1 |
| 24–26 years | 12 | 7.1 |
| 27–29 years | 3 | 1.8 |
| 30 years or more | 1 | 0.6 |
| **Gender** | **Frequency** | **%** |
| Male | 23 | 13.5 |
| Female | 147 | 86.5 |
| **Official certificate of English (C1 or above)** | **Frequency** | **%** |
| No | 157 | 92.4 |
| Yes | 13 | 7.6 |
| **Total Sample** | **170** | **100.0** |

Table 1 clearly shows that the distribution of students according to the different university degrees in the sample obtained was balanced. The large majority of learners were in their third year (62.4%), aged under 24 years (90.6%), and were mostly female (86.5%). In addition, the vast majority of research participants did not hold a certificate of English level equivalent to C1 or higher (92.4%).

### *2.3. Study Variables and Evaluation Instrument*

An ad hoc questionnaire was designed to gather information for the purposes of this research. This instrument was implemented in order to obtain the specifics of the attributive variables evaluated and the reflections of the study participants concerning the items set. It was a short questionnaire with 17 closed-ended questions, distributed into three sections. The first 6 questions (Section 1) addressed the attributive variables of the study: *sex*, *age*, *university degree*, *academic year*, and *undergraduate specialisation*, together with *self-perceived proficiency in English*. Section 2, comprising items 7 to 14, enquired about the trainees' level of English, their background as EFL learners, and their general understanding of the regional BP. Finally, Section 3 intended to map the variety of positions

adopted by the trainees regarding the regional BP (items 15 and 16), as well as the impact of the programme on the pre-service teachers' professional careers (item 17). The evaluation instrument was previously validated by a group of three qualified university experts in the field of education in April 2021. The reviewers determined that items were adequate, sufficient, and relevant and did not contain biased content or common errors such as leading, confusing, or double-barrelled questions. Later, the authors piloted and validated the questionnaire on a small sample of 15 respondents. The questionnaire was administered in Spanish to encourage the student teachers' participation.

Regarding the intervening variables, this study collected information on 15 different variables, of which 3 were nominal (*university degree*, *gender of participants*, and *school type in primary and secondary education*), 5 were dichotomous with Yes/No answers (*English as favourite subject*, *English as the first choice of undergraduate specialisation*, *C1 certificate in English*, *stay in an English-speaking country*, and *bilingual schooling in primary and secondary education*), and 7 were ordinal with three or more response options (*age*, *academic year*, *self-perceived proficiency in English*, *do you consider your level of English is adequate to develop your future teaching career as a teacher in a bilingual school*? *Do you think that the curricular contents in Primary Education are learned worse*, *better*, *or in the same way in English than in Spanish*? *Does the English BP improve or worsen the overall learning experience* (*i.e.*, *cognitive*, *affective*, *linguistic, and subject knowledge skills*) *in pre-primary Education*? *Do you think that the existence of the BP can help you obtain a permanent teaching position more quickly*?).

It should be noted that those variables in which the participant was asked about two different education stages (pre-primary or primary education, primary or secondary education) included two items to define each variable. These are the cases that follow: *school type in primary and secondary school*, and *bilingual schooling in primary and secondary education*.

## 3. Results

This study departed from the assumption that the self-perceived level of English and education background of participants may have an influence on their position concerning the regional BP. Thus, having described the profile of the students of the official degrees in education in Spain in the Methods section, the first research objective is tackled. To that end, the frequency of distribution of the sample in the different self-assessment items related to bilingual education is shown in Table 2 below.

From the analysis of Table 2, it was possible to confirm that English was the favourite subject for only 21.2% of the sample, showing very similar percentages as the participants' interest in pursuing an English undergraduate specialisation (17.1%). A total of 41.8% of the students considered their English level to be intermediate, a level that according to less than half of the total respondents is insufficient to work as a teacher in a BP (44.7%). Very few of the participants had been on an international stay in a country with English as the official language (7.6%), and the schools where they studied in primary and secondary education are mostly state (52.4% and 55.3%, respectively) and non-bilingual (70% and 68.2%, respectively) schools. Regarding their specific opinion of bilingual education depending on the educational stage, it was possible to state that there were opposite results. Students were mostly of the opinion that it improves the learning experience in pre-primary education (70%); however, in line with Coonan's [21] and Alonso-Belmonte and Fernández-Agüero's [20] findings, they considered it to worsen the acquisition of content in primary education (64.7%). Finally, 45.3% of the selected sample thought that the existence of the BP could help them obtain a permanent teaching position more quickly.

In response to the second research objective, assessing the possible connection between the participants' self-perceived level of English and their position towards the regional BP, Table 3 below shows the correlations of all the variables evaluated with respect to the self-perceived English level using *Chi-square* and *Somers' D* significance.

**Table 2.** Self-assessment responses on different aspects related to bilingual education.

| **University Degree** | **Frequency** | **%** |
|---|---|---|
| No | 134 | 78.8 |
| Yes | 36 | 21.2 |
| **Interest in English as the first choice for the undergraduate specialisation** | **Frequency** | **%** |
| No | 141 | 82.9 |
| Yes | 29 | 17.1 |
| **In your opinion, what is your current English level?** | **Frequency** | **%** |
| Elementary (A1-A2) | 30 | 17.6 |
| Intermediate (B1) | 71 | 41.8 |
| Upper intermediate (B2) | 56 | 32.9 |
| Advanced (C1-C2) | 13 | 7.6 |
| **International stay in an English-speaking country** | **Frequency** | **%** |
| No | 157 | 92.4 |
| Yes | 13 | 7.6 |
| **School type (primary education)** | **Frequency** | **%** |
| Private | 6 | 3.5 |
| Subsidised | 75 | 44.1 |
| State | 89 | 52.4 |
| **School type (secondary education)** | **Frequency** | **%** |
| Private | 14 | 8.2 |
| Subsidised | 62 | 36.5 |
| State | 94 | 55.3 |
| **Bilingual schooling (primary education)** | **Frequency** | **%** |
| No | 119 | 70.0 |
| Yes | 51 | 30.0 |
| **Bilingual schooling (secondary education)** | **Frequency** | **%** |
| No | 116 | 68.2 |
| Yes | 54 | 31.8 |
| **Do you consider that your level of English is sufficient to develop your future career as a teacher in a bilingual school?** | **Frequency** | **%** |
| No | 76 | 44.7 |
| Not sure | 29 | 17.1 |
| Yes | 65 | 38.2 |
| **Does the English BP improve or worsen the overall learning experience (i.e., cognitive, affective, linguistic, and subject knowledge skills) in pre-primary education?** | **Frequency** | **%** |
| Worsens | 42 | 24.7 |
| Does not affect | 9 | 5.3 |
| Improves | 119 | 70.0 |
| **Do you think that the curricular contents in primary education are learned worse, better, or in the same way in English than in Spanish?** | **Frequency** | **%** |
| Worse | 110 | 64.7 |
| Does not affect | 60 | 35.3 |
| Better | 0 | 0 |
| **Do you think that the existence of the BP can help you obtain a permanent teaching position more quickly?** | **Frequency** | **%** |
| No | 40 | 23.5 |
| Maybe | 53 | 31.2 |
| Yes | 77 | 45.3 |
| **Total Sample** | **170** | **100.0** |

**Table 3.** Contingency table of self-perceived English level and the rest of the variables evaluated.

| Variable | *Chi-Square* | *Somers' D* |
|---|---|---|
| University degree | 21.87 ** | 0.269 ** |
| Academic year | 20.361 * | 0.26 ** |
| Age range | 4.901 | −0.106 |
| Gender of participants | 2.363 | 0.024 |
| English as a favourite subject | 36.237 ** | 0.353 ** |
| Interest in English as an undergraduate specialisation (first choice) | 39.787 ** | 0.305 ** |
| Official certificate of English (C1 or above) | 14.248 ** | 0.175 ** |
| International stay in an English-speaking country | 7.154 | 0.169 ** |
| School type (primary education) | 8.525 * | −0.022 |
| School type (secondary education) | 3.615 | 0.018 |
| Bilingual schooling (primary education) | 19.426 ** | 0.288 ** |
| Bilingual schooling (secondary education) | 13.083 ** | 0.25 ** |
| Do you consider that your level of English is sufficient to develop your future career as a teacher in a bilingual school? | 3.981 | 0.006 |
| Does the BP in English improve or worsen the overall learning experience (i.e., cognitive, affective, linguistic, and subject knowledge skills) in pre-primary education? | 9.973 | −0.151 * |
| Do you think that the curricular contents in primary education are learned worse, better, or in the same way in English than in Spanish? | 4.898 | −0.006 |
| Do you think that the existence of the BP can help you obtain a permanent teaching position more quickly? | 12.866 * | 0.126 |

* Correlation is significant at the 0.05 level/** correlation is significant at the 0.01 level.

Table 3 clearly reveals how the self-perceived English level correlated directly and significantly with the university degree; in particular, there was a higher self-perceived level in the joint degree students than in the rest of the student groups, and the lowest self-perceived level was found in pre-primary education students. As regards the academic year, it was found that the higher the course, the higher the self-perceived English level. There were other variables, such as English as a favourite subject, interest in an English undergraduate specialisation, holding a certificate in advanced English (C1), or bilingual schooling, either in primary or secondary education, which also showed significant correlation with the self-perceived level of English. On the other hand, *Chi-square* showed significant results in the school type in primary education and the variable *do you think that the existence of the BP can help you obtain a permanent teaching position more quickly*? However, it did not occur using *Somers' D* test. This fact means that both variables covaried in a non-random way with the self-perceived English level and that their relationship was not linear, but curvilinear, which means that *Somers' D* did not show significance.

Finally, it should be noted that, unlike *Chi-square* testing, *Somers' D* is significant for the variables of *international stay in an English-speaking country* and *does the BP in English improve or worsen the overall learning experience in pre-primary education*? This finding may lead us to interpret that the relationship between the two variables was not significant, but there was a clear indication of a trend. In the case of the latter variable, our study sample shows that the lower the level of English, the higher the pre-service teachers rank the regional BP as concerns the overall learning experience. In the case of *international stay in an English-speaking country*, this fact may be explained due to the disproportionality of the sample, but in the case of the variable *does the BP in English improve or worsen the overall*

*learning experience in pre-primary education?*, we can deduce that the inverse orientation is not strong enough to be significant.

## 4. Discussion and Conclusions

This paper analysed the beliefs of 170 pre-service teachers concerning the effectiveness of the regional BP. Regarding the first research objective, the description of the opinions of prospective pre-primary and primary teachers in Spain in relation to bilingual education considering their self-perceived level of English and education background, it can be stated that, according to the percentages observed in the study, the self-perceived English level has a strong influence on participants' perception of the BP. Holding a certificate of English level equivalent to C1 or higher (7.6%), which is a compulsory requirement to teach the Advanced English curriculum, correlates significantly with other variables, such as having an interest in pursuing the English undergraduate specialisation (17.1%), valued very positively for working in bilingual courses, or having been on an international stay in an English-speaking country (7.6%), which usually help students develop foreign language proficiency and intercultural communicative competence [23].

Further, the pre-service teachers' perception of their self-perceived proficiency in the target language is strongly determined by previous bilingual schooling at pre-primary and/or primary stages (70% and 68.2%, respectively), that is, those participants who have been enrolled in BPs prior to their university studies tend to give a better assessment of their own English level, supporting its effectiveness. In line with the previous research findings [9–12], the implementation of CLIL helps develop a positive attitude towards multilingual instruction, alongside higher motivation towards foreign language learning and the promotion of intercultural competence. Dual immersion programmes therefore have a direct impact on a student's performance in English. As a consequence of the sample population profile obtained, not many participants (38.2) saw themselves as prepared to meet the level required to work in a bilingual school. This finding complies with the EF EPI report [22], which concluded that Spain falls behind other countries in the global ranking as regards English proficiency.

It is interesting to note that although the learning experience at the pre-primary stage was valued positively by students (70%), the acquisition of the curricular contents in primary education was seen as negatively affected due to English-medium instruction (64.7%). According to Coonan's [21] study, content simplification is a result of the limited linguistic performance of certain teachers and/or their learners. There is a tendency to emphasise conceptual understanding over the practical application of the learning tasks, leading to a decrease in the students' own cognitive development. Alonso-Belmonte and Fernández-Agüero [20] also found that due to a lack of linguistic expertise, teachers tend to apply curricular adaptations that oversimplify the learning contents. Despite this, early exposition to English is generally seen as beneficial for learners' learning experience, understood as the development of cognitive, affective, linguistic, and subject knowledge skills [13]. The study participants did not question the relevance of exposing learners to two different languages during their critical period, which has not just been proven to be beneficial for the acquisition of the target language, but also for the development of cognitive flexibility and the performance of executive functions.

As concerns the second research objective, assessing the correlations between their self-perceived level of English and different factors related to bilingual education, the findings indicate that according to *Chi-square* and *Somers' D* tests, the self-perceived proficiency correlated directly and significantly with most of the variables analysed. It particularly had a strong connection with the following factors analysed in this research: interest in an English undergraduate specialisation, advanced certificate, and bilingual schooling. It is reasonable to think that students' perception of their own level of English directly connects with their education background, as well as their interest in their own subject specialisation during their undergraduate studies. Moreover, a curvilinear relationship was found between this variable and the fact of getting a permanent position more quickly due

to the existence of the BP. This finding may be explained considering that the higher the level of English, the more possibilities students consider themselves to have in acquiring a vacancy due to the foreign language requirements for entry into a CLIL centre. Finally, there was a clear indication of a trend between self-perceived English level and the assessment of the way the BP affects the overall learning experience in pre-primary education. As already indicated, participants with a low level of English tended to rank the learning experience derived from the regional BP more highly than the rest of the research groups. From this finding, we may conclude that pre-service teachers with high language proficiency appeared to be more critical of the BP, despite acknowledging they will probably have more job opportunities due to their high command of English.

In conclusion, the results obtained are satisfactory as they helped to complete the vision about the BP by analysing pre-service teachers' perspectives, which are highly significantly correlated with their self-perceived proficiency in English. As already mentioned, exploring student teachers' perceptions is critical as their professional career in Madrid will closely connect with the BP. The research findings underline the ongoing need to improve the methodological training of teachers working in CLIL in order to reverse the trend of stagnation in Spain as regards English proficiency levels. The accreditation procedure established in the region to work in bilingual sections should also incorporate passing a training course before the teacher joins the BP, providing educators with the required standards to cater to CLIL demands. As seen above, navigating lessons about curricular content via a foreign language usually poses a challenge for those teachers with a lower level of English. Some educators tend to opt for code-switching and/or content simplification as unique teaching strategies, to the detriment of CLIL learners. However, if they were given clear directions on how to scaffold students to become more independent and active learners and on the best way to show them the practical connection between what is taught in class and their real-life experience, learners would profit maximally from CLIL instruction. The results also reveal that there is still a long way to go to establish the necessary conditions for the real and effective implementation of bilingual education that might result in more motivated and communicatively competent L2 speakers. To that end, school administrators should create more meaningful opportunities for cross-curricular coordination between content-knowledge teachers and language specialists for the full integration of CLIL components. The introduction of financial incentives related to the completion of short-term retraining courses to upskill teachers' English proficiency and/or their methodological skills, with an emphasis on CLIL instruction, could also benefit the regional BP. To conclude, this research purports to have broadened the scope of study in this domain, emphasising the continuing need to improve teaching training in bilingual education.

**Author Contributions:** Conceptualization, L.M.-P.; methodology, R.S.-C.; software, R.S.-C.; validation, L.M.-P.; formal analysis, L.M.-P. and R.S.-C.; investigation, L.M.-P. and R.S.-C.; resources, L.M.-P.; data curation, R.S.-C.; writing—original draft preparation, L.M.-P.; writing—review and editing, L.M.-P.; visualization, L.M.-P.; supervision, L.M.-P. and R.S.-C.; project administration, L.M.-P. and R.S.-C.; funding acquisition, L.M.-P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Ministry for Science and Innovation (Ministerio de Ciencia e Innovación, MCIN), grant number PID2020-119102RB-I00.

**Institutional Review Board Statement:** Not applicable.

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

**Data Availability Statement:** Not applicable.

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

## References


1. Madrid Regional Authorities. Acreditación en Inglés Avanzado 2022. 2021. Available online: https://www.comunidad.madrid/servicios/educacion/acreditacion-ingles-avanzado-2022 (accessed on 5 February 2022).
2. Coyle, D.; Hood, P.; Marsh, D. *CLIL: Content and Language Integrated Learning*; Cambridge University Press: Cambridge, UK, 2010. [CrossRef]
3. Madrid Regional Authorities. Datos y Cifras de la Educación 2021–2022. 2021. Available online: https://edicion.comunidad.madrid/sites/default/files/doc/educacion/sgea_datosycifras_2021-22_1.pdf (accessed on 5 February 2022).
4. Madrid Regional Authorities. Instrucciones de la Dirección General de Bilingüismo y Calidad de la Enseñanza Sobre la Organización de las Enseñanzas en Colegios Públicos e Institutos Bilingües Español-Inglés de la Comunidad de Madrid Para el Curso Escolar 2021–2022. 2021. Available online: https://www.educa2.madrid.org/documents/35607646/0/INSTRUCC_INIC_CURSO_DGBC_2021_2022.pdf/f2df39c4-10c2-d79e-0781-c13ba77902b4 (accessed on 5 February 2022).
5. Jiménez-Catalán, R.M. Vocabulary profiles in English as a foreign language at the end of Spanish primary and secondary education. *RLA Rev. Lingüística Teórica Apl.* **2016**, *54*, 37–50. [CrossRef]
6. Lahuerta, A. Analysis of accuracy in the writing of EFL students enrolled on CLIL and non-CLIL programmes: The impact of grade and gender. *Lang. Learn. J.* **2020**, *48*, 121–132. [CrossRef]
7. Ruiz de Zarobe, Y.; Sierra, J.M.; Gallardo del Puerto, F. *Content and Foreign Language Integrated Learning. Contributions to Multilingualism in European Contexts*; Peter Lang: Bern, Switzerland, 2011; Available online: https://www.peterlang.com/document/1051918 (accessed on 26 March 2022).
8. Van de Craen, P.; Lochtman, K.; Ceuleers, E.; Mondt, K.; Allain, L. An interdisciplinary approach to CLIL learning in primary schools in Brussels. In *Empirical Perspectives on CLIL Classroom Discourse*; Dalton-Puffer, C., Smit, U., Eds.; Peter Lang: Bern, Switzerland, 2007; pp. 253–274. Available online: https://www.peterlang.com/document/1053615 (accessed on 27 March 2022).
9. Arigita-García, A.; Sánchez-Cabrero, R.; Barrientos-Fernández, A.; Mañoso-Pacheco, L.; Pericacho-Gómez, F.J. Pre-eminence of determining factors in second language learning: An educator's perspective from Spain. *Heliyon* **2021**, *7*, e06282. [CrossRef] [PubMed]
10. Lasagabaster, D.; Sierra, J. Language attitudes in CLIL and traditional EFL classes. *Int. CLIL Res. J.* **2009**, *2*, 4–17. Available online: http://www.icrj.eu/index.php?vol=12&page=73 (accessed on 27 March 2022).
11. Roiha, A.; Sommier, M. Viewing CLIL through the eyes of former pupils: Insights into foreign language and intercultural attitudes. *Lang. Intercult. Commun.* **2018**, *18*, 631–647. [CrossRef]
12. San Isidro, X.; Lasagabaster, D. Students' and families' attitudes and motivations to language learning and CLIL: A longitudinal study. *Lang. Learn. J.* **2020**, *50*, 119–134. [CrossRef]
13. Coyle, D. Listening to learners: An investigation into 'successful learning' across CLIL contexts. *Int. J. Biling. Educ. Biling.* **2013**, *16*, 244–266. [CrossRef]
14. Moreno, M.C.; Carlos, U.R.J.; Paz-Albo, J. Bilingual cognitive and language development in the early years. *Enero-junio* **2020**, *35*, 151–161. [CrossRef]
15. Ferjan, N.; Kuhl, P.K. The brain science of bilingualism. *Young Child.* **2017**, *72*, 38–44. Available online: https://ilabs.uw.edu/sites/default/files/2017_FerjanRamirez_Kuhl_NAEYC.pdf (accessed on 2 April 2022).
16. Jafarigohar, M.; Divsar, H.; Etemad, P. The effect of CLIL context on the primary education EFL receptive and productive lexical growth. *Lang. Teach. Res.* **2022**. [CrossRef]
17. Lasagabaster, D. The effects of three bilingual education models on linguistic creativity. *Int. Rev. Appl. Linguistics Lang. Teach.* **2000**, *38*, 213–228. [CrossRef]
18. Otto, A.; Cortina-Pérez, B. *Content and Language Integrated Learning in Pre-Primary Education: Moving Towards Developmentally Appropriate Practices*; Springer: Berlin/Heidelberg, Germany, 2022.
19. Campillo, J.M.; Sánchez, R.; Miralles, P. Primary Teachers' Perceptions of CLIL Implementation in Spain. *Engl. Lang. Teach.* **2019**, *12*, 149–156. [CrossRef]
20. Alonso-Belmonte, I.; Fernández-Agüero, M. Teachers' narratives of resistance to Madrid's bilingual programme: An exploratory study in secondary education. *Linguist. Educ.* **2021**, *63*, 100925. [CrossRef]
21. Coonan, C.M. Insider views of the CLIL class through teacher self-observation–introspection. *Int. J. Biling. Educ. Biling.* **2007**, *10*, 625–646. [CrossRef]
22. EF EPI. *EF English Proficiency Index*; Education First: Westerville, OH, USA, 2021; Available online: https://www.ef.com.es/epi (accessed on 2 April 2022).
23. European Commission. *Education & Training: About Education Policies: About Education and Training Policies*; Publications Office of the European Union: Luxembourg, 2020; Available online: https://ec.europa.eu/education/policies/about-education-policies_en (accessed on 2 April 2022).
24. Barrientos-Fernández, A.; Sánchez-Cabrero, R.; Arigita-García, A.; Mañoso-Pacheco, L.; Pericacho-Gómez, F.J.; Novillo-López, M.Á. Measurement of different types of intelligence (general, verbal vs. non-verbal, multiple), academic performance and study habits of secondary students at a Music Integrated Centre. *Data Brief* **2019**, *25*, 104124. [CrossRef]
25. Murillo, F.J.; Almazán, A.; Martínez-Garrido, C. La elección de centro educativo en un sistema de cuasi-mercado escolar mediado por el programa de bilingüismo. *Rev. Complut. De Educ.* **2021**, *32*, 89–97. [CrossRef]

26. Ferrer, A.; Gortazar, L. Diversidad y libertad. Reducción la segregación escolar respetando la capacidad de elección de centro. *ESADE. EsadeEcPol—Center for Economic Policy. Save the Children.* 2021. Available online: https://www.esade.edu/ecpol/es/publicaciones/segregacion-escolar-esadeecpol (accessed on 2 April 2022).
27. Pena-Díaz, C.; Porto-Requejo, D. Teacher beliefs in a CLIL education project. *Porta Ling.* **2008**, *10*, 151–161. [CrossRef]
28. Senra-Silva, I. A study on CLIL secondary school teachers in Spain: Views, concerns and needs. *Complut. J. Engl. Stud.* **2021**, *29*, 49–68. [CrossRef]
29. Gierlinger, E.M. I Feel Traumatized: Teachers' beliefs on the roles of languages and learning in CLIL. In *Integrating Content and Language in Higher Education: Perspectives on Professional Practice*; Valcke, J., Wilkinson, R., Eds.; Peter Lang: Bern, Switzerland, 2016; pp. 97–116. Available online: https://www.peterlang.com/document/1050797 (accessed on 2 April 2022).
30. Hüttner, J.; Dalton-Puffer, C.; Smit, U. The power of beliefs: Lay theories and their influence on the implementation of CLIL programmes. *Int. J. Biling. Educ. Biling.* **2013**, *16*, 267–284. [CrossRef]
31. Viebrock, B. The Situation in the CLIL Classroom Is Quite Different—Or is it? Teachers' mindsets, Methodological Competences and Teaching Habits. In *Quality Interfaces: Examining Evidence and Exploring Solutions in CLIL*; Marsh, D., Meyer, O., Eds.; Eichstaett Academice Press: Eichstätt, Germany, 2012; pp. 79–90.
32. Wegner, A. Seeing the bigger picture: What students and teachers think about CLIL. *Int. CLIL Res. J.* **2012**, *1*, 29–35.
33. Jääskelä, P.; Häkkinen, P.; Rasku-Puttonen, H. Teacher beliefs regarding learning, pedagogy, and the use of technology in higher education. *J. Res. Technol. Educ.* **2017**, *49*, 198–211. [CrossRef]
34. Wood, G. Do religious beliefs have a place within an 'epistemically naturalized' cognitive system? *Sophia* **2017**, *56*, 539–556. [CrossRef]
35. Borg, S. The impact of in-service education on language teachers' beliefs. *System* **2011**, *39*, 370–380. [CrossRef]
36. Sánchez-Cabrero, R.; Estrada-Chichón, J.L.; Abad-Mancheño, A.; Mañoso-Pacheco, L. Models on Teaching Effectiveness in Current Scientific Literature. *Educ. Sci.* **2021**, *11*, 409. [CrossRef]
37. Skinnari, K.; Bovellan, E. CLIL teachers' beliefs about integration and about their professional roles: Perspectives from a European context. In *Conceptualising Integration in CLIL and Multilingual Education*; Nikula, T., Dafouz, E., Moore, P., Smit, U., Eds.; Multilingual Matters: Bristol, UK, 2016; pp. 145–274. [CrossRef]

education sciences

MDPI

*Article*

# Teachers' Perceptions and Appropriation of EFL Educational Reforms: Insights from Generalist Teachers Teaching English in Mexican Rural Schools

**Miguel Hernández Hernández * and Jesús Izquierdo ***

División Académica de Educación y Artes, Universidad Juárez Autónoma de Tabasco Villahermosa, Villahermosa 86040, Mexico

* Correspondence: miguel.hernandezh@ujat.mx (M.H.H.); jesus.izquierdo@mail.mcgill.ca (J.I.)

**Abstract:** Educational reforms and educational policy changes have favored the learning of English as a foreign language (EFL) in public education. Empirical research has examined how EFL specialist teachers in urban public schools perceive these changes or the extent to which they adopt a new curriculum. Nonetheless, the new EFL policies have also had an impact on rural schools where generalist teachers are forced to teach English along with other areas of the curriculum. In this context, little research has explored teachers' perceptions and appropriation of ongoing curricular changes. The present study explored this issue among generalist rural secondary school teachers in the southeast of Mexico. To this end, an explanatory sequential mixed method was adopted with a sample of 216 generalist teachers. During the quantitative phase, the participants completed two Likert scale questionnaires. Then, a semi-structured interview was conducted with a subsample of participants who obtained high ($n$ = 7) or low ($n$ = 7) results in the perceptions and appropriation questionnaires. The statistical analyses showed a weak but positive correlation between perceptions and appropriation. The qualitative data provide some insights that explain the weakness of the correlation.

**Keywords:** English Language Teaching (ELT); curricular reforms; teachers' perception; curricular appropriation; generalist teachers; rural schools

**Citation:** Hernández Hernández, M.; Izquierdo, J. Teachers' Perceptions and Appropriation of EFL Educational Reforms: Insights from Generalist Teachers Teaching English in Mexican Rural Schools. *Educ. Sci.* **2023**, *13*, 482. https://doi.org/10.3390/educsci13050482

Academic Editor: Martin Howard

Received: 4 April 2023
Revised: 5 May 2023
Accepted: 8 May 2023
Published: 10 May 2023



## 1. Introduction

Many countries around the world have experienced significant changes due to the impact of supranational financial, environmental, and sociopolitical challenges. These challenges, alongside the imperialism of English as a global and useful international language, have led many countries to undergo major educational reforms that sanction the learning of English as a second/foreign language (EFL) in public education [1,2]. In these reforms, curricular changes constitute the main axis of educational development and, in the case of EFL education, instantiate the renewal of the day-to-day teaching and learning practices [3].

Through educational reforms, public educational systems require teachers to adopt instructional models that aim to help learners develop particular and general EFL competencies [4]. However, the implementation of reforms is not straightforward. Their success greatly depends on teachers' willingness to accept, adapt, and implement a curricular change; nonetheless, these actions demand a critical reorganization of well-established teaching habits [4]. Some authors affirm that the critical reorganization of teaching habits is partly influenced by the positive or negative perceptions that teachers hold with respect to the educational reforms [1,5]. Moreover, for educational changes to occur, teachers need to appropriate the educational practices outlined in the new curriculum. This effect implies that, vis-à-vis the educational reforms, teachers need to become educational policy enactment agents who perceive and appropriate curricular changes [6].

The constructs of teachers' perception and appropriation of EFL curricular changes have received attention in previous second language research [5]. Nonetheless, they have been separately examined by a handful of studies; often, these studies have been conducted with EFL specialist teachers who deliver language instruction across different levels of public education in urban areas [7]. It should be noted though that the educational policies have not only sanctioned the learning of English in urban areas. They have also made the learning of English obligatory in rural settings where there is a lack of specialists in English language teaching or other areas of the curriculum [2]. Therefore, in rural schools, one generalist teacher is compelled to teach all areas of the public school curriculum, including English [8–11], to the same group throughout the school day.

In the context of the current study, the national curriculum of the public education of Mexico states that English needs to be taught in the three grades of secondary education. Vis-à-vis this policy, in urban schools, EFL education is often delivered by language specialists. In rural areas, however, generalist teachers are obligated to teach English in addition to the other areas of the secondary school curriculum. The professional profile of this type of educator often includes undergraduate studies in general education or pedagogy. In their schools, they deliver EFL instruction without formal language teacher education and minimal language competencies [12,13]. In addition to their teaching duties, generalist teachers mentor learners, manage their school, and plan school logistics [14,15].

Based on the aforementioned issues, the objective of the present study was to explore the perceptions of generalist teachers about current EFL curricular changes and the appropriation of the recommended teaching practices. To this end, a sequential explanatory mixed-methods study was carried out; it addresses three research questions: How do generalist teachers perceive curricular changes for the teaching of English in rural education? What is the level of appropriation of curricular changes in their English language teaching practice? What are the factors that influence their perception and appropriation of curricular changes for English language teaching?

## 2. Background to the Study

In this section, the central constructs of the perception and appropriation of EFL curricular reforms are first presented. Then, the section comes to an end by discussing the need to explore the interplay between these two constructs among generalist teachers who deliver EFL education in public rural schools.

### *2.1. Teachers' Perception of EFL Curricular Reforms*

From a psychological viewpoint, the construct of perception refers to understanding how a context is perceived. In the field of education, perceptions are conceived psychological notions that encompass, for instance, thoughts, emotions, behavior, beliefs, and cognition [16–18]. In second language research, teachers' perception is conceptualized as a cognitive process that is based on what a teacher feels, creates, thinks, and understands and how a teacher behaves with respect to a particular aspect of language education—in this case, a curricular change [5,16,19]. This cognitive process is nurtured by a set of external and internal factors [18]. While the former includes school policies and the teachers' social, cultural, and professional background [19], the latter spans ideologies, knowledge, and attitudes [2,5,19]. Teacher's perceptions about curricular changes develop from internal ideologies that emerge from observation, knowledge, and discernment of (new) teaching approaches and curricular guidelines [20]. The construct of teachers' perception builds upon a wide array of factors, but theoretical and empirical researchers agree upon the dimensions of opinions, beliefs, behaviors, and emotions. During curricular changes, these dimensions allow teachers to interpret educational policies and teaching demands [5,21,22].

Opinions constitute a set of subjective interpretations that teachers develop in a specific context or toward an issue [23]. In language teaching, opinions are considered positive or negative constructions that teachers build on, taking a stance in favor of or against the English teaching philosophy [5,24], curricular guidelines, teaching strategies, content,

or other elements of the language educational reforms [5]. Beliefs are assumptions that emerge from knowledge [25]. In second language teaching, beliefs [6,19] are part of teachers' cognition [5] and encompass multiple aspects of language education, such as pedagogical processes, learning processes, and evaluation processes [18,25]. Teachers prioritize beliefs of language teaching and learning based on their educational environment and their professional background experiences [5]. Behaviors are complex processes, as they imply that cognition transfers into actions [5]. In turn, these actions are known as pedagogical practices that teachers consider relevant for the teaching and learning process [26]. The behaviors that teachers display could be connected to all aspects of the curricular changes, for example, language policies, guidelines, content, teaching practices' impact on students, and other aspects of the curriculum that are tightly connected to the classroom [5,27]. Emotions rely on different conceptual psychosocial or psychoeducational conditions. [17]. In language teaching education, teachers experience positive or negative emotions, for example, excitement, joy, motivation, dissatisfaction, fear, and burnout, as a result of their practice or curricular reforms [18,28].

### *2.2. Teachers' Appropriation of EFL Teaching Practices*

The construct of appropriation refers to teachers' adoption of curricular changes [29]. Appropriation is achieved when teachers understand the new curricular demands and reorient their practice accordingly [3]. Depending on the level of adoption, five levels of teachers' appropriation of curricular changes can be identified [30]: (1) absence; (2) superficial; (3) surface; (4) conceptual underpinning; and (5) engagement. Two relevant aspects of the appropriation encompass regular teaching practices and evaluation processes [30].

The teaching practices are connected to the development of a lesson; thus, this teaching process is supported by the bases of curricular, disciplinary, or pedagogical knowledge that teachers need for the implementation of learning activities in the classroom [29]. However, a teaching practice could be limited when teachers are non-specialists or lack teacher training [31]. Teaching practices are framed by experiences on learning and teaching and are complemented by teachers' professional development [32]. In the context of curricular reforms, a teaching practice is connected to teaching activities that are built upon curricular guidelines. As for the evaluation dimension, it is a complex aspect of appropriation. It goes beyond the adoption of measurement and testing, which implies systematic measures of a specific competence. Instead, evaluation encompasses judgments of the learning progress and the achievement of the learners within the curriculum [2,26]. Moreover, it provides teachers with valuable information about the effectiveness of their teaching practice, their commitment to curricular demands [33], the appropriateness of the material, and the role of the learning context, among other elements [2]. In English Second Language (ESL) or EFL, teachers often conduct evaluations using a language framework and guidelines about the teaching, learning, and evaluation processes in their curriculum [34].

### *2.3. The Interphase of Perception and Appropriation*

In the field of second language education, perception is associated with behaviors, beliefs, thoughts, and emotions tightly connected with the process of learning or teaching a new language [5,18,35]. Appropriation, however, encompasses the adoption, adaptation, and interpretation of new curricular changes [30,36]. Some studies have examined these two constructs separately [35], while other studies have explored a possible interphase among them.

In this regard, some studies from Libya, Iran, Argentina, and other countries have examined how language teachers perceive and react to curricular changes [37,38]. These studies have examined teachers' perceptions of curricular changes using a qualitative approach through interviews, written reflections, and observational data. Their findings show a negative perception of curricular changes. In the same way, other qualitative studies in China, Vietnam, and Libya have focused on the construct of appropriation and analyzed how language teachers implemented or appropriated the language curricular

reforms to innovate in their practices [39–41]. The qualitative interview and observation data from these studies indicate that the participants showed difficulties in adopting the curricular changes.

While the aforementioned studies have explored these constructs separately, empirical evidence has revealed a harmonious interphase [41,42], which supports teachers' understanding of the curricular changes and reorientation of their pedagogical practices [3,8,22]. In Hong Kong, for instance, using a mixed methods approach, a pioneer study [42] analyzed how English teachers from public schools appropriated the curriculum. The interview and observational data showed that teachers implement the curriculum based on their training and experiences due to a constant interaction with stakeholders. In turn, the quantitative results from an attitude Likert scale revealed a positive tendency to the challenge of language curricular innovation [42]. In Taiwan, a research study [7] collected quantitative and qualitative data to obtain a holistic appreciation of the appropriation of curricular changes. The quantitative findings revealed that teachers were aware of and showed positive perceptions. However, the qualitative data revealed a lack of appropriation of the language curricular guidelines [7]. This finding is in line with those from a qualitative study in Argentina, where Banegas (2016) found a remarkable incongruence between teachers' practices and curricular guidelines [38].

The above evidence provides valuable insights into the perception and appropriation of curricular changes among EFL teacher specialists who work in urban public schools. Nonetheless, there is a growing interest in exploring the perceptions and appropriation of language curricular changes among teachers who are neither language teachers nor specialists in teaching English [2]. This interest emerges from the fact that, in a rural context, there is often a lack of language specialists to deliver second language instruction [2,15]. Thus, non-specialized teachers are obligated to teach English using the educational policies established by the reforms [15]. These teachers are often generalist teachers [2] or educators whose teacher education focuses on the teaching of subject matter from different areas of the public curriculum (i.e., first language literacy, geography, mathematics, etc.) [2]. In their rural schools, they often need to deal with overcrowded school groups, a lack of teaching resources, and the teaching of multiple school grades in the same classroom. Their students have agricultural responsibilities that limit their time to study and cause absenteeism [12]. With respect to the teaching of English, generalist teachers often have very low language proficiency levels and have not attended language teacher training. This fact leads them to implement EFL tasks based on their own language learning experience [2].

## 3. Materials and Methods

### *3.1. Research Design*

Based on the aforementioned issues, the present study was conducted using an explanatory sequential mixed methods design [43] to answer three research questions:

1. How do generalist teachers perceive curricular changes for the teaching of English in public rural education?
2. What is the level of appropriation of the ELT curricular guidelines?
3. What are the factors that contribute to their perception and appropriation of the ELT curricular guidelines?

The first and second research questions were answered during a quantitative phase using a descriptive design [44]. In this phase, three instruments were administered to generalist secondary school teachers using non-probabilistic sampling: a survey and two Likert scale questionnaires. The first Likert scale questionnaire was created considering four dimensions of teachers' perception: opinions, beliefs, pedagogical practices, and emotions. The second Likert scale questionnaire was created considering two dimensions of appropriation: teaching practice and evaluation. All of these dimensions were conceptualized and operationalized based on the literature reviewed in the previous section. Furthermore, the construct, content, and ecological validity and reliability of the questionnaires were verified. In turn, the quantitative data allowed us to test the following hypothesis:

**H1.** *The perceptions of generalist school teachers, about ongoing curricular changes in English language learning and teaching, can have an impact on the appropriation of the ELT practices that are sanctioned by public education reforms.*

The third research question was answered during a qualitative phase, using a multiple-case design. During this phase, two subsamples of participants were selected, using their responses from the quantitative instruments [45]. The selected participants [46] completed a semi-structured interview [47], where they elaborated upon their answers in the quantitative instruments. The interview data helped us explore the following research assumption:

*Different factors underpin the relationship between teachers' perceptions and appropriation of curricular changes for English language teaching.*

### *3.2. Context and Participants*

In Mexico, secondary education is mandatory and can be completed at public schools in urban or rural areas [15]. The urban and rural schools follow the same national curriculum and EFL teaching guidelines [15]. All students need to complete three hours of EFL instruction per week [2,15]. Currently, the three grades of secondary education work with the 2017 curriculum [46]. This curriculum has undergone different reforms (i.e., 2017, 2011, 2006, 1993) over the last 30 years [15,48]. In terms of EFL learning, the reforms aim to favor a transition from a structure-based to communicative approach. According to the curriculum, English should be seen not as the object of study, but as a means of communication [2,15]. The EFL teaching guidelines are based on the Common European Framework of Language Reference and the National English Program [48]. While the language content and curricular guidelines are the same across all school types [15], the implementation of EFL teaching is different in urban and rural schools.

In urban areas, public secondary education is usually delivered at general and technical schools. In rural areas, secondary education is mostly delivered at telesecondary schools. In general and technical secondary schools, each area of the curriculum is taught by a content specialist. In the case of English, the lessons are delivered by language teachers who move from classroom to classroom across the three grades of secondary education. In the telesecondary schools, however, the teaching conditions are completely different. In these schools, a generalist teacher works with the same group throughout the school year and teaches all of the curricular areas: first-language literacy, arts, history, sciences, mathematics, and English [2]. To deliver the EFL lessons, the generalist teachers should project a 15-min video-recorded EFL lesson, taught by an EFL teacher. The National Ministry of Education broadcasts the EFL lessons nationwide through satellite TV or the Internet [15]. Then, the generalist teacher needs to build a 45-minute lesson based on the video recording [24,48]. In order to implement this lesson, the generalist teacher needs to follow up on the TV program content, adhere to the EFL teaching guidelines in the curriculum, create the necessary learning tasks, provide feedback, and evaluate students' learning [48]. In the absence of language training, these teachers implement individual initiatives to meet the EFL curricular demands, reduce the level of learner attrition, and compensate for their language deficiencies through the use of web resources [2].

### *3.3. Participants*

The present study was conducted in the southeast of Mexico, where the majority of learners complete secondary education in the rural areas of the state of Tabasco. In this state, telesecondary schools have an approximate population of 50,715 learners. This population is served by approximately 2262 generalist teachers, 60 percent of whom are female teachers and 40 percent of whom are male teachers. The teachers are distributed in 459 telesecondary schools across 17 municipalities, which are clustered in 5 geopolitical regions.

### *3.4. Sample*

Due to substantial differences in teacher profiles, teaching conditions, and school organization in rural areas, the participant selection was based on sampling criteria that have been used in previous research for the homogenization of schools' and teachers' conditions in this context [6]. The first condition was geographical proximity. Thus, the participating teachers were located in the central geopolitical region of the state and their schools were close to rural villages that could be reached by car or boat. In terms of teaching conditions, the participants' schools needed a minimum of two groups per grade. Based on these two conditions, 29 telesecondary schools were selected. Finally, to homogenize the knowledge of the school curriculum among the participants, the generalist teachers needed at least five years of experience in the telesecondary school system, to have had more than one year of experience at their current school, and were currently teaching one school grade only. Furthermore, the participants were excluded if they held administrative appointments (e.g., school district directors, school principals, school coordinators, etc.).

To gain access to the selected schools and participants, a consent letter was summitted to the telesecondary school department of the Ministry of Education of the state. The department answered with an official acceptance letter, granting access to the selected schools. In each school, a meeting was held with the school administrators and teachers to invite them to participate in the research. In this meeting, the objective, implementation of instruments, and ethical principles were presented. In turn, only 216 generalist teachers agreed to participate in the research and signed a consent letter. The survey data showed a balanced distribution of the participants between first (34%), second (35%), and third school grades (31%). These participants teach students whose age ranges between 12 and 15 years old [42], and English language proficiency is quite low.

In terms of professional background, the majority of the participants held an undergraduate degree in education (see Figure 1). Moreover, the survey data revealed that some participants (37%) had a master's degree, and a few of them (3%) held a PhD in their discipline. Likewise, the survey data showed that 8% of the generalist teachers had been abroad for pleasure, and only a few of them had done so for academic purposes.

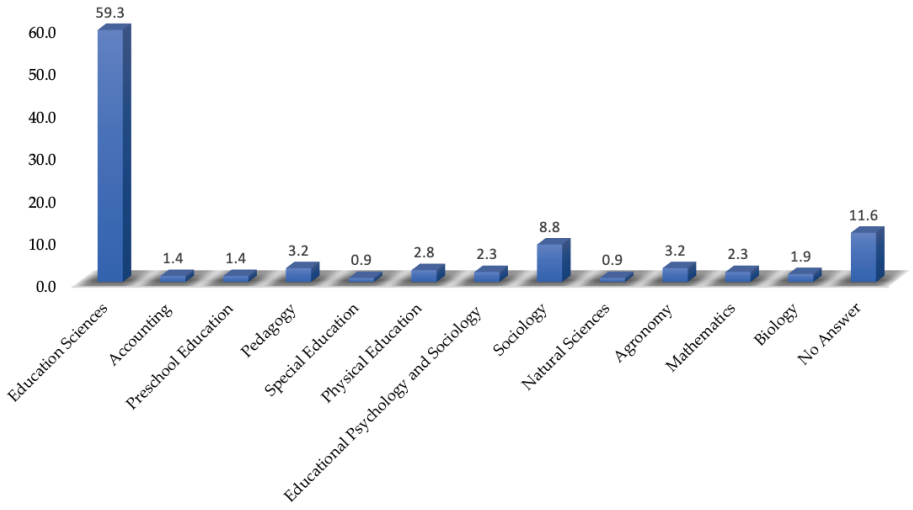



**Figure 1.** Bachelor's degrees of the generalist teacher population.

For the qualitative component, two subsamples of teachers were selected using z-scores [49–51]. These subsamples represented extreme-case [52] teachers who showed a positive or negative pattern of answers in both the perception and appropriation Likert scale questionnaires [51]. The participants in the first subsample held at least one z-score above the group mean [53] in terms of perceptions and appropriation. This meant that their perception and appropriation scores were statistically higher in comparison to their peers. The participants in the second subsample held at least one z-score below the group mean in terms of perceptions and appropriation. This meant that their perception and appropriation

scores were lower in comparison to those of other teachers. Out of the 216 teachers who voluntarily completed the quantitative instruments, the z-score procedure allowed us to select 14 teachers that represented extreme cases [52] and consented to participate in an interview. It should be noted though that a larger subsample was considered for the interview. However, most of them clearly stated their objection to be interviewed.

Based on these selection criteria, the participants in Tables 1 and 2 were interviewed. They were from 13 schools. Their schools were located in remote rural areas of five municipalities.

**Table 1.** Participant subsample with high perception and appropriation scores.

| Alias | Age | Perception z-Score | Appropriation z-Score |
|---|---|---|---|
| Samuel | 47 | 1.3 | 1.1 |
| Grecia | 43 | 1.5 | 1.1 |
| Alex | 50 | 1.6 | 1.5 |
| Jonas | 50 | 1.8 | 1.5 |
| Tana | 41 | 2.2 | 1.8 |
| German | 59 | 2.4 | 1.0 |
| Addy | 33 | 2.1 | 1.2 |

**Table 2.** Participant subsample with low perception and appropriation scores.

| Alias | Age | Perception z-Score | Appropriation z-Score |
|---|---|---|---|
| Willy | 55 | −1.6 | −1.9 |
| Carme | 35 | −1.6 | −1.5 |
| Cando | 46 | −1.6 | −1.2 |
| Nicole | 41 | −1.3 | −1.4 |
| Natalia | 35 | −1.3 | −1.3 |
| Maya | 45 | −1.2 | −1.5 |
| Liz | 48 | −2.5 | −1.3 |

Samuel had worked for 23 years in the telesecondary system. He had a BA in education. He reported that he had not taken any ELT training and had never received information about the national English program. He indicated adapting his own EFL teaching techniques and material according to the students' level.

Grecia indicated having worked in the system for 17 years. She had a BA in education, a specialization in teaching Spanish, and an MA in education. She reported participating in conferences to learn about the 2017 curriculum but implemented the 2011 curriculum instead. She indicated that she was struggling with teaching English.

Alex had worked in the telesecondary system for 23 years. He had a BA in sociology. He reported having difficulties with English and not receiving training to teach English in telesecondary schools. He said he was using the 2016 English syllabus and books of the 2006 curriculum due to a lack of knowledge of the 2017 curriculum.

Jonas had worked for 18 years in the telesecondary system. He had a BA in social sciences. He indicated receiving information and participating in a conference to learn about the 2017 curriculum. Moreover, he had received information about EFL learning and teaching from the Ministry of Education. He indicated having difficulty teaching English; therefore, he had taken private English lessons to prepare his EFL lessons. He reported working with the 2006 curriculum and materials and using dictionaries.

Tana, who reported having a BA in education, had worked in the system for 16 years. She affirmed using the 2006 and 2011 curricula and books. She had travelled abroad, specifically to the United States. Nonetheless, she reported having difficulty when teaching English due to lack of EFL teacher training; therefore, she used online translators during her classes.

German had a BA in pedagogy. He had worked in the telesecondary system for 27 years and had information about the 2017 curriculum but had not participated in

conferences. He reported having difficulty when teaching English. He indicated using the 2011 curriculum and books.

Addy had a BA in education. She had worked in the telesecondary system for 13 years. She reported not receiving materials about the 2017 curriculum. Moreover, she had not received information or training to teach English in telesecondary schools. Similar to other teachers, she indicated having difficulty teaching English. She reported using the 2006 curriculum, a dictionary, and a translator to teach English.

The second subsample of teachers had lower perception and appropriation scores, as their questionnaire data placed them at least one z-score below the group mean. Table 2 shows their pseudonyms and personal and professional data. This table indicates that this subsample also included seven participants, whose age varied from 35 to 55 years old. Moreover, their teaching experience in telesecondary schools ranged from 5 to 27 years.

Willy had 27 years of experience in the telesecondary system. He had two undergraduate degrees: one in social sciences and one in pedagogy. Moreover, he indicated having little information about the 2017 curriculum and did not have information about the national English program. Thus, he reported implementing the 2011 and 2017 curricula. He indicated having difficulties when teaching English.

Carmen, with a BA and PhD in education, had worked in the telesecondary system for 9 years. He indicated not having information or participating in a conference about the 2017 curriculum. Moreover, he reported having neither knowledge of the national English program nor having had support for teaching English in telesecondary schools; he used translators to teach English due to language proficiency issues.

Cando held two undergraduate degrees, one in physical education and one in telesecondary education. He had worked in the telesecondary system for 22 years. He reported a lack of information about the 2017 curriculum and had not attended ELT workshops. Thus, he indicated using the 2011 curriculum. He reported taking private English lessons to prepare for his English classes.

Nicole had a BA in social sciences. She had worked in rural secondary schools for 22 years, and she had little information about the 2017 curriculum; thus, she had organized a study group to learn about the new curriculum. She had little information about EFL teaching courses offered by the Ministry of Education. She affirmed having low proficiency in English. Therefore, she had paid for private English lessons. In turn, she reported working with the 2006 curriculum and books.

Natalia had one BA in education and one in elementary education. She had 5 years of experience in the telesecondary system. She had taken courses on the 2017 curriculum. Nonetheless, she specified that she was using the books that were based on the 2006 curriculum. She indicated not having had support from the Ministry of Education for teaching English. Moreover, she reported having difficulties teaching English.

Maya knew neither the 2017 curriculum nor the national English program; thus, she was using the 2011 curriculum. She had worked for 20 years in the telesecondary system. She had a BA in education. She had difficulties teaching English and used a translator during her English classes.

Liz had worked in telesecondary schools for 19 years. She had a BA and MA in education. She reported using the 2006 curriculum because she had received neither materials of the 2017 curriculum nor training about the national English program. Similar to other teachers, she had paid for private English lessons.

### *3.5. Quantitative Instruments*

In the quantitative phase, a survey, a Likert scale questionnaire for perceptions and a Likert scale questionnaire for appropriation were administered. These instruments emerged from the literature review and were constructed under the principles of the classical testing theory [53–55]. Moreover, the design of the perception and the appropriation Likert scale questionnaires considered different methodological criteria: the nature of the scale, the unidimensionality, the univocity of items, [53], and the semantic direction (positive or

negative) of the items. Likewise, content, construct, and ecological validity were checked to verify the central construct as well as the dimensions and their pertinence [50]. Further, the consistency and accuracy of the collected data were verified through reliability procedures.

#### 3.5.1. Survey Design

The purpose of the survey was to explore the generalist teachers' knowledge of language curricular changes. This instrument considered six categories or sections that included close-ended, open-ended, and multiple-choice questions [44,54,55]. The participants' answers were analyzed using frequency counts [50,56]. In total, the survey included 31 items that were grouped into six sections. Section 1 elicited sociodemographic, academic, and experience abroad information using seven items. Section 2 elicited information about their knowledge of the public education system employing three items. Section 3 elicited information about professional development and knowledge of the 2017 curricular changes using four items. Section 4 focused on their ELT professional development considering five items. Section 5 elicited information about the implementation of ELT lessons and their EFL proficiency using nine items. Section 6 elicited information about the implementation of the ELT curriculum in their teaching using six items.

#### 3.5.2. Design of the Likert Scale Questionnaires

The second instrument consisted of an available published and validated Likert scale questionnaire that explores generalist teachers' perceptions about EFL curricular changes [18]. This instrument aimed at identifying the generalist teachers' position about their knowledge, belief, thought, behavior, and emotions about curricular changes [50,55]. It included four dimensions and a four-degree agreement scale [51,55] to identify the participant position (negative or positive) for each dimension of the constructs of interest, which were explained in the background to the study section. The first dimension, opinions, tapped into their thoughts about curricular changes using seven items. The second section, beliefs, focused on what generalist teachers believed about the language curricular changes using five items. The third dimension, pedagogical practices, elicited information about behaviors that align with the language curriculum through nine items. The fourth dimension, emotion, included seven items about what teachers feel about the ELT curricular changes. All items relied on the methodological criteria for the use of ordinal variables and differential gradients of opinion [56–58]. Hence, a 4-degree agreement scale was implemented [56,58]. Each of the scale's options was assigned a numerical value: strongly agree = 4 points, agree = 3 points, disagree = 2 points, and strongly disagree = 1 point [24,58]. The instrument did not include a neutral point [58] to avoid confusing participants [56,58,59]. The absence of a neutral point pushed the respondents to indicate a positive or negative judgement [24,54] about the ELT curricular changes.

The third quantitative instrument identified the position of participants in regard to the two dimensions of the appropriation of curricular guidelines through items using ordinal scales [59]. Through the items, the occurrence of the recommended teaching practice and an evaluation of learning were verified. All items of the appropriation Likert scale questionnaire were based on the literature of appropriation included in the background to the study section. This Likert scale appropriation questionnaire, considered two dimensions—teaching practice using 18 items and the evaluation of learning—using 16 items. These dimensions were conceptualized and operationalized using the definitions presented in the literature review. This instrument registered the frequencies of factual information related to both dimensions through a differential gradient scale of frequency. Each gradient was assigned a numerical value: very often = 4 points, often = 3 points, rarely = 2 points, and never = 1 point [50,50–60]. Based on this characteristic, the instrument was treated with an additive property [24,59].

### 3.6. Qualitative Instrument

For the qualitative phase, a semi-structured interview [61] was designed to expand the participants' answers obtained through the perception and appropriation Likert questionnaires. In turn, a trustworthiness verification process was considered using the epistemological principles of qualitative research [52]. The interview included a series of open-ended questions for each dimension of the Likert scale questionnaires. For the exploration of perception, the interview considered three questions for the dimension of opinions. For the dimension of beliefs, four questions were prepared. Regarding the dimension of pedagogical practices, three questions were considered. Then, for the emotion dimension, two questions were asked. Likewise, to follow up on the participants' answers in the appropriation questionnaire, the interview included six questions for the teaching practice dimension. Finally, for the evaluation of the learning dimension, three questions were asked. During the interviews, in addition to the pre-set group of prompts, personalized questions were asked. These questions allowed for a deeper exploration of the constructs of interest for each participant. Upon completion of the interview, the researchers thanked the participants for their contribution to the present study.

### 3.7. Quantitative Instruments' Validity and Reliability Procedures

For the quantitative instruments, content, construct, and ecological validity were considered [50,55]. As for reliability, the stability and internal consistency of the Likert scale questionnaires were independently tested for each dimension.

#### 3.7.1. Validity Procedures

The survey was subject to ecological validity only. To this end, the survey was administered to a group of generalist teachers who had knowledge about the language curricular changes. These participants commented on the appropriateness of the questions and suggested changes that were incorporated in the final version of the survey.

In the Likert scale questionnaires of perception and appropriation, content and construct validity were checked [24]. For content validity [50], the items of each dimension were created considering the literature reviewed. Following this procedure, four groups of items were developed for the perception questionnaire and two groups of items for the appropriation questionnaire. Then, with the use of expert judgement procedures, each group of items was analyzed by a group of researchers in the field of second language teaching. These judges were provided with the definitions of the two constructs and their dimensions. Then, they made comments on the congruence between the items and the scale, the congruence among the items, and the pertinence of the items within each dimension. With this procedure, the unidimensionality of the items in all dimensions was verified [51,52]. Likewise, ecological validity was checked using a subsample of eight generalist teachers who assessed the relevance of the items.

During the design of a Likert scale questionnaire, construct validity is fundamental because it allows the researchers to identify the dimensions of the construct under observation. When the items are not created with a set of dimensions of a construct in mind, the items are subject to exploratory factor analyses. These analyses group the items into factors that can be interpreted by the researchers [51]. However, when the researchers have a clear and warranted theoretical basis for each of the dimensions of interest, construct validity can be verified using other procedures. First, the groups of items are subject to content validity. Once the judges validate the conceptual interdependence of the items and their adherence to the dimensions of the construct, the correlation coefficient between the group of items is verified through Cronbach alpha analyses [51]. These analyses are run separately for each of the group of items. Then, the relationship between the dimensions of the construct is checked through convergent validity [50]. In order to test convergent validity, correlation analyses are performed between the dimensions of the construct. In the current study, since the items for each dimension were created independently, construct validity was not verified through exploratory factor analyses. Instead, it was verified

through content validity, Cronbach alpha analyses for each group of items, and correlation analyses between the dimensions of each construct (see [62], for the use of these procedures for the verification of construct validity).

#### 3.7.2. Reliability Procedures

To check reliability, two parallel forms [12] of the perception and appropriation questionnaires were simultaneously administered to two independent groups of generalist teachers. Both forms included the same items, but in reverse order. Then, the statistical software SPSS v.25 was used to run Mann–Whitney tests to identify differences between the results that were obtained from the test forms, item by item. The differences were analyzed using an alpha of 0.05. A $p$ greater than 0.05 assured that both versions collected similar data [51].

The stability and internal consistency of the Likert scale questionnaires were independently tested for each dimension. To this end, a Cronbach's alpha greater than 0.80 would confirm the internal consistency of each dimension. Then, the corrected correlation coefficient of each item was checked. Only items with a coefficient greater than 0.3 were retained. Finally, the interdependence between the questionnaire dimensions was explored using a Spearman correlation test with an alpha of 0.05 [51]. Then, the correlation strength was identified [63] as weak ($0.20 < r_s < 0.39$), moderate ($0.40 < r_s < 0.59$), or strong ( $r_s > 0.60$).

#### 3.7.3. Validity and Reliability Results: Perception Questionnaire

For the perception questionnaire, the Mann–Whitney analyses showed that the answers to the items of the dimensions of opinion and beliefs were similar across the questionnaire versions, as all items obtained a $p > 0.05$. Therefore, all items in dimensions 1 and 2 were retained for the analysis.

For dimension 3, pedagogical practice, item 1 yielded a significant difference between questionnaire versions. This implied that this item would elicit different answers depending on the version; therefore, it was excluded from future analyses. For dimension 4, emotion, a significant difference between test versions was obtained for items 4, 5, and 7; thus, they were excluded.

The internal consistency analyses yielded a favorable Cronbach's alpha value for dimensions 1 ($\alpha = 0.926$), 2 ($\alpha = 0.806$), and 3 ($\alpha = 0.747$). Moreover, the correlation coefficient analyses yielded a value higher than 0.3 for the items that were retained for dimensions 1, 2, and 3. Nonetheless, dimension 4, emotions, yielded an internal consistency alpha value of 0.696; therefore, this dimension was excluded from future analyses. Table 3 summarizes the validity and reliability results, presenting the dimensions and number of items of the perception questionnaire that were retained for this study.

**Table 3.** Items retained for the perception questionnaire after the validity and reliability tests.

| Dimension | Initial Number of Items | Internal Consistency $\alpha$ = | Correlation Coefficient $p$-Value | Final Number of Items |
|---|---|---|---|---|
| Opinions | 7 | 0.926, | >0.3 | 7 |
| Beliefs | 5 | 0.806, | >0.3 | 5 |
| Pedagogical practices | 9 | 0.747 | >0.3 | 8 |
| Emotions | 7 | 0.696 | <0 | 0 |

In order to test convergent validity, correlation analyses were run on the dimensions of the perception questionnaire using the Spearman test, as the data were not normally distributed. The correlation between the retained dimensions of the perception questionnaire yielded a weak but significant positive correlation between dimension 1, opinion, and dimension 2, beliefs ($p = 0.014$; $r_s = 0.165$), and between dimension 2, beliefs, and dimension 3, pedagogical practice ($p < 0.001$; $r_s = 0.255$). Nonetheless, the Spearman test

showed a lack of significant correlation between dimension 1, opinion, and pedagogical practice ($p$ = 0.298).

3.7.4. Validity and Reliability Results: Appropriation Questionnaire

Regarding the appropriation questionnaire, the Mann–Whitney analyses of dimension 1, teaching practices, indicated that items 1, 6, and 9 elicited different answer patterns between versions. Therefore, these items were excluded from future analyses. In dimension 2, evaluation, the nonparametric analysis results showed that all items had a $p$-value higher than 0.05. Consequently, all items were retained.

During the reliability analyses, the internal consistency of dimension 1, teaching practices, achieved a Cronbach's alpha value of $\alpha$ = 0.904, and dimension 2, evaluation, achieved a Cronbach's alpha value of $\alpha$ = 0.907. Moreover, the coefficient correlation analysis of all the items for dimensions 1 and yielded a value higher than 0.3. In light of these results, all items were retained for future analyses. Table 4 summarizes the validity and reliability results, presenting the dimensions and number of items that were retained for the study.

**Table 4.** Items retained for the appropriation questionnaire after the validity and reliability tests.

| Dimension | Initial Number of Items | Internal Consistency $\alpha$ = | Correlation Coefficient $p$-Values | Final Number of Items (after Validity and Reliability Procedures) |
|---|---|---|---|---|
| Teaching practices | 18 | 0.904 | >0.3 | 15 |
| Evaluation of learning | 16 | 0.907 | >0.3 | 16 |

*3.8. Qualitative Instrument Validation Procedures*

Regarding the validity and reliability of the qualitative instrument, the interviews were transcribed using word processing software. Moreover, the transcripts were verified to assure the verbatim transcription of 14 interviews. Then, all transcripts were entered into Atlas.ti version 8.4.5. A cross-case thematic analysis [61] was implemented with the 14 informants' transcripts to identify emerging themes [52]. First, common themes were identified using two analysis cycles [61]. The first cycle focused on the identification of themes connected to the research questions. The second cycle focused on reorganizing and reducing categories and subcategories. This process analyzed excerpts about the perceptions of curricular changes and the appropriation of new ELT pedagogical practices.

## 4. Results

*4.1. Quantitative Results: Perception Questionnaire*

In order to identify the type of perceptions that teachers hold about the curricular reforms, frequency analyses were run on the items in each of the three remaining dimensions of the questionnaire. In these analyses, every item was treated independently to identify positive or negative perceptions of participants. Table 5 shows the distribution of participants across the possible Likert scale choices.

For dimension 1, opinions on the new curriculum, the teachers tended to take a negative position. For example, in item 1, the majority of the generalist teachers indicated that the pedagogical principles in the 2017 educational curriculum for English language teaching did not favor learning effectively in secondary education. Further, in item 2, a high number of teachers considered that the suggested strategies for teaching the English language in the 2017 educational model were not suitable for students in telesecondary schools. In item 4, more than a half of the participants agreed that the suggested strategies for teaching English in the 2017 educational curriculum could not be easily adapted by teachers in telesecondary schools. In item 5, the majority of the participants indicated that the amount of thematic content in English hardly allowed generalist teachers to properly cover the curricular content with their groups. However, in item 7, a large number of

participants considered that the teaching strategies they used help them to teach English as required by the 2017 educational model (see Table 5).

**Table 5.** Distribution of answers in dimension 1, opinions about the new educational model, in the perception questionnaire.

| | Dimension 1 Items | No Answer | Strongly Disagree | Disagree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|
| (1) | The pedagogical principles in the 2017 educational model for English language teaching favor learning effectively in secondary education. | 12% | 10% | 37% | 39% | 2% |
| (2) | The suggested strategies for teaching the English language in the 2017 educational model are suitable for students in telesecondary schools. | 13% | 20% | 46% | 19% | 2% |
| (3) | The suggested strategies for teaching English in the 2017 educational model are suitable for teachers in telesecondary schools. | 14% | 15% | 47% | 22% | 2% |
| (4) | The suggested strategies for teaching English in the 2017 educational model can be easily adapted by teachers in telesecondary schools. | 12% | 15% | 41% | 29% | 3% |
| (5) | The amount of thematic content in English allows telesecondary teachers to develop them properly with their groups. | 12% | 11% | 50% | 25% | 2% |
| (6) | The amount of thematic content in the English courses allows telesecondary students to achieve a certain learning competence in this language. | 14% | 15% | 43% | 26% | 2% |
| (7) | The teaching strategies that I possess help me to teach English as required by the 2017 educational model. | 14% | 8% | 33% | 40% | 5% |

In dimension 2, beliefs about teaching English, the majority of generalist teachers indicated for items 1, 3, 4, and 5 that the parents, colleagues, principals, and leaders in the educational systems did not influence their beliefs about learning English; nonetheless, they believed that their students' opinions played a key role in their teaching (see Table 6).

**Table 6.** Distribution of answers in dimension 2, beliefs about the new educational model, in the perception questionnaire.

| | Dimension 2 Items | No Answer | Strongly Disagree | Disagree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|
| (1) | The opinion of students' parents about learning English influences my perception of the importance of English for the telesecondary students. | 2% | 13% | 41% | 33% | 11% |
| (2) | My students' opinion about learning English influences my perception of the importance of the English language for telesecondary students. | 1% | 8% | 30% | 48% | 13% |
| (3) | My colleagues' opinions about learning English influence my perception of the importance of this language for telesecondary students. | 1% | 12% | 39% | 40% | 8% |
| (4) | My principal's opinions about learning English influence my perception of the importance of this language for telesecondary students. | 1% | 13% | 38% | 43% | 5% |
| (5) | The opinions of the educational stakeholders about learning English in telesecondary school influence my perception of the importance of this language for telesecondary students. | 3% | 13% | 38% | 39% | 7% |

For dimension 3, pedagogical practice, the results in Table 7 show that generalist teachers' practices were not influenced by their coworkers or the curricular materials

(items, 2, 7, and 8). The majority of the teachers tended to positively react to item 3; more than half of the teachers agreed that the recommended English material for telesecondary education had a positive influence when they taught English. Further, they considered that their language proficiency and their students contributed to what they do in the classroom (items 4 and 5). Moreover, their practices were nurtured by their previous language learning and experience and their learners' expectations (items 6 and 9).

**Table 7.** Distribution of answers in dimension 3, pedagogical practice, in the perception questionnaire.

| | Dimension 3 Items | No Answer | Strongly Disagree | Disagree | Agree | Strongly Agree |
|---|---|---|---|---|---|---|
| (1) | The way my telesecondary coworkers teach English influences how I teach English to my students. | 2% | 10% | 56% | 28% | 4% |
| (2) | The English books that are recommended for telesecondary education influence how I teach English. | 2% | 9% | 32% | 48% | 9% |
| (3) | The level of English that I possess influences the activities that I implement with my students. | 1% | 1% | 12% | 63% | 23% |
| (4) | The English level of my students influences the activities I implement with them. | 1% | 3% | 13% | 55% | 28% |
| (5) | The way I learned English influences how I teach English in telesecondary schools. | 2% | 3% | 16% | 60% | 19% |
| (6) | My colleagues' opinions about English language teaching influence how I teach the language to my students. | 1% | 9% | 51% | 34% | 5% |
| (7) | My principal's expectations about teaching English in telesecondary school influence how I teach English to my students. | 1% | 14% | 49% | 33% | 3% |
| (8) | My students' expectations about how to study English influence how I teach my students the language. | 2% | 6% | 23% | 55% | 14% |

### *4.2. Quantitative Results: Appropriation Questionnaire*

In order to identify the level of appropriation of curricular changes, frequency analyses were run on the items in each of the two dimensions of the questionnaire. In these analyses, every item was treated independently to identify positive or negative appropriation patterns among the participants. Table 8 shows the distribution of participants across the possible Likert scale choices.

In dimension 1, teaching practice, the responses to most of the items (5, 7, 8, 11, and 12) in Table 8 indicated that the generalist teachers had a high level of appropriation of teaching practice, as the largest number of participants opted for the high frequency level. The participants indicated that they often implemented music, sports, movies, games, and other activities to favor the use of English using their learning experiences. Nonetheless, items 10, 14, and 18 showed that just less than one half of teachers, from 42% to 49%, rarely implemented speaking activities and did not consider the teaching of linguistic aspects or promote the learning of the cultural aspects of English at school events. Nonetheless, as explained above, items 1, 6, and 9 revealed different answers between versions; therefore, these items were excluded.

**Table 8.** Distribution of answers in dimension 1, teaching practice, in the appropriation questionnaire.

| | Dimension 1 Items | No Answer | Never | Rarely | Often | Very Often |
|---|---|---|---|---|---|---|
| (1) | I implement activities for students to communicate in English using their personal experiences. | 1% | 6% | 40% | 44% | 9% |
| (2) | I implement activities for students to communicate in English based on their interest in music, sports, movies, games, and other activities. | 1% | 3% | 40% | 43% | 13% |
| (3) | The English activities I implement favor the use of the language for a social purpose. | 2% | 3% | 28% | 55% | 12% |
| (4) | In the English class, I provide and explain the instructions that students need to achieve the development of activities. | 1% | 1% | 8% | 64% | 26% |
| (5) | I implement communicative activities (e.g., dialogs, conversations, presentations) in English where students have to retrieve information from previously learned lessons. | 2% | 4% | 28% | 52% | 14% |
| (6) | I implement activities where I use the English language orally. | 1% | 6% | 42% | 41% | 10% |
| (7) | I implement activities to show students strategies on how an English speaker can use the language in oral communication. | 1% | 10% | 42% | 41% | 6% |
| (8) | I implement activities where I show students different strategies that promote written communication in English. | 2% | 4% | 30% | 56% | 8% |
| (9) | I implement activities where differences or similarities between English and Spanish are discussed with the students. | 1% | 5% | 31% | 53% | 10% |
| (10) | I implement activities where the communicative functions of English are discussed with the students. | 1% | 8% | 38% | 46% | 7% |
| (11) | I implement activities where linguistic aspects of the English language are discussed with my students. | 1% | 7% | 46% | 39% | 7% |
| (12) | I select material, printed or multimedia, that promotes the student's contact with the English language. | 1% | 0% | 19% | 59% | 21% |
| (13) | I implement activities to raise awareness about the importance of the language. | 1% | 1% | 21% | 62% | 15% |
| (14) | In my activities, I promote an environment of respect for students to practice English. | 1% | 1% | 12% | 58% | 28% |
| (15) | I promote the learning of cultural aspects of English in school events. | 1% | 13% | 49% | 31% | 6% |

In the analysis of the answers to the items in dimension 2, evaluation of learning, all items showed a high level of appropriation for the evaluation suggested in the curriculum. The number of participants ranged from 55% to 70% (see Table 9). The results indicated that the teachers implemented formative evaluation, where they considered different qualitative and quantitative aspects that provided evidence of their learners' learning progress and continuous performance. Moreover, they indicated that their evaluation focused on the communication competencies that they gradually developed during the school year.

These generalist teachers indicated that they often analyzed the congruence of the evaluation process, evaluation material, and the students' interaction during English lessons. Furthermore, the results show that teachers reported evaluating competences established in the curriculum that they were using. Moreover, they often evaluated their learners using continuous, permanent, and formative evaluation.

**Table 9.** Distribution of answers in dimension 2, evaluation, in the appropriation questionnaire.

| | Dimension 2 Items | No Answer | Never | Rarely | Often | Very Often |
|---|---|---|---|---|---|---|
| (1) | I provide information about the degree of progress for each student in the English class. | 0% | 4% | 24% | 55% | 17% |
| (2) | I identify the progress of my students in English proficiency at the end of each period. | 0% | 1% | 10% | 68% | 21% |
| (3) | I analyze the congruence between the evaluation activities I implement and the purpose of English for the grade I teach. | 1% | 1% | 14% | 68% | 16% |
| (4) | I analyze the congruence between the evaluation materials I implement and the purpose of English for the school grade I teach. | 1% | 1% | 13% | 70% | 15% |
| (5) | I consider the interaction that my students have in English during the class during evaluation. | 1% | 1% | 20% | 64% | 14% |
| (6) | I consider the student's performance during the development of the L2 activities in the evaluation English. | 0% | 1% | 8% | 67% | 24% |
| (7) | I consider the progress made by the students in the period with respect to their own starting point in the evaluation. | 2% | 1% | 12% | 70% | 15% |
| (8) | For English learning evaluations, I consider the level of achievement that is established for each school cycle in the grade I teach. | 1% | 4% | 22% | 58% | 15% |
| (9) | I consider the competences that are established in each school cycle in the grade I teach to evaluate English learning. | 1% | 4% | 21% | 63% | 11% |
| (10) | I consider the development of communicative skills in English for each student to evaluate them. | 1% | 3% | 20% | 57% | 19% |
| (11) | To evaluate English learning, I carry out a continuous, permanent, and formative evaluation of each of my students. | 1% | 0% | 9% | 57% | 33% |
| (12) | To evaluate English learning, I consider evidence of learning in the products students delivered to me. | 0% | 1% | 6% | 62% | 31% |
| (13) | To evaluate the English learning of my students, I consider qualitative aspects for each one of them. | 1% | 1% | 13% | 65% | 20% |
| (14) | To evaluate the English learning for my students, I consider the strengths of each of them. | 0% | 1% | 9% | 67% | 23% |
| (15) | To evaluate the English learning of my students, I consider the weaknesses of each one of them. | 1% | 3% | 14% | 60% | 22% |
| (16) | To evaluate my students' English learning, I select instruments that allow me to have clear results regarding their learning. | 0% | 1% | 11% | 64% | 24% |

### *4.3. Quantitative Results: Perception and Appropriation Interphase*

In order to examine the potential interphase between the two central constructs of the study, Spearman correlation analyses were run on the global scores that the participants obtained in both questionnaires. These analyses yielded a significant correlation between the results of the perception questionnaire and appropriation questionnaire ($p$ = 0.024; $r_s$ = 0.153).

The analysis results indicated that this correlation was positive but weak. In other words, the analysis results suggested that as the perception of the curriculum among the participants improved, the appropriation of the recommended curricular guidelines became more systematic. Based on this statistical result, the alternative hypothesis was retained.

### 4.4. Qualitative Results: Interview Data

With the use of Atlas.ti software, during the qualitative analyses, 10 categories and 80 subthemes emerged during the first cycle analysis. During the second cycle, the themes were condensed into three macro-categories based on the three research questions. Table 10 shows that the condensed themes centered upon the ELT challenges, curricular transition, and training. Across the macro-categories (themes) and subthemes, the interview data pointed to different positive and negative factors that underpin the relationship between perception and appropriation.

**Table 10.** Themes and subthemes in the second cycle qualitative data analyses.

| Theme | Subtheme | Definition of Theme | Answer to Question |
|---|---|---|---|
| ELT Challenges | | Issues that teachers face in the implementation of the curricular reforms. | |
| | Deficiency. | | 3 |
| | Implementing the language content of the new educational model. | | 2 |
| | Interpretation of the language curriculum. | | 2 |
| | New educational model's material. | | 3 |
| | Teachers' perception. | | 1 |
| | Teaching barrier. | | 3 |
| | Teaching material support. | | 3 |
| | Teaching strategies. | | 2 |
| | Using technology. | | 2 |
| Curricular Transition | | The transition process from the previous to the current curriculum. | |
| | Adaptation and modification of language curriculum. | | 2 |
| | Knowledge of the 2017 curriculum. | | 2 |
| | Lack of personal initiative to implement the 2017 curriculum. | | 1 |
| | Changes from the old to new curriculums. | | 3 |
| | Imposition of new curricular reforms. | | 3 |
| | Resistance to implement the 2017 curriculum. | | 3 |
| | Being dependent on certain transitional curriculum. | | 3 |
| Training | Deficiency | Development of pedagogical skills | 3 |

On the positive side, the qualitative data revealed that some teachers tended to accept the curricular reforms. To this end, they adapted and modified the English syllabus through consensus with peers in their school board. Moreover, they indicated that although a few teachers avoided their responsibilities for teaching English, they all needed to assume the responsibility to teach English to their students in rural schools. They considered that the adaptation of the curriculum would enhance their students' learning, as the next excerpts illustrate:

> Excerpt 1 . . . *I think that learning English for telesecondary students is complex. And yes, it may be, because students come from schools where they have not taken English lessons. Here in the telesecondary school is the first time they have English lessons. They come from elementary schools where they never studied English, so we want them to learn English, that is the reason we modify the programs to help our students*. (Addy)

> Excerpt 2 . . . *to help our students we made a consensus to adapt and modify the language curriculum because when we reviewed the English material, we noticed that the contents were so complicated, I mean very complex* . . . (Cando)

Moreover, these generalist teachers assumed that the curricular adaptations helped their students complete significant learning activities using their own context. Likewise, teachers selected material with content that helped their students learn; for example, they reused English books that they used during the implementation of previous curricula. In addition to the use of elements from the various curricula, teachers considered the use of tutorials, songs, translators, and games that helped students interact with the English language. In addition to the adaptations, teachers showed willingness toward professional development courses in order to implement the new curriculum in their telesecondary school.

> Excerpt 3 *I am working with the 2017 curriculum, using strategies of the 2011 syllabus. And I am working with English books from 2006 to guide me when teaching English* (Carmen)

> Excerpt 4 . . . *the new curriculum is not clear to me. . . . but I like to do the best I can for example I use a Duolingo app, translator and I use apps for pronunciation, so I try to implement some aspects of the new curriculum. We have sung in English for example the Beatles' songs* (German)

> Excerpt 5 . . . *We know that this 2017 plan is good, everything is supposed to be planned and I think the program is good, but we as teachers . . . we need training on the educational tools the Ministry of Education gives us. Why? In order to improve what we do in the classroom. . . . they deliver the training courses online at four or five in the afternoon . . . and you have to be at home with your computer, paying for your internet, uploading evidence and talking, taking the class, right?* (Wilbert)

Nonetheless, on the negative side, it was observed that teachers had many limitations when implementing the 2017 curriculum. The transition and challenge of implementing the 2017 curriculum seemed to bring about issues that were not acknowledged in current curricular reforms. The following excerpts illustrate some of these issues.

> Excerpt 6. *I think that the new curriculum is complex to be implemented in my English lessons, because, on one side I have little knowledge about it. I mean, the language demands of the 2017 curriculum imply a high level of English proficiency that the students in our context don't have; that is why I can't implement the curriculum at all* (Lizbeth)

The previous excerpt indicates that the students' language proficiency was a major challenge. In addition, the teachers also acknowledged that their own proficiency was a negative factor that needed addressing. They considered that the curricular language demands did not consider the extent to which the linguistic abilities of the students and the teachers hindered the learning of English. Moreover, the personal financial implications of those demands were not accounted for, as teachers needed to defray the expenses of pedagogical and language training using their own resources.

> Excerpt 7. . . . *we received a yearlong training course from the Ministry of Education, but it did not focus on English, we are not trained in that aspect. . . . I like to learn English to teach it, so I pay for English training because I am interested in how to teach English* (Grecia)

Nevertheless, teachers considered that the transition of the 2006, 2011, and 2017 models had had a negative impact on their English pedagogical practices. Moreover, the little knowledge they had about the 2017 model made them modify the language curriculum or in some cases reject it.

> Extract 8 . . . *in fact we have not seen, we do not take into account the English subject. I mean that teaching English is not taken into account, and we did not review the current plans and programs of study in-depth. Because we were worried about what was coming with the 2017 reform. Thus, we only took a look at the English language content. We do nothing, because we were truly more concerned about our assessment* (Natalia)

> Extract 9 . . . *I've been . . . how will I say? . . . negligent in that aspect, I don't like the word negligent very much, but I have to accept it . . .* (German)

> Extract 10 . . . *I would love to do dialogues but I cannot speak or pronounce them, so I cannot say that I am going to have a dialogue in the English class, so I cannot do what I want to do . . . We are not sure of what we teach in English. That is our problem. You know, when we teach English, we are afraid of having writing, grammar, or pronunciation mistakes when we use English in the classroom. That is so alarming for us. But we do not show that in class. But I have reflected on it, so I have realized that there are teachers who do know a little but they try it. And also I think that my students don't know English, so why should I be afraid of teaching English, right?* (Addy)

The excerpts above show that the teachers experienced confusion vis-à-vis the ongoing language curricular changes. It seems that this negative behavior was embedded in their low proficiency in English and the low proficiency levels of their students. In light of these issues, they felt unable to cover the content of the curriculum and implement proper teaching English strategies. To counteract this reality, they modified and adapted the curriculum to the best of their knowledge and within their language proficiency.

### *4.5. Summary of Findings*

The quantitative analyses yielded a significant positive correlation between opinions and appropriation; this suggests that as teachers' perceptions became more positive, the level of appropriation increased. Nonetheless, this correlation was weak. The qualitative findings provide some insights to better understand the weakness of the correlation. For instance, the interview data indicated that the generalist teachers experienced fear when teaching their English lesson because of their English proficiency and that of their students. This proficiency issue made them feel uncomfortable regarding the teaching of the productive skills and the provision of corrective feedback. Moreover, while the teachers were aware of the existence of the 2017 curriculum, they had no training on its content and the implementation of proper language learning tasks. Therefore, they included elements of the curricula they already knew and with which they felt more comfortable.

## 5. Discussion

Regarding research question 11, the quantitative results show that generalist teachers had a negative opinion about the ELT curricular changes. This finding was corroborated during the interviews, where they expressed disagreement with the new curriculum. This dissatisfaction constitutes a well-known challenge for the acceptance of EFL educational reforms in other international contexts such as Libya [32,40] Taiwan [7], Vietnam [39], Turkey, and Argentina [38]. It should be noticed, though, that while the opinion of our participants is not in line with their expectations of the curriculum, their beliefs and pedagogical practices showed a positive tendency in the quantitative results (for similar results, see [3,7,17,19,29,32,41]). Nonetheless, in order to adhere to the guidelines, generalist teachers have to deal with issues such as L2 proficiency, teacher training, and pedagogical support in the same way as specialized English teachers do in rural areas [12,23].

Our qualitative and quantitative data indicate that the generalist teachers value some aspects of the curriculum and, thus, modify the content of the curriculum and use different materials. Teachers' beliefs about what they should teach is nurtured by their students' opinions, context, and English proficiency. Previous research has revealed that, due to border-crossing issues, the generalist teachers in Mexican rural areas make efforts to adapt the curriculum content in order to help children with immigrant parents or relatives [2]. The curricular adaptations of generalist teachers in Mexico diverge from empirical evidence in other Latin-American studies where rural teachers showed a passive engagement in teaching English [23]. Furthermore, rural teachers in other countries were found not to value the learning of English as much as Mexican rural generalist teachers do. This finding brings about questions on how the border sharing conditions between Mexico and the

United States influence the perception of rural teachers on the national EFL curriculum, its demands, and its curricular guidelines.

In regard to research question 2, two appropriation dimensions were considered: teaching practices and the evaluation of learning. These dimensions explored how teachers interpret and adopt the curriculum [22,30]. The quantitative and qualitative results showed that generalist teachers exhibited a positive trend in teaching practice and evaluation process. For the appropriation of teaching practices, the questionnaire data show that the teachers adopted the new curriculum based on their students' contextual reality. The interview data allowed us to see that the teachers mixed the teaching strategies recommended across the various curricular reforms. They often did so despite technological, pedagogical, and linguistic limitations. Nonetheless, two areas of the EFL curriculum that the teachers did not consider were oral and written production. Due to their low English proficiency, it was difficult for them to prepare fluent conversations [2], and when they implemented oral production activities, they were hesitant on the accuracy of the language they were delivering. Moreover, they felt limited in terms of the amount and type of feedback they provided.

The quantitative data revealed a high level of appropriation in terms of the evaluation recommended in the curriculum. During the interview, the teachers explained that, as recommended in the curriculum, they adhered to formative evaluation. Throughout the course, they considered the content that learners grasp and their performance in the activities. While our participants indicated adherence to the type of evaluation stated in the curriculum, a discrepancy was observed between what should be evaluated and what is evaluated. The curriculum states that EFL education should focus on the use of the target language for communicative purposes. Nonetheless, the teachers' evaluation activities centered upon word identification, sentence making, verb conjugation, and sentence ordering. These findings instantiate that the lack of EFL proficiency and formal teacher education not only hinders the adoption of teaching practices but also affects the evaluation process, despite the willingness of generalist teachers to comply with curricular guidelines.

Regarding research question 3, although the quantitative findings confirmed the alternative hypothesis, the qualitative data provided evidence of a positive interaction and a negative interaction between teachers' perceptions and the appropriation of curricular changes, respectively. This, in turn, can explain the weak correlation between the two constructs of interest in the study. For example, generalist teachers modify and adapt the language curriculum based on students' needs and context. However, this modification is made using the 2006, 2011, and 2017 curricular content. Moreover, many generalist teachers develop their classes based on the 2011 curriculum and implement strategies from the 2017 syllabus and books from the 2006 model. This finding shows congruence with Park and Sung's (2013) international research. Using interviews, these authors showed that teachers interpreted the curriculum by selecting certain content and teaching strategies during the curricular transition [64]. Moreover, Taylor and Marsden (2014) found, using qualitative and quantitative data, that teachers interpret the curricular changes based on their teaching experience and beliefs [20].

In regard to negative factors, our evidence shows that generalist teachers received little information about the 2017 curriculum. This finding is similar to that in other international [32,38,40] and national studies [2,24,65–67] that reveal that generalist and language specialist teachers implement the new curriculum without training. To overcome this absence of curricular knowledge, some generalist teachers pay for training in ELT and information and communication technology. Thus, they undertake professional development for the enactment of the new curricular reforms.

Another aspect that counteracts the interphase between teachers' perception of the curricular changes and appropriation of educational practices is the discomfort that teachers feel about the language teaching demands. The teachers often feel overwhelmed by the content of the language curriculum [2]. However, this finding is not particular to generalist

teachers with low proficiency levels of EFL. Some studies indicate that even generalist teachers who completed English language courses and EFL teacher training also feel overwhelmed by the EFL curriculum.

Fear is an additional issue in ELT among generalist teachers when they need to implement curricular changes. This factor could be interpreted, at the emotional level, as anxiety [28]. Our qualitative findings indicate that generalist teachers do not feel secure about how and what they teach. In the interview data, the teachers reported that they are preoccupied about how they handle critical issues, for example, social issues, such as family violence, economic resources, psychological problems, sexual violence, hunger, and agricultural responsibilities, that the curriculum does not consider for the organization of their lessons [2]. These teachers considered that these factors hinder English language teaching and learning in rural areas but are disregarded in curricular reforms. Other studies [23] showed similar results among Nicaraguan rural teachers who faced similar issues. Therefore, these findings confirm that generalist teachers struggle to implement their English classes at the emotional level. In turn, all of these eventualities keep generalist teachers in a state of emotional instability [17].

Our study provides some valuable information about the constructs of interest. For instance, our empirical evidence shows that the construct of perception is built upon different dimensions, opinions, beliefs, and pedagogical practices. Nonetheless, our findings indicate that the dimension of emotions showed a lack of stability due to a fluctuation process. Thus, while we were able to identify teacher disagreement in regard to the language curriculum, questions arise about how teachers feel teaching an aspect of the curriculum they are not ready for. Moreover, an interesting aspect of our study was that the statistical results revealed a positive correlation between perception and appropriation. Nonetheless, this correlation was weak. The qualitative evidence provided some information to have an initial idea of the factors that hinder the correlation between the two constructs. Nevertheless, we considered that a longitudinal observational classroom study might provide further information on the level of appropriation and the implementation of the curricular changes. While this type of study is desirable and valuable, researchers might encounter that only a few teachers are willing to participate in longitudinal studies due to time constraints and a fear of observation [2].

Finally, the three research questions were answered, but some methodological modifications could have helped us collect more informative data. For example, probabilistic sampling was not possible due to the geographical location of the rural schools. While the quantitative and qualitative validity procedures instantiate the internal validity of the results, the use of probabilistic sampling would be better for future research, as it could enhance the representativeness and generalizability of our findings. Finally, we consider that our qualitative instrument was not sensitive enough to facilitate a deeper exploration of the participants' reality. Hence, other criteria should have been considered during the organization of the interview questions in order to increase the trustworthiness of the qualitative findings. Although the open-ended nature of the interview questions allowed us to go deeper into the answers of the participants, the use of personalized interviews could have elicited individual data to better understand how each participant deals with EFL education and the implementation of the new curriculum vis-à-vis the reality of their students.

## 6. Conclusions

Language curricular changes bring about substantial challenges for English language teaching in public education. When these changes are enforced in rural areas, generalist teachers face major challenges [2]. Nonetheless, their challenges have been particularly underestimated. Our quantitative and qualitative data showed that this kind of teacher population believes that adopting curricular content and teaching strategies could help their students learn English. Moreover, rural generalist teachers are convinced that English is essential for their students. Thus, they are willing to invest in their teaching practice and

language competencies. These findings contrast with those from specialist English teachers who consider that public EFL education is of little value to their students [65–67]. In turn, the generalist teachers make extra effort to comply with EFL curricular demands, and their efforts need to be considered by stakeholders and policymakers. Due to their willingness to enhance the ELT process, generalist teachers could be included in mentoring projects that help them develop their ELT practice. Moreover, they could be part of collaborative projects with specialized EFL teachers. Based on the evidence from this study and other studies conducted with generalist teachers who are obligated to teach English in rural schools, policy planners should pay attention to these teachers and be willing to engage in bottom-up curricular development processes. Generalist teachers have proven to be knowledgeable about the challenges of the ELT curricular reforms. However, above all, they have provided ample evidence of commitment and engagement with the teaching of a discipline that is far beyond their own professionalization and training. Thereafter, their voice should be heard during the organization of curricular reforms and the conceptualization of the public education ELT curriculum.

**Author Contributions:** Both authors equally contributed during the conceptualization and realization of the study, as well as during the preparation of the manuscript. All authors have read and agreed to the publish 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 local Ministry of Education as shown by approval letter number 1016/2019 dated 7 May 2019.

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

**Data Availability Statement:** To obtain access to the data and instruments, please contact the authors.

**Acknowledgments:** We are grateful to the participants and the local institution for their support during the realization of the project. We regret that their names cannot be openly acknowledged due to anonymity and ethical concerns.

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

## References

1. Nunan, D. The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *Tesol Q.* **2003**, *37*, 589–613. [CrossRef]
2. Izquierdo, J.; Zúñiga, S.P.A.; Martínez, V.G. Foreign language education in rural schools: Struggles and initiatives among generalist teachers teaching English in Mexico. *Stud. Second Lang. Learn. Teach.* **2021**, *11*, 133–156. [CrossRef]
3. Yuan, R. Appropriating national curriculum standards in classroom teaching: Experiences of novice language teachers in China. *Int. J. Educ. Res.* **2017**, *83*, 55–64. [CrossRef]
4. Kaniuka, T. Toward an understanding of how teachers change during school reform: Considerations for educational leadership and school improvement. *J. Educ. Chang.* **2012**, *13*, 327–346. [CrossRef]
5. Borg, S. *Teacher Cognition and Language Education. Research and Practice*; Bloomsbury: New York, NY, USA, 2006.
6. Cassels, J.; Freeman, R. Negotiation of language education policies guided by educators' experiences or identity (individual): Appropriating language policy on the local level. In *Negotiating Language Policies: Educators as Policymakers*, 1st ed.; Menken, K., García, O., Eds.; Routledge: New York, NY, USA, 2010; pp. 13–31.
7. Luo, W.H. Teacher perceptions of teaching and learning English as a lingua franca in the expanding circle: A study of Taiwan. *Engl. Today* **2016**, *33*, 2–11. [CrossRef]
8. Ardzejewska, K.; Mcmaugh, A.; Coutts, P. Delivering the primary curriculum: The use of subject specialist and generalist teachers in NSW. *Issues Educ. Res.* **2010**, *20*, 203–219.
9. Carr, W. Fear of teaching French: Challenges faced by generalist teachers. *Can. Mod. Lang. Rev.* **1999**, *56*, 155–178. [CrossRef]
10. Hansen-Thomas, H.; Richins, L.G. ESL mentoring for secondary rural educators: Math and science teachers become second language specialists through collaboration. *Tesol J.* **2015**, *6*, 766–776. [CrossRef]
11. Serrano, J.; Lera, A.; Contreras, O. Maestros generalistas vs especialistas: Claves y discrepancias en la reforma de la formación inicial de los maestros de primaria. *Rev. Educ.* **2007**, *344*, 247–250.
12. Roldán, A.M.; Peláez Henao, O.A. Pertinencia de las políticas de enseñanza del inglés en una zona rural de Colombia: Un estudio de caso en Antioquia. *Ikala Rev. Leng. Cult.* **2017**, *22*, 121–139. [CrossRef]

13. Russo, J.; Corovic, E.; Hubbard, J.; Bobis, J.; Downton, A.; Livy, S.; Sullivan, P. Generalist Primary School Teachers' Preferences for Becoming Subject Matter Specialists. *Aust. J. Teach. Educ.* **2022**, *74*, 38–57. [CrossRef]
14. Zein, M.S. Elementary English education in Indonesia: Policy developments, current practices, and future prospects. *English Today* **2017**, *33*, 53–59. [CrossRef]
15. Secretaría de Educación Pública, SEP. *Aprendizajes Clave para la Educación Integral*, 2nd ed.; Secretaría de Educación Pública: Mexico City, Mexico, 2017.
16. Samdal, O.; Wold, B.; Bronis, M. Relationship between students' perceptions of school environment, their satisfaction with school and perceived academic achievement: An international study. *Sch. Eff. Sch. Improv.* **1999**, *10*, 296–320. [CrossRef]
17. Saunders, R. The role of teacher emotions in change: Experiences, patterns and implications for professional development. *J. Educ. Change* **2013**, *14*, 303–333. [CrossRef]
18. Borg, S. Language teacher cognition: Perspectives and debates. In *Second Handbook of English Language Teaching*, 2nd ed.; Gao, X., Ed.; Springer: Cham, Switzerland, 2019; Volume 1, pp. 1–23. [CrossRef]
19. Pajares, F. Teachers' beliefs and educational research: Cleaning up a messy construct. *Rev. Educ. Res.* **1992**, *62*, 307–332. [CrossRef]
20. Taylor, F.; Marsden, E. Perceptions, attitudes, and choosing to study foreign languages in England: An experimental intervention. *Mod. Lang. J.* **2014**, *98*, 902–920. [CrossRef]
21. Zhang, Y.; Hu, G. Between intended and enacted curricula: Three teachers and a mandated curricular reform in Mainland China. In *Negotiating Language Policies: Educators as Policymakers*, 1st ed.; Menken, K., García, O., Eds.; Routledge: New York, NY, USA, 2010; pp. 123–142.
22. Brown, C. Advancing policy makers' expertise in evidence-use: A new approach to enhancing the role research can have in aiding educational policy development. *J. Educ. Chang.* **2014**, *15*, 19–36. [CrossRef]
23. Coelho, F.; Henze, R. English for what? Rural Nicaraguan teachers' local responses to national educational policy. *Lang. Policy* **2014**, *13*, 145–163. [CrossRef]
24. Hernández, M.; Izquierdo, J. Cambios curriculares y enseñanza del inglés: Cuestionario de percepción docente. *Sinéctica Rev. Electron. Educ.* **2020**, *54*, 1–22. Available online: https://sinectica.iteso.mx/index.php/SINECTICA/article/view/1042 (accessed on 25 June 2020). [CrossRef]
25. Rulison, S.R. Teachers' Perceptions of Curricular Change. Ph.D. Thesis, University of Oklahoma, Norman, OK, USA, 2012. Available online: http://ezproxy.depaul.edu/login?url=https://search-proquest-com.ezproxy.depaul.edu/docview/1015169185?accountid=10477 (accessed on 10 November 2020).
26. Niu, R.; Andrews, S. Commonalities and discrepancies in L2 teachers' beliefs and practices about vocabulary pedagogy: A small culture perspective. *Tesol J.* **2012**, *6*, 134–154.
27. Aquino, Z.S.; Izquierdo, J.; Echalaz, B. Evaluación de la práctica educativa: Una revisión de sus bases conceptuales. *Rev. Actual. Investig. Educ.* **2013**, *13*, 1–22. Available online: http://www.scielo.sa.cr/pdf/aie/v13n1/a02v13n1 (accessed on 25 June 2020).
28. Oxford, R.L. Emotion as the amplifier and the primary motive: Some theories of emotion with relevance to language learning. *Stud. Second Lang. Learn. Teach.* **2015**, *5*, 371–393. [CrossRef]
29. Tsang, K.; Lianjiang, J. Sociological understandings of teachers' emotions in second language classrooms in the context of education/curricular reforms. In *Emotions in Second Language Teaching: Theory, Research and Teacher Education*; Martínez, J., Ed.; Springer: Cham, Switzerland, 2018; pp. 53–72. [CrossRef]
30. Grossman, L.; Smagorinsky, P.; Valencia, S. Tools for teaching English: A theoretical framework for research on learning to teach. *Am. J. Educ.* **1999**, *108*, 1–29. [CrossRef]
31. Peláez, O.; Usma, J. The crucial role of educational stakeholders in the appropriation of foreign language education policies: A case study. *Profile* **2017**, *19*, 121–134. [CrossRef]
32. Altaieb, S.R. Teachers' Perception of the English Language Curriculum in Libyan Public Schools: An Investigation and Assessment of Implementation Process of English Curriculum in Libyan Public High Schools. Doctor of Philosophy. Ph.D.Thesis, University of Denver, Denver, CO, USA, 2013. Available online: https://digitalcommons.du.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1025&context=etd (accessed on 5 May 2020).
33. Tao, J.; Gao, X. Teacher agency and identity commitment in curricular reform. *Teach. Teach. Educ.* **2017**, *63*, 346–355. [CrossRef]
34. Heyworth, F. The Common European Framework. *ELT J.* **2006**, *60*, 181–183. [CrossRef]
35. Boag, N.H. Teacher Perception of Curricular Change. Ph.D. Thesis, The University of Alberta, Edmonton, AB, Canada, 1980.
36. Graves, K.; Garton, S. An analysis of three curriculum approaches to teaching English in public-sector schools. *Lang. Teach.* **2017**, *50*, 441–482. [CrossRef]
37. Moradi, K.; Sepehrifar, S.; Khadiv, T.P. Exploring Iranian EFL teachers' perceptions on supervision. *Procedia Soc. Behav. Sci.* **2014**, *98*, 1214–1223. [CrossRef]
38. Banegas, L. Exploring perceptions of curriculum change in initial English language teacher education: A case in Argentina. *Estud. Sobre Educ.* **2016**, *31*, 75–96. [CrossRef]
39. Nguyen, H.T.M.; Bui, T. Teachers' agency and the enactment of educational reform in Vietnam. *Curr. Issues Lang. Plan.* **2016**, *17*, 88–105. [CrossRef]
40. Orafi, S.M.S.; Borg, S. Intentions and realities in implementing communicative curriculum reform. *System* **2009**, *37*, 243–253. [CrossRef]

41. Yan, C. "We can only change in a small way": A study of secondary English teachers' implementation of curriculum reform in China. *J. Educ. Change* **2012**, *13*, 431–447. [CrossRef]
42. Carless, D.R. A case study of curriculum implementation in Hong Kong. *System* **1998**, *26*, 353–368. [CrossRef]
43. Creswell, J.W. *A Concise Introduction to Mixed Methods Research*; Sage: Thousand Oaks, CA, USA, 2015.
44. Mackey, A.; Gass, S.M. *Second Language Research: Methodology and Design*, 1st ed.; Lawrence Erlbaum: Mahwah, NJ, USA, 2005.
45. Hoy, W.K. *Quantitative Research in Education*; Sage: Washington, DC, USA, 2010.
46. Noordzij, M.; Tripepi, G.; Dekker, F.W.; Zoccali, C.; Tanck, M.W.; Jager, K. Sample size calculations: Basic principles and common pitfalls. *Nephrol. Dial. Transplant.* **2010**, *25*, 1388–1393. [CrossRef] [PubMed]
47. Creswell, J.W.; Creswell, J.D. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2018.
48. Secretaria de Educación Pública, SEP. Programa Nacional de Inglés en Educación Básica, Segunda Lengua: Inglés. Fundamentos Curriculares. 2011. Available online: ttps://www.gob.mx/sep/documentos/programa-ingles-primaria?state=published (accessed on 15 May 2019).
49. Boyle, J.; Fisher, S. *Educational Testing: A Competence-Based Approach*; Blackwell: Oxford, UK, 2007. [CrossRef]
50. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*, 8th ed.; Routledge: New York, NY, USA, 2018.
51. Field, A. *Discovering Statistics Using SPSS*, 5th ed.; Sage: Oaks, CA, USA, 2017.
52. Yin, R.K. *Case Study Research Design and Methods*, 5th ed.; Sage: Oaks, CA, USA, 2014.
53. Glass, G.; Hopkins, K. *Statistical Methods in Education and Psychology*; Allyn and Bacon: Boston, MA, USA, 1996.
54. Hodge, D.R.; Gillespie, D.F. Permanency for children in foster care: A competing Risks Analysis. *J. Soc. Serv. Res.* **2007**, *33*, 1–12. [CrossRef]
55. Dörnyei, Z. *Questionnaires in Second Language Research*, 1st ed.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2003. [CrossRef]
56. Fabila Echauri, A.M.; Minami, H.; Izquierdo, J. La escala de Likert en la evaluación docente: Acercamiento a sus características y principios metodológicos. *Perspect. Docentes* **2012**, *50*, 31–40. [CrossRef]
57. Leung, S.O. A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *J. Soc. Serv. Res.* **2011**, *37*, 412–421. [CrossRef]
58. Johnson, P.O. Development of the sample survey as a scientific methodology. *J. Exp. Educ.* **1959**, *27*, 167–176. [CrossRef]
59. Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* **2010**, *15*, 625–632. [CrossRef]
60. Guy, R.F.; Norvell, M. The neutral point on a Likert scale. *J. Psychol. Interdiscip. Appl.* **1977**, *95*, 199–204. [CrossRef]
61. Ravitch, S.M.; Carl, N.M. *Qualitative Research. Bridging the Conceptual, Theoretical, and Methodological*; Sage: Los Angeles, CA, USA, 2016.
62. Dauzón-Ledesma, L.; Izquierdo, J. Language learning investment in Higher Education: Validation and implementation of a Likert-scale questionnaire in the context of compulsory EFL Learning. *Educ. Sci.* **2023**, *13*, 370. [CrossRef]
63. Williams, F.; Monge, P. *Reasoning with Statistics: How to Read Quantitative Research, 5th, ed.*; Thomson Wadsworth: Belmont, CA, USA, 2001.
64. Park, M.; Sung, Y.K. Teachers' perceptions of the recent curriculum reforms and their implementation: What can we learn from the case of Korean elementary teachers? *Asia Pac. J. Educ.* **2013**, *33*, 15–33. [CrossRef]
65. Padilla, L.; Espinoza, L. La práctica docente del profesor de inglés en secundaria. Un estudio de casos en escuelas públicas. *Sinéctica* **2015**, *44*, 1–18. Available online: https://sinectica.iteso.mx/index.php/SINECTICA/article/view/164/157 (accessed on 1 May 2020).
66. Banks, L. English Teachers in Mexico: Initial Preparation and the Realities of Practice. Doctor of Education. Ph.D. Thesis, California State University, San Diego, CA, USA, 2017. Available online: https://csusmdspace.calstate.edu/bitstream/handle/10211.3/193053/BanksLeticia_Spring2017.pdf?sequence=1 (accessed on 5 September 2020).
67. Trujeque, A.; Jiménez, M. *An Observational Study in an English Class in the Public Secondary School Context in Tabasco*; Bachelor's Research Monograph; Universidad Juárez Autónoma de Tabasco: Villahermosa, Mexico, 2013.

education sciences

MDPI

*Article*

# Developing Second Language Learners' Sociolinguistic Competence: How Teachers' CEFR-Related Professional Learning Aligns with Learner-Identified Needs

**Katherine Rehner [1,*] and Ivan Lasan [2]**

[1] Department of Language Studies, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada
[2] Department of Curriculum, Teaching and Learning, OISE, University of Toronto, 252 Bloor Street West, Toronto, ON M5S 1V6, Canada
* Correspondence: katherine.rehner@utoronto.ca

**Abstract:** This article explores how teachers' professional learning about the Common European Framework of Reference (CEFR) can re-orient their reported teaching practices to meet learner-identified sociolinguistic needs. To this end, the article first examines learners' sociolinguistic needs by exploring the extent to which post-secondary French-as-a-second-language (FSL) learners, who completed their elementary and secondary schooling in Ontario, Canada, believe that they have successfully developed sociolinguistic competence in their target language. Specifically, it considers the learners' assessment of their sociolinguistic abilities, the types of sociolinguistic skills they wish to develop further, a comparison with their actual sociolinguistic performance, and the ways in which they hope to develop the sociolinguistic skills they feel they lack. Second, the article explores Ontario elementary- and secondary-school FSL teachers' reported focus on sociolinguistic competence in their teaching after having engaged in intensive and extensive CEFR-oriented professional learning. Specifically, it considers how the teachers' professional learning influences the sociolinguistic relevance of their planning, classroom practice, and assessment and evaluation. The article concludes by considering whether the degree of "fit" between the learners' self-identified needs and the teachers' reports of their re-oriented practices is poised to improve the sociolinguistic outcomes of Ontario FSL learners.

**Keywords:** CEFR; sociolinguistic competence; French as a second language; teacher practice; learners' self-identified needs

**Citation:** Rehner, K.; Lasan, I. Developing Second Language Learners' Sociolinguistic Competence: How Teachers' CEFR-Related Professional Learning Aligns with Learner-Identified Needs. *Educ. Sci.* **2023**, *13*, 282. https://doi.org/10.3390/educsci13030282

Academic Editor: Martin Howard

Received: 14 February 2023
Revised: 24 February 2023
Accepted: 2 March 2023
Published: 7 March 2023



## 1. Introduction

Sociolinguistic agency is the "socioculturally mediated act of recognizing, interpreting, and using the social and symbolic meaning-making possibilities of language". It is predicated on an understanding that "the use of one linguistic variant or another simultaneously reflects and creates the context in which it is used" and is "a performance of one's social identity at the time of utterance" [1] (p. 237). As such, to help language learners enact sociolinguistic agency in their target language interactions, it is necessary to support their development of sociolinguistic competence [1,2]. Sociolinguistic competence consists of the knowledge of social contexts and the socio-stylistic value of the range of variants associated with these contexts. In the field of second language acquisition (SLA), such competence has long been recognized as an integral component of communicative competence [3]. The development of sociolinguistic competence and sociolinguistic agency is one of the explicit aims of the Common European Framework of Reference (CEFR) in that it views learners as "language users and social agents" and language as "a vehicle for communication rather than as a subject to study". To achieve this aim, the framework "proposes an analysis of

learners' needs" in order to help them develop the ability "to act in real-life situations" in a variety of social contexts [4] (p. 29).

With this aim of the CEFR in mind, the present paper addresses the following research questions in relation to the context of French-as-a-second-language (FSL) education in the Canadian province of Ontario:

1. What needs do FSL learners identify concerning their development of sociolinguistic competence? Specifically:
    a. What are their beliefs about the extent of their sociolinguistic abilities?
    b. What types of sociolinguistic skills would they like to develop further?
    c. How do their beliefs compare to their actual sociolinguistic performance, as captured during a semi-structured interview in French?
    d. In what ways would they prefer to acquire the sociolinguistic skills they feel they lack?
2. In what ways do FSL teachers feel that their CEFR-oriented professional learning is making sociolinguistic competence more central to their teaching practices? Specifically:
    a. What impacts do they report on their planning practices?
    b. What impacts do they report on their classroom practices?
    c. What impacts do they report on their assessment and evaluation practices?
3. Is the degree of "fit" between the learners' self-identified sociolinguistic needs and the teachers' reports of their reoriented teaching practices likely to lead to increased sociolinguistic competence for Ontario FSL learners?

## 2. Literature Review

This review focuses, first, on the literature addressing FSL learners' sociolinguistic development, specifically their metalinguistic knowledge and patterns of sociolinguistic variants use. It then considers the literature examining FSL teachers' practices, particularly as concerns the CEFR in FSL teaching in Ontario, Canada, and sociolinguistically oriented FSL teaching.

### *2.1. FSL Learners' Metasociolinguistic Knowledge*

Studies investigating FSL learners' metasociolinguistic knowledge have focused on exploring the effectiveness of providing learners with explicit instruction and/or engaging learners in metalinguistic reflection on the use of sociolinguistic variants. In such research, a clear distinction emerged between the knowledge of sociolinguistic variants and their actual use; in other words, between competence and performance [5–7]. The findings indicated that although explicit instruction was effective in helping learners develop competence, it did not automatically translate into performance. However, these studies stressed that the absence of sociolinguistic variation in the learners' second language (L2) production did not necessarily indicate a lack of sociolinguistic awareness [1,6]. They also suggested that with appropriate pedagogical support, newly acquired sociolinguistic knowledge can result in the productive use of sociolinguistic variants over time [1].

The findings of such studies further suggested that learners need to become aware of sociolinguistic variation before they can engage in the process of developing receptive and productive knowledge of it [8]. Additionally, they suggested that it is beneficial for learners to start acquiring sociolinguistic knowledge early to avoid needing to counteract the entrenched invariable use of sociolinguistic variants at later stages [6,8]. Finally, these studies suggested that it should be expected that learners will develop personal preferences in the use of sociolinguistic variants, whether reflecting the use of variants in their educational input, their personal L2 sociolinguistic goals or their overall L2-related goals, or other factors [6,7].

*2.2. Patterns of Learners' Sociolinguistic Variant Use*

In a volume synthesizing studies focused on the acquisition of sociolinguistic variants by FSL learners [9], four primary categories of variants were identified, namely vernacular variants (e.g., *'ien que*, *mas*, *nous-autres*, *fait que*, *job*, *rester*), informal variants (e.g., *juste*, *je vas*, *on*, *so*), formal variants (e.g., *seulement*, *je vais*, *nous*, *emploi*, *travail*, *habiter*, *alors*, *donc*), and hyper-formal variants (e.g., *ne . . . que*, *poste*, *demeurer*). These variants include the following: (i) expressions of restriction meaning 'only' (i.e., *'ien que/juste/seulement/ne . . . que*), for example, il (n')y a *'ien que/juste/seulement/que trois autos* (there are only three cars); (ii) forms of the first-person singular of the verb aller, meaning 'to go' (i.e., *mas/je vas/je vais*), for example, *mas/je va/je vais te dire quoi faire* (I am going to tell you what to do); (iii) first-person plural personal subject pronouns meaning 'we' (i.e., *nous-autres/on/nous*), for example, *nous-autres/on/nous sommes/est ici* (we are here); (iv) lexical expressions meaning 'job' (i.e., *job/emploi/travail/poste*), for example, *il a trouvé un(e) job/emploi/travail/poste* (he found a job); (v) lexical expressions meaning 'to live' (i.e., *rester/habiter/demeurer*), for example, *je reste/habite/demeure à Toronto* (I live in Toronto); and (vi) expressions of consequence meaning 'therefore' (i.e., *fait que/so/alors/donc*), for example, *il est tard fait que/so/alors/donc je vais me coucher* (it is late therefore I will go to bed). This synthesis revealed that classroom-instructed learners use vernacular and hyper-formal variants only marginally or not at all, informal variants less frequently than first-language (L1) speakers of French, and formal variants substantially more frequently than L1 speakers. One reason for these trends was the learners' educational input. The learners were approximating patterns that research has documented in the in-class speech of FSL teachers and in FSL instructional materials.

A number of factors that lead to learners' increased use of a wider array of differentially marked sociolinguistic variants have been identified in the literature. These include, in particular, increased extracurricular exposure to the authentic use of French in Francophone environments and increased long-term or targeted curricular exposure to the target language [9–14]. These factors also include increased engagement in learning French, which is measured by the frequency, intensity, intellectual demand, and immediacy of the use of French at present (e.g., current curricular and extracurricular exposure) and in the future (e.g., the intent to live and work in a Francophone environment) [15,16].

Research has also suggested that some sociolinguistic variants may be acquired more or less easily than others, as is the case with other linguistic features. For example, research focused on the acquisition of morphosyntactic features in French has revealed that important factors include the frequency with which a particular form naturally occurs in the language and the complexity of the rules governing its use [17]. Concerning the acquisition of sociolinguistic variants specifically, an additional factor can be a variant's socio-stylistic status. For example, some studies have found that both FSL learners and teachers avoided the use of vernacular and hyper-formal variants that were strongly socially marked but made highly frequent use of many formal variants and non-negligible use of certain informal variants that were less socio-stylistically marked [9]. However, other studies have found that the Canadian regional variant of lax-i was rare in the speech of FSL learners, even though the variant is used with high frequency by L1 speakers, has a fairly easy pronunciation, and is socio-stylistically neutral [18]. (In the study, a distinction is made between a tense-i, which is pronounced categorically in open syllables and stressed syllables closed by a lengthening consonant, as in the French verbs *vivre* (to live) and *rire* (to laugh), and a lax-i, which is pronounced in stressed syllables closed by other (non-lengthening) consonants, as in the French nouns *site* (area) and *ride* (wrinkle)). This suggests that the above combination of factors is not always sufficient to make a variant easy to acquire.

Although the nature of certain sociolinguistic variants can make their acquisition more difficult, research has shown that instruction can be effective in helping learners overcome such challenges. An important and necessary step toward the acquisition of sociolinguistic variation is the learners' awareness of its existence. Such awareness, especially if acquired through explicit, systematic, and recurrent instruction, can lead learners to develop control over sociolinguistic variant use [19]. However, research has also shown that this is not

a uniform process for all learners, even within the same learning conditions [12,20]. For example, one study investigated the acquisition of three morphosyntactic sociolinguistic variables (retention/deletion of *ne*, *nous/on*, and expression of futurity) and two phonological variables (/l/ retention/deletion, liaison) by five learners of French in a study-abroad context [20]. It documented individual differences among the learners, even though they were believed to be extremely similar both in their study-abroad experience and in their attitude, motivation, and desire to learn French.

### *2.3. The CEFR in FSL Teaching in Ontario, Canada*

The introduction of the CEFR in Canada, which is relatively recent in comparison to certain parts of the world [21], dates back to the recommendation of its use in 2006 to the Council of Ministers of Education, Canada, which officially endorsed its adoption in 2010 [22]. Initial interest in the framework is linked to one of the goals formulated by the Government of Canada in 2003 to double the proportion of English-French functionally bilingual high school graduates, with the framework serving as a primary assessment tool. Since then, the adoption of the CEFR has varied across Canada, where primary and secondary school curricula are regulated at the provincial level. In Ontario, the adoption of the framework has been aided by several province-wide movements.

One such movement concerns the development of a *Framework for FSL in Ontario Schools: Kindergarten to Grade 12* [23]. In this framework, the Ontario Ministry of Education placed the CEFR at the center of FSL education in the province after recognizing the framework as "a valuable asset for informing instruction and assessment practices" (p. 4). It envisions FSL learners as having "the confidence and ability to use French effectively in their daily lives" (p. 8) by calling for the implementation of the action-oriented approach, the use of "authentic, meaningful, interactive, and relevant tasks", and the prioritization of "the functional use of the language" (p. 18). This framework may be viewed as part of a larger province-wide movement related to the CEFR, which involves offering FSL teachers numerous professional learning opportunities designed to familiarize them with the CEFR. These opportunities include, for example, provincial CEFR-focused meetings and web conferences, regional learning events, school board conferences and workshops, self-directed and job-embedded learning, and *Diplôme d'études en langue française* (DELF) assessor training. The *Diplôme d'études en langue française* (DELF) is a CEFR-aligned exam of FSL proficiency offered at four CEFR levels: A1 and A2 (basic users), and B1 and B2 (independent users). The exam tests skills in four major areas: oral comprehension and production, and written comprehension and production. The diplomas are issued by the French Ministry for National Education. Assessors must receive specific training before they are qualified to score the DELF exam.

These types of professional learning opportunities helped to address teachers' initial confusion about the CEFR, which they saw as a positive though challenging addition to FSL education in the province. For example, one study of Ontario teachers' perceptions of CEFR-informed instruction conducted prior to the development of the new FSL framework in 2013 found that the teachers saw many benefits of the new approach [24]. However, they also felt that to be able to orient their teaching to the framework successfully, they needed to develop a thorough understanding of its principles and identify practical ways for applying these principles in their classroom practice. That same year, another study reported that one effective way for teachers to address these types of challenges was through professional development opportunities, such as teachers working together in groups designed as professional learning communities [25]. In this study, a group of L2 teachers in the Canadian province of New Brunswick engaged in formal and informal meetings about the CEFR. Over the course of their engagement, the teachers identified what the framework was asking them to do and how, created CEFR-informed instructional materials, and designed a plan for their implementation.

Another CEFR-related movement in Ontario has been to encourage graduating Grade 12 students to take the DELF. A study of Grade 12 students' DELF results showed that their

areas of greatest strength were written comprehension skills and, within their productive skills, the ability to follow instructions and provide information [26]. In contrast, oral comprehension skills and written and oral production skills, particularly their application of grammar and vocabulary in context, were identified as the areas in greatest need of improvement. The study explored the link between the students' test results and their confidence in their French skills and found that the students were generally more confident about their receptive than their productive skills, more confident about their written than their oral skills, and, specifically, that they felt most confident about their reading skills and least confident about their conversational skills. The results thus showed that speaking skills were an area that the students needed to improve in terms of both proficiency and confidence.

The push toward the DELF in Ontario has meant that the teachers who undertake the training to serve as DELF assessors develop a solid understanding of the CEFR. For example, the majority of teachers in one study reported that the listening, speaking, reading, and writing tasks they use in their classrooms were similar to those found on the DELF test [27]. The teachers' comments further revealed that they believed their understanding of the DELF and the CEFR has made them focus more on helping their students develop oral comprehension skills and use more activities that are oral, interactive, and require the use of critical thinking skills. In terms of assessment, the teachers believed that knowledge of the test's benchmarks helped them understand more clearly what their students were expected to be able to do in French at various levels of proficiency.

Finally, learners, for their part, have also been found to positively view the type of instruction that the CEFR promotes. A study exploring Grade 12 students' preferences in L2 learning found that the students prioritized as a goal the ability to use the target language in the real world [28]. They believed that L2 instruction should be focused on helping them develop skills that are practical and applicable outside of the classroom. They reported that one way to bring the real world into the classroom was to use authentic L2 materials, such as music videos, news, or interviews, which they saw as useful in exposing them to different varieties of the language and to 'slang'.

### *2.4. Sociolinguistically-Oriented FSL Teaching*

Studies exploring effective approaches to the development of classroom-based FSL learners' sociolinguistic competence have suggested that explicit instruction is necessary and is most effective when integrated in the curriculum so that its delivery is systematic and recurrent and follows the awareness-practice-feedback sequence [14,19]. There is evidence to suggest that such explicit instruction is possible and desirable even with lower-proficiency learners [29]. It does not appear to be overwhelming for new learners and has the advantage of preventing learners from developing habits of invariant use that may be difficult to counteract later [30]. Studies also show that the effectiveness of explicit instruction is increased, first, if the focus is placed primarily on teaching learners to understand sociolinguistic variation "conceptually" rather than on acquiring rigid rules for the use of specific sociolinguistic variants [31]. Second, it is increased when learners are allowed to develop a personal stance toward the sociolinguistic concepts they are acquiring. Such an approach allows learners, on the one hand, to apply their sociolinguistic knowledge (such as understanding the link between formality and social distance) more broadly and, on the other hand, to enact their sociolinguistic agency by making choices that reflect their current personal or social identity. Concerning the use of authentic materials, the use of films and the process of scriptwriting have been proposed, as well as more recently the use of television series and subtitles [32]. These allow learners to observe a character's choice of sociolinguistic variants in interactions that differ in terms of their conversation partner, setting, and communication purpose—importantly, also at various stages of relationship development.

### *2.5. Links to the Present Study*

As we have seen, the CEFR calls for the analysis of learners' needs with the aim of helping them develop the ability to act in real-life situations. This ability relies on their

capacity to enact their sociolinguistic agency. Supporting learners in this endeavor can be accomplished through pedagogical approaches that prioritize the use of authentic materials and communicative tasks modeled on real-world interactions. For this reason, the present article, as mentioned, draws on learner and teacher data to explore the extent to which the learners' self-identified sociolinguistic needs align with the teachers' self-reported reoriented practices following CEFR-oriented professional learning.

## 3. Methods

The present study draws on datasets collected as part of two larger projects. The first project examined the sociolinguistic knowledge base of Ontario FSL learners at the university level. Using data from university-level learners for the present study offers the advantage of students who can provide a retrospective perspective on their entire FSL journey, from Kindergarten to Grade 12, and beyond. The second project examined the impact of CEFR-informed professional learning on the pedagogical practice of Ontario FSL teachers from Kindergarten to Grade 12. Details of these two datasets are presented below.

### *3.1. Learners*

The learner data referenced in this article and summarized in Table 1 were collected in 2012 from 44 undergraduate students in years one through five at two Canadian universities—26 at an English-medium institution and 18 at an English/French bilingual one. The 44 students represent a non-probabilistic convenience sample of students who were enrolled in undergraduate FSL courses, either as part of a French Major or French Specialist program, and had studied French at elementary or secondary schools in Ontario in French immersion programs or non-immersion programs.

**Table 1.** Distribution of FSL Learners.

| Categories | | n | % |
|---|---|---|---|
| University | English-Medium | 26 | 59 |
| | Bilingual | 18 | 41 |
| Year of Study | 1st | 5 | 11 |
| | 2nd | 11 | 25 |
| | 3rd | 10 | 23 |
| | 4th | 12 | 27 |
| | 5th | 6 | 14 |
| Elementary/Secondary Schooling | Non-Immersion | 14 | 32 |
| | Immersion | 30 | 68 |

Two data collection instruments were used: an English interview and a French interview. The English interview that the students participated in was an ad hoc semi-directed interview, in which they answered questions about what they know about levels of formality in French, how they learned about such (in)formality, what else they would like to know about this topic, and how their FSL courses could better develop their sociolinguistic abilities. The English interview also asked the following specific questions:

1. What do you know about sociolinguistic variation in French (including any specific sociolinguistic variants you know, and how you acquired them)?
2. To what extent do you feel able to perceive the identity and intentions of others in French?
3. To what extent do you feel able to express your own identity and intentions in French?

The French interview that the students participated in was a semi-directed interview in French. Following the standard methodology of previous sociolinguistic research [9,10,13], the present study used this French interview to examine the students' use of sociolinguistic variants. The list of interview questions was drawn from previous sociolinguistic research [9]. These questions broached a variety of formal topics, such as the importance of religion in their life, their views on political issues, and their views about different

regional varieties of French. They also addressed a variety of informal topics, such as what television programs they enjoy watching, a funny trick they played on a teacher, and plans for summer vacation.

### 3.2. Teachers

The teacher data referenced in this article were collected in 2017. Thirty-six Ontario school boards were each asked to invite five teachers to participate in a study exploring the practices of teachers who had participated in CEFR-related professional development opportunities. The participating teachers from this non-probabilistic sample are, thus, likely to be those who are the most positively oriented to the CEFR and who had made the greatest efforts to align their teaching practices with the framework. A total of 103 teachers responded to this call. The teachers taught in Core French programs (i.e., where French is the subject of study), French Immersion programs (i.e., where French is the medium of instruction for other subject areas), and/or Extended French programs (i.e., a type of intensive French program akin to delayed-immersion and preceded by Core French courses). As Table 2 shows, they are elementary and secondary school teachers teaching FSL classes at multiple grade levels from Kindergarten (students aged three to five years old) to Grade 12 (students aged 17 or 18 years old).

**Table 2.** Grades Taught by FSL Teachers.

| Grades Taught | % of Teachers |
|---|---|
| 12 | 42 |
| 11 | 43 |
| 10 | 42 |
| 9 | 51 |
| 8 | 20 |
| 7 | 22 |
| 6 | 13 |
| 5 | 13 |
| 4 | 12 |
| 3 | 12 |
| 2 | 8 |
| 1 | 8 |
| Kindergarten | 6 |

As can be seen in Figure 1, the teachers were fairly evenly distributed based on their number of years of FSL teaching experience, except for a dearth of novice teachers.

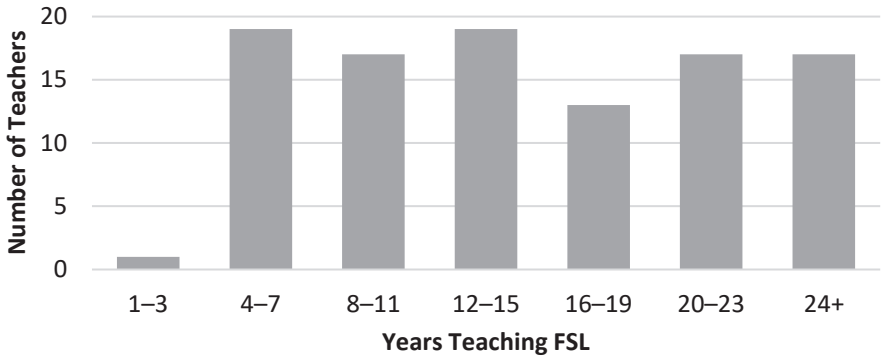



**Figure 1.** FSL Teachers' Years of Teaching Experience.

Figure 2 shows that the majority of the teachers reported having participated in CEFR-related professional learning for four years or more, with approximately half reporting four to five years of such training.


**Figure 2.** FSL Teachers' Years of CEFR-Related Professional Learning.

As shown in Table 3, all of the teachers had completed DELF corrector training and nearly all had attended CEFR-related conferences and/or workshops hosted by their schools or their Boards as part of their array of CEFR-related professional learning experiences.

**Table 3.** FSL Teachers' Professional Learning Opportunities.

| Professional Learning Opportunities | % of Teachers |
|---|---|
| DELF corrector training | 100 |
| School/board conferences, workshops | 93 |
| DELF corrector "refresher" | 77 |
| Job-embedded professional learning | 56 |
| Provincial Web-conferences | 55 |
| CEFR regional learning events | 54 |
| Self-directed | 51 |
| Other conferences or workshops | 50 |
| CEFR provincial meetings | 49 |
| Coaching/mentoring | 39 |

Each of the 103 teachers completed a two-part online survey. The first part asked them to provide information about their teaching experience (e.g., grades taught and number of years of teaching experience) and their CEFR-related professional learning (e.g., the number of years and types of CEFR-related professional learning). The second part of the survey asked them to provide self-reports about their teaching practices before (i.e., retrospectively) versus after (i.e., currently) their CEFR-related professional learning. The survey tapped their practices concerning planning, classroom delivery, and assessment and evaluation. With respect to their planning practices, the teachers were asked to comment on how they approached developing their students' proficiency and what proportion of time they devoted to each of the four basic skills. They were also asked about what aspects of the CEFR they considered most important, how their experience scoring the DELF exam improved their understanding of the CEFR and impacted their FSL planning, and what changes they had made to their FSL teaching resources. In regard to their classroom delivery, the teachers were invited to reflect on the extent to which they used a variety of traditional practices (e.g., immediate error correction, a focus on forms) versus CEFR-informed practices (e.g., activities related to everyday life, student-driven needs analysis). They were also asked about the balance they placed across linguistic, sociolinguistic, and pragmatic competences; how they presented language in their classrooms; and the effective activities they used to teach grammar and vocabulary and encourage authentic and spontaneous student-to-student interaction. This section of the survey ended with a question asking the teachers to identify a change in their practice that they felt had the most impact on increasing their students' proficiency. Finally, with respect to their assessment and evaluation practices, the teachers were asked a series of questions about whether learning goals, success criteria, and feedback prioritized form versus function, and about the aspects targeted in their feedback. The teachers were also asked what proportion of their summative evaluations they devoted to each of the four skills and which change

in their assessment practices they believed had the greatest impact on increasing their students' proficiency.

### 3.3. Data Analysis

This study adopts a qualitative approach to the analysis of the data, allowing for the emergence of themes or categories within the students' and teachers' responses, as reported in the results section. This approach is supplemented, where appropriate, with descriptive statistics in the form of raw numbers and percentages. Since the goal of this qualitative study is to compare the students' self-identified needs and the teachers' self-reported practices, rather than identifying differences across subgroups within each dataset, a statistical analysis has not been pursued.

## 4. Results

The results are presented below in response to the three research questions guiding the present study. First, the results related to the learners' self-identified sociolinguistic abilities and needs are reported. Then, the results related to the teachers' self-reports about their CEFR-informed teaching practices are presented. Finally, the extent of the match between the learners' self-identified sociolinguistic needs and the teachers' reports of their re-oriented practice is discussed.

### 4.1. Research Question #1: Learners' Self-Identified Sociolinguistic Abilities and Needs

#### 4.1.1. Ability to Express One's Own Identity and Intentions and Perceive That of Other French Speakers

As Table 4 shows, despite having studied French between 10 and 18 years at the elementary and secondary school level, only half (50%) of the students reported feeling as though they had developed both the ability to express their own identity and intentions in French and the ability to perceive the identity and intentions of other French speakers. Less than half (42%) of the students reported having developed only the receptive ability, and 8% reported not having developed either of the two abilities. The students who felt they had developed both abilities pointed to their capacity to discern other speakers' attitudes and highlighted their ability to express their personality through their linguistic choices. The students who had developed only the receptive ability reported that when it came to their expressive ability, they had to prioritize the communication of their message over that of their personality. Finally, the students who felt they had not developed either ability commented explicitly about not being familiar with vocabulary that shows a speaker's personality and/or reported that they needed to devote their attention primarily to ensuring their ideas were understood.

**Table 4.** Distribution of Learners by Self-reported Ability to Express and Perceive Identity and Intentions in French.

| Abilities Reported | Sample Comments |
| --- | --- |
| Both Abilities<br>50% | "I can understand when people are being more serious or when they're being more laid back".<br>"I feel like I can express myself . . . I feel like for the most part they can pick up on my personality". |
| Receptive Only<br>42% | "Listening, I find I don't have an issue with listening to people and understanding them um, and even understanding their personality [ . . . but . . . ]"<br>"I'm sort of thinking so far in advance about what I'm going to say that I don't get to throw in any of myself into it". |
| Neither Ability<br>8% | "I don't know the words that are used to express someone's personality or how to talk for yourself".<br>"Because you're stressing about making sure the person understands what you're trying to say, you can't really show your personality". |

#### 4.1.2. Elements Learners Report Missing to Develop Their Expressive and/or Receptive Sociolinguistic Ability

Table 5 summarizes the missing elements that the learners identified as obstacles to being able to develop their expressive and/or receptive sociolinguistic abilities in French (i.e., the ability to express their own identity and intentions and perceive that of other French speakers). Their responses revealed that in addition to a lack of confidence, they believed a major issue was their lack of knowledge of stylistically varied vocabulary, particularly with regard to the very formal and very informal registers, including an understanding of their appropriate and nuanced use according to the context.

**Table 5.** Learner-Identified Missing Elements Preventing Development of Expressive and Perceptive Sociolinguistic Ability.

| Missing Elements | Sample Comments |
| --- | --- |
| Confidence and vocabulary in action | "Confidence, definitely vocab, always vocab. I will have learned a word for four years and I will not know how to use it".<br>"I think sometimes it's just like the confidence to speak in French and other times like I know the vocab, when I hear it I understand it, but actually using it . . . " |
| Experience and nuanced expression | "I think it's experience with the language. I think it's understanding the nuances within the language like the registers, like idiomatic expressions or sayings".<br>"If I wanted to be super informal and be super cool you know with all my friends, I don't think I could do that". |
| Understanding of how language changes with formality | "I would find it beneficial to learn about French formality and the various different registers that can be used to express oneself. I find that this is a component that is overlooked in French teaching and would aid students".<br>"I wish I learned a lot of [variants] that I can use in regards to the different social settings I will be exposed to". |
| Ability to use the lower and upper registers and their markers | "I would like to know more about very formal language that goes deeper than just using *vous*". *<br>"Colloquial, slang. Mostly just when to use it and where it comes from". |

* *Vous* is a first-person plural personal pronoun ('you') whose use can indicate (i) more than one addressee or (ii) a single addressee in formal social situations and/or situations characterized by greater social distance.

#### 4.1.3. Learners' Perceptions of Their Ability to Use Upper and Lower Register Markers

The learners' frequency of use of vernacular, informal, formal, and hyper-formal variants during their French-language interview was measured according to their self-reported French proficiency level. Six sociolinguistic variables were examined:

1. Expressions of restriction meaning "only" (i.e., *'ien que/juste/seulement/ne . . . que*);
2. First person singular of the verb *aller* "to go" (i.e., *mas/je vas/je vais*);
3. First person plural personal subject pronouns (i.e., *nous-autres/on/nous*);
4. Lexical expressions for a "job" (i.e., *job/emploi/travail/poste*);
5. Lexical expressions for "to live" (i.e., *rester/habiter/demeurer*);
6. Expressions of consequence meaning "therefore" (i.e., *fait que/so/alors/donc*).

The students' French proficiency was arrived at using their self-assessments of how well they felt they could speak, listen, read, and write in French using a five-point scale for each ability (one = not at all, two = a little, three = fairly well, four = very well, five = fluently). Each student's reported fluency for the four language abilities were combined to provide a total out of 20 points, and students were regrouped into the following proficiency levels: low (4–8 points), mid (9–15 points), or high (16–20 points).

As Table 6 shows, the students predominantly used the formal variants regardless of their self-reported level of proficiency (53–64%). The students who rated their French proficiency the highest made the greatest use of informal variants (45% versus 37/35%). As for the hyper-formal and vernacular variants, the students made nil to marginal use of these regardless of their proficiency level. These results are consistent with the students'

self-reported lack of knowledge of different registers in French, particularly with regard to very formal and very informal vocabulary in French (see Table 5).

**Table 6.** Learners' Use of Sociolinguistic Variants According to their Self-Rated French Proficiency.

| Self-Rated Proficiency | Vernacular Variants (i.e., *'ien que; mas; nous-autres; fait que; job; rester*) | Informal Variants (i.e., *juste; je vas; on; so*) | Formal Variants (i.e., *seulement; je vais, nous; emploi; travail; habiter; donc, alors*) | Hyper-Formal Variants (i.e., *ne . . . que; poste; demeurer*) |
|---|---|---|---|---|
| Low | 1% | 37% | 62% | 0% |
| Mid | 1% | 35% | 64% | 0.2% |
| High | 2% | 45% | 53% | 0.2% |

#### 4.1.4. Learners' Preferred Ways of Developing Sociolinguistic Abilities

Table 7 summarizes what the learners reported about how they would prefer to develop their sociolinguistic abilities in French. Their responses show that they would welcome more of a sociolinguistic focus in their French classes, particularly to practice speaking and listening at both the high school and university levels. They also pointed to a desire for supervised conversations and speaking activities that would encourage them to speak both within and outside of class. The use of authentic interactions and authentic materials was also high on their list of desired ways to develop their sociolinguistic abilities. For instance, students identified interacting with French speakers and watching movies with no subtitles as important means to improve this aspect of the target language. Finally, students expressed a desire to access more exchange programs and immersive experiences as a way to improve their sociolinguistic abilities, noting that there is "no substitute" for out-of-class exposure.

**Table 7.** Learner-Identified Preferred Ways of Developing Sociolinguistic Abilities.

| Learners' Preferences | Sample Comments |
|---|---|
| French classes | "There could be more courses focused on how to speak in a given context".<br>"In a classroom setting, directed by a teacher (with opportunities for practice) would be the best way to learn, listen and practice".<br>"It would be fantastic to learn this in courses in both high school and university". |
| Supervised conversations and speaking activities | "There's no supervised speaking ever. We are not encouraged to talk in class."<br>"I'd like to say like oral classes/conversation". |
| Authentic interactions | "I would like to learn that in interactions with others".<br>"I would love to learn this by speaking to French speakers!" |
| Authentic materials | "Watching movies with no subtitles or listening to music or magazines".<br>"Through guest speakers or watching movies". |
| Exchange programs and immersive experiences | "I am going on an exchange to France so maybe I can learn there".<br>"I find that there is no substitute for practice experience, authentic immersion in a cultural milieu". |

In sum, the learners' self-assessments of their sociolinguistic competence revealed that only half of them believed they had developed the ability to both express their own identity and intentions and perceive that of other French speakers. Students also believed that they lacked knowledge of very formal and very informal vocabulary in French, which was confirmed by their preference for the use of formal sociolinguistic variants in their speech. To address the gap in their knowledge of stylistically-varied vocabulary and to improve their sociolinguistic skills in general, the students expressed a preference for an

increased amount of supervised conversation and speaking activities, the use of authentic instructional materials, and opportunities to engage in authentic interactions. To gain a sense of how well these learners' self-identified needs can be met by instructional practices informed by CEFR principles, in the following section we explore the importance of sociolinguistic competence in the teachers' self-reports of their CEFR-informed approach to planning, classroom practice, and assessment and evaluation.

### *4.2. Research Question #2: Teachers' CEFR-Informed Instructional Practices*

#### 4.2.1. Teachers' Self-Reported CEFR-Informed Approach to Planning

As Figure 3 illustrates, the teachers reported that their CEFR-related professional learning has encouraged them to plan for instruction that places a more balanced focus on sociolinguistic, pragmatic, and linguistic competence using authentic communicative situations presented in tasks that are action-oriented and individualized. Action-oriented tasks are authentic, open-ended, and involve meaningful interaction. For example, one teacher reported that they use tasks that "incorporate open-ended situations where [the students] have to give their opinions. When an issue has a personal connection to the students, they want to share their ideas and thoughts on the matter". This more balanced focus on all three major competences replaced the teachers' reported previous focus on primarily developing their students' linguistic competence, especially through the prioritization of written skills and grammatical forms. For example, one teacher reported, "I am less strict with certain structures and focus more on [the students'] communicative ability".

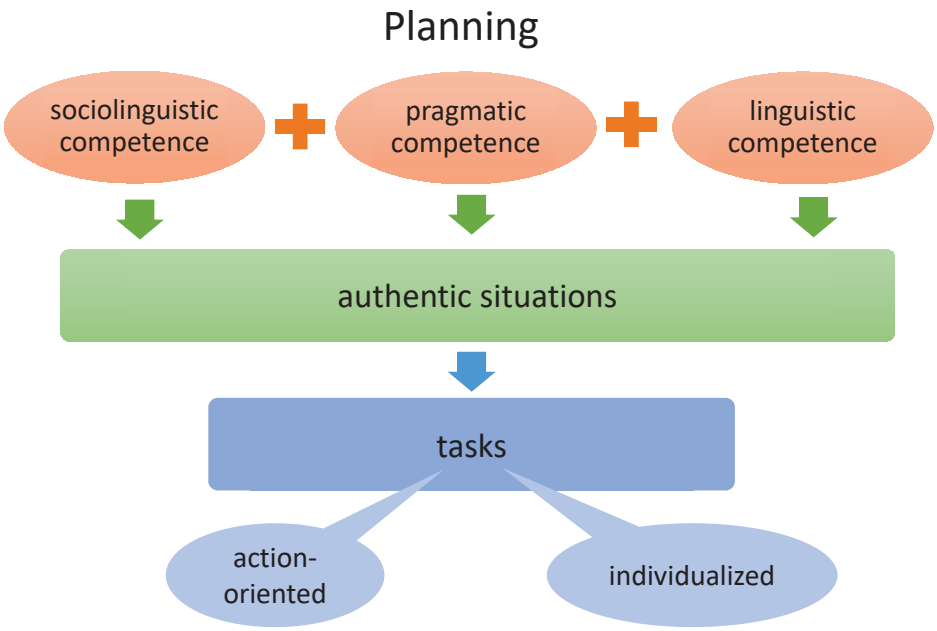



**Figure 3.** Teachers' Self-Reported CEFR-oriented Approach to Planning.

#### 4.2.2. Teachers' Self-Reported CEFR-Informed Approach to Teaching

As summarized in Figure 4, the teachers reported that in their teaching, their CEFR-related professional learning inspired them to present language in relation to speech acts and on-demand by focusing on real-life, everyday uses of French. This provides their students with more opportunities to practice using the language in tasks that are sociolinguistically relevant (e.g., role-plays, structured and unstructured discussions), and to make use of online resources, such as a variety of authentic documents that exemplify stylistic varieties. For example, one teacher reported, "Going shopping in a store—this role-playing activity is a good time to review vocabulary associated with clothing, sizing, money, conditional tense (polite requests), asking questions". Another teacher reported, "Opinion sharing in small group situations with little or no preparation—however, they do have access to a guide-sheet

with specific sociolinguistic structures of focus". This new approach to introducing language situated in real-life contexts replaced the teachers' reported previous approach to presenting language that was theme-based, isolated, and disconnected.

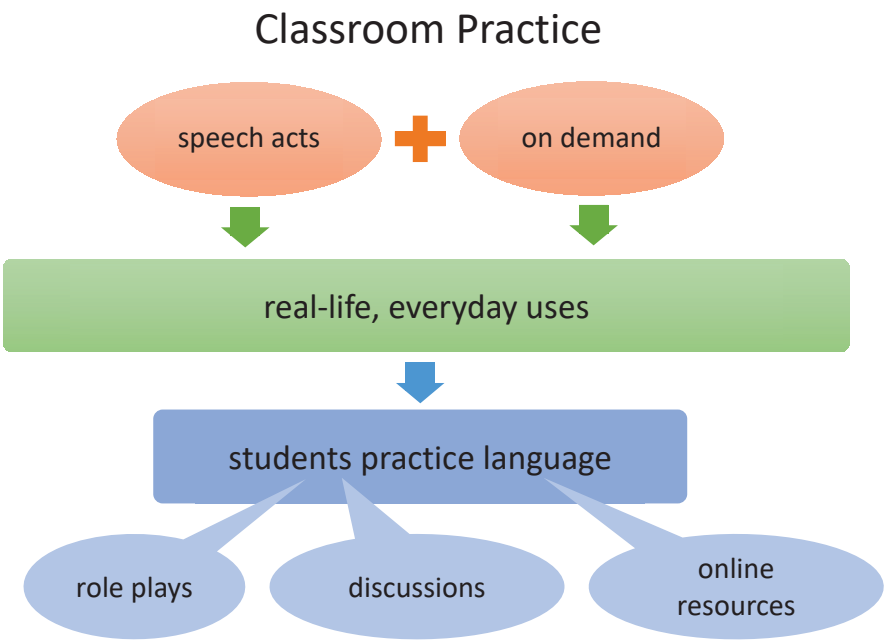



**Figure 4.** Teachers' Self-Reported CEFR-oriented Approach to Teaching.

4.2.3. Teachers' Self-Reported CEFR-Informed Approach to Assessment and Evaluation

As can be seen in Figure 5, the teachers reported that their CEFR-related professional learning led them to focus their assessment and evaluation on their students' ability to produce and understand communication in French and on the quality of their use of the language. They reported that this new approach is reflected in their attention to sociolinguistic appropriateness, pragmatic appropriateness, and functional competence, as well as in their use of a variety of assessment and evaluation strategies. These include, among others, assigning speaking a more prominent role, using a variety of test formats, and creating authentic oral exams and other authentic assessment and evaluation tasks. For example, one teacher reported that their new approach consisted of "carefully designing authentic tasks and selecting specific success criteria". Concerning speaking, one teacher not only reported a change in their approach, but also commented on one of the benefits that they perceived in this new approach: "Making speaking worth more has not only increased the students' marks, but also their confidence". This benefit corresponds directly to a lack of confidence, which was one of the learner-identified obstacles in relation to their development of sociolinguistic abilities (see Table 5). The teachers' new expanded and varied approach to assessment and evaluation replaced their previous reported attention to orthographic control, grammatical accuracy, and phonological control.

In sum, the teachers' self-reports of the place of sociolinguistic competence in their CEFR-informed approach to planning, classroom practice, and assessment and evaluation revealed that they prioritize helping their students develop the ability to communicate in French by focusing not only on developing their linguistic competence, but also their ability to express themselves in sociolinguistically and pragmatically appropriate ways. To this end, the teachers reported increasing the prominence of speaking skills, working with authentic documents that expose their students to real-life varieties of French, and using activities that allow their students to practice communicating in authentic interactions.

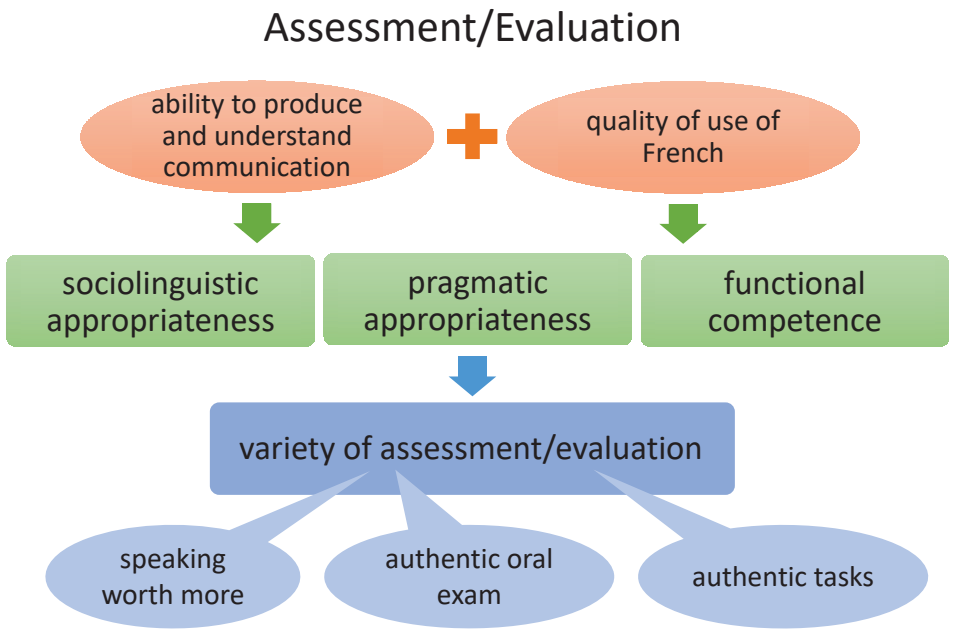



**Figure 5.** Teachers' Self-Reported CEFR-oriented Approach to Assessment and Evaluation.

### *4.3. Research Question #3: Overlap between Learners' Sociolinguistic Needs and Teachers' CEFR Practices*

In this section, we explore how well these teachers' CEFR-informed instructional practices meet the learners' self-identified sociolinguistic needs. Table 8 summarizes both the "what" and "how" of the learners' self-identified sociolinguistic needs and the teachers' self-reported CEFR-informed instructional practices. As can be seen, the overlap between the learners' needs and the teachers' reoriented practices is considerable. In terms of the "what", the learner-identified need to develop sociolinguistic confidence is matched by the teachers' reported attention to develop their students' ability to understand and engage in communication in French through a focus on students' communication needs and via speech acts. Similarly, the students' desire to learn how to express themselves (in)formally in contextually appropriate ways through the use of variants ranging from vernacular variants, slang, and colloquial expressions to hyper-formal variants, is matched by the teachers' reported re-balancing of their focus to include sociolinguistic and pragmatic competences, alongside linguistic competence. In these ways, both groups are oriented to developing students' sociolinguistic competence, including various abilities and skills, and particularly the ability to use and understand the full scale of sociolinguistic variants in French.

With respect to the "how", the learner-identified preferred ways of developing their sociolinguistic competence and the students' preferred instructional approach, resources, and methods focus on a call for increased class time spent on oral production using authentic materials and for additional opportunities for real-world use of the language. These preferred means appear to match remarkably well with the teachers' reported new instructional practices, which include the prioritizing of speaking, making use of guided conversations, and drawing on authentic materials and authentic action-oriented tasks, including role-plays of real-world interactions. As such, the students and teachers are both focused on developing conversational skills by engaging in guided class discussions, using authentic materials, and completing communicative tasks set in a variety of contexts with a variety of real-life, everyday, authentic purposes.

This substantial extent of "fit" between the learners' self-identified sociolinguistic needs and the teachers' focus on sociolinguistic competence in their self-reported CEFR-informed approach to teaching is not surprising. This is because one of the central aims of the CEFR is to make sure that L2 teaching leads students to transition successfully from L2

learners into autonomous L2 speakers as "social agents", which makes the framework well positioned to help learners develop their sociolinguistic competence.

**Table 8.** Extent of Match between Learners' Sociolinguistic Needs and Teachers' CEFR-Informed Practices.

| | Learners' Self-Identified Sociolinguistic Needs | Teachers' Self-Reported Approach to Teaching | Overlap |
|---|---|---|---|
| What | • sociolinguistic confidence<br>• expression of (in)formality and use of context-appropriate sociolinguistic variants<br>• very formal variants and very informal variants (including slang and colloquialisms) | • ability to understand and engage in communication<br>• students' communication needs and speech acts<br>• balanced development of sociolinguistic, pragmatic, and linguistic competence | • sociolinguistic knowledge, abilities, and skills<br>• vernacular to hyper-formal variants |
| How | • increased French class opportunities<br>• supervised conversation and speaking activities<br>• authentic materials, such as movies, music, and magazines<br>• exchange programs | • prioritizing of speaking<br>• guided conversations<br>• authentic materials<br>• authentic, action-oriented tasks<br>• role-plays of real-life interactions | • increased attention to conversational skills in guided class discussions<br>• use of authentic materials<br>• communicative tasks set in a variety of contexts with a variety of real-life, everyday, authentic purposes |

## 5. Discussion

As we have seen, only half of the FSL learners in the present study reported having developed both the sociolinguistic ability to express their own identity and intentions, and perceive the identity and intentions of other French speakers. Nevertheless, their interview responses revealed that they have developed sufficient sociolinguistic awareness to accurately assess their sociolinguistic competence. Their comments pertaining to the gaps they perceive specifically in their knowledge of sociolinguistic variants further suggest that they are aware of the full range of sociolinguistic registers available in French—and such awareness is a necessary prerequisite for the development of receptive and productive knowledge of sociolinguistic variation [8]. In addition, the students' identification of gaps regarding their knowledge of the upper and lower register markers (i.e., hyper-formal and vernacular sociolinguistic variants) is accurate in that it matches the patterns documented in their productive use of sociolinguistic variants, which featured the use primarily of formal and, to a lesser extent, informal sociolinguistic variants. This overuse of formal sociolinguistic variants is reminiscent of the findings of past studies [9] and may well reflect their educational input, their L2 sociolinguistic goals or overall L2-related goals, or other factors [6,7]. Regarding the learners' identification of what is missing in order for them to develop sociolinguistic competence, it is interesting to note that their calls for increased exposure to (authentic) use of French match the findings of previous research that has found it to be an effective way to develop such competence [10–12,17]. This finding suggests that learners have a good idea of what it is they need to learn and how to do it successfully.

The FSL teachers' self-reported re-oriented planning, teaching, and assessment and evaluation practices provide evidence that CEFR-related professional development opportunities are an effective way for FSL teachers to become familiar with the framework and adopt its approach in their teaching, as has been suggested in previous research [25]. The teachers made it clear that their instructional practices have changed to reflect the CEFR-inspired call for being attentive to the learners' needs. Their practices have also changed to reflect their new focus on teaching learners not about the language, but how to use the language—no longer only in terms of linguistic accuracy (i.e., vocabulary and grammar), but also in sociolinguistically and pragmatically appropriate ways. The teachers reported that this new approach is reflected in their focus on authentic and interactive activities, their increased use of authentic materials, and their increased attention to speaking skills. Importantly, the use of authentic instructional materials [32], paired with explicit instruction, including a conceptual understanding of sociolinguistic variation (which allows learners

to make situation-specific decisions in relation to various contextual factors rather than follow rigid rules), has not only been proposed but also found to be an effective way to help learners develop sociolinguistic competence [19,29].

Analysis of the FSL learners' self-reported sociolinguistic needs and the FSL teachers' self-reported CEFR-informed instructional practices in the present study revealed a substantial amount of overlap. This overlap was in terms of content (i.e., what the learners report they need in order to develop their sociolinguistic competence and what the teachers report as their instructional focus) and method (i.e., how the learners would prefer to acquire their sociolinguistic knowledge, abilities, and skills, and the types of materials, activities, and tasks that the teachers report prioritizing). An obvious limitation of this study, however, is that it relies on the self-reports of a group of highly motivated teachers rather than on direct observations of their instructional practices or the instructional practices of teachers who have yet to embrace this new framework. As such, the data from the teachers may be best understood as a reflection of their intentions to change rather than actual changes to their practices as a result of the CEFR-informed professional learning. However, the dedication of the teachers in this study to adopt the CEFR's new approach to teaching students the ability to communicate in socially appropriate ways in a variety of contexts appears to be strong. Because the framework envisions language teaching as leading learners to become confident, legitimate, and autonomous speakers of the target language, the development of sociolinguistic competence is both its implicit and explicit aim. If the teachers act on their intentions to focus on their students' development of sociolinguistic competence through the use of action-oriented, real-life-based, authentic tasks and instructional materials, alongside explicit instruction on sociolinguistic variation, then the teachers' ongoing intensive and extensive CEFR-informed professional learning is well positioned to help teachers meet the sociolinguistic needs of Ontario's FSL students.

**Author Contributions:** Conceptualization, K.R. and I.L.; writing—original draft preparation, I.L. and K.R.; writing—review and editing, K.R. and I.L.; funding acquisition, K.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Social Sciences and Humanities Research Council (SSHRC) (for student data) and by the Ontario Ministry of Education and l'Association canadienne des professionnels de l'immersion (ACPI) (for teacher data).

**Institutional Review Board Statement:** The study was conducted in accordance with the University of Toronto's policies on Research Ethics for studies involving human participants.

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

**Data Availability Statement:** Data from this study is unavailable due to privacy or ethical restrictions.

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

## References

1. van Compernolle, R.A.; Williams, L. Reconceptualizing Sociolinguistic Competence as Mediated Action: Identity, Meaning-Making, Agency. *Mod. Lang. J.* **2012**, *96*, 234–250. [CrossRef]
2. Lasan, I.; Rehner, K. Expressing and perceiving identity and intentions in a second language: A preliminary exploratory study of the effect of (extra)curricular contact on sociolinguistic development. *Int. J. Biling. Educ. Biling.* **2018**, *21*, 632–646. [CrossRef]
3. Canale, M.; Swain, M. Theoretical bases of communicative approaches to second language teaching and testing. *Appl. Linguist.* **1980**, *1*, 1–47. [CrossRef]
4. Council of Europe Common European Framework of Reference: Learning, Teaching, Assessment. Companion Volume. Available online: https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4 (accessed on 18 September 2022).
5. van Compernolle, R.A. Constructing a Second Language Sociolinguistic Repertoire: A Sociocultural Usage-based Perspective. *Appl. Linguist.* **2019**, *40*, 871–893. [CrossRef]
6. van Compernolle, R.A.; Williams, L. Metalinguistic explanations and self-reports as triangulation data for interpreting second language sociolinguistic performance. *Int. J. Appl. Linguist.* **2011**, *21*, 26–50. [CrossRef]
7. Yang, J.; Rehner, K. Learner beliefs about sociolinguistic competence: A qualitative case study of four university second language learners. *Lang. Learn. High. Educ.* **2015**, *5*, 157–180. [CrossRef]

8. Beaulieu, S.; Woll, N.; French, L.M.; Duchemin, M. Language learners' metasociolinguistic reflections: A window into developing sociolinguistic repertoires. *System* **2018**, *76*, 210–218. [CrossRef]
9. Mougeon, R.; Nadasdi, T.; Rehner, K. *The Sociolinguistic Competence of Immersion Students*; Multilingual Matters: Bristol, UK, 2010. [CrossRef]
10. Blondeau, H.; Lemée, I. Future Temporal Reference in L2 French Spoken in the Laurentian Region: Contrasting Naturalistic and Instructed Context. *Can. Mod. Lang. Rev.* **2020**, *76*, 70–90. [CrossRef]
11. Donaldson, B. Aspects of interrogative use in near-native French: Form, function, and register. Linguistic Approaches to Bilingualism. *Linguist. Approaches Biling.* **2016**, *6*, 467–503. [CrossRef]
12. Donaldson, B. Negation in Near-Native French: Variation and Sociolinguistic Competence. *Lang. Learn.* **2016**, *67*, 141–170. [CrossRef]
13. Regan, V.; Howard, M.; Lemée, I. *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*; Multilingual Matters: Bristol, UK, 2009. [CrossRef]
14. French, L.M.; Beaulieu, S. Effects of sociolinguistic awareness on French L2 learners' planned and unplanned oral production of stylistic variation. *Lang. Aware.* **2016**, *25*, 55–71. [CrossRef]
15. Mougeon, F.; Rehner, K. Engagement portraits and (socio)linguistic performance: A transversal and longitudinal study of advanced L2 learners. *Stud. Second. Lang. Acquis.* **2015**, *37*, 425–456. [CrossRef]
16. Ruivivar, J. Best Grad Competition: Engagement, Social Networks, and the Sociolinguistic Performance of Quebec French Learners. *Can. Mod. Lang. Rev.* **2020**, *76*, 243–264. [CrossRef]
17. Rehner, K.; Mougeon, R.; Mougeon, F. Variation in choice of prepositions with place names on the French L1–L2 continuum in Ontario, Canada. In *Variation in Second and Heritage Languages*; Bailey, R., Preston, D., Li, X., Eds.; John Benjamins: Amsterdam, The Netherlands, 2022; pp. 223–252.
18. Nadasdi, T.; Vickerman, A. Opening up to Native Speaker Norms: The Use of/I/in the Speech of Canadian French Immersion Students. *Can. J. Appl. Linguist.* **2017**, *20*, 18–31. [CrossRef]
19. van Compernolle, R.A.; Williams, L. The effect of instruction on language learners' sociolinguistic awareness: An empirical study with pedagogical implications. *System* **2013**, *41*, 298–306. [CrossRef]
20. Howard, M. The Advanced Learner's Sociolinguistic Profile: On Issues of Individual Differences, Second Language Exposure Conditions, and Type of Sociolinguistic Variable. *Mod. Lang. J.* **2012**, *96*, 20–33. [CrossRef]
21. Arnott, S.; Brogden, L.M.; Faez, F.; Péguret, M.; Piccardo, E.; Rehner, K.; Taylor, S.K.; Wernicke, M. The common European framework of reference (CEFR) in Canada: A research agenda. *Can. J. Appl. Linguist.* **2017**, *20*, 31–54.
22. Vandergrift, L. *Nouvelles Perspectives Canadiennes: Proposition d'un Cadre Commun de Référence Pour Les Langues Pour le Canada*; Patrimoine Canadien: Ottawa, ON, Canada, 2006. Available online: https://www.caslt.org/pdf/Proposition_cadre%20commun_reference_langues_pour_le_Canada_PDF_Internet_f.pdf (accessed on 18 September 2022).
23. Ontario Ministry of Education. *A Framework for French as a Second Language in Ontario Schools: Kindergarten to Grade 12*; Ontario Ministry of Education: Toronto, ON, Canada, 2013.
24. Faez, F.; Majhanovich, S.; Taylor, S.K.; Smith, M.; Crowley, K. The power of "Can Do" statements: Teachers' perceptions of CEFR-informed instruction in French as a second language classrooms in Ontario. *Can. J. Appl. Linguist.* **2011**, *14*, 1–19.
25. Kristmanson, P.; Lafargue, C.; Culligan, K. From action to insight: A professional learning community's experiences with the European Language Portfolio. *Can. J. Appl. Linguist.* **2011**, *14*, 53–67.
26. Rehner, K. *French as a Second Language (FSL) Student Proficiency and Confidence Pilot Project 2013–2014: A Report of Findings*; Ontario Ministry of Education and Curriculum Services Canada: Toronto, ON, Canada, 2014.
27. Vandergrift, L. The DELF in Canada: Perceptions of Students, Teachers, and Parents. *Can. Mod. Lang. Rev.* **2015**, *71*, 52–74. [CrossRef]
28. Kristmanson, P.; Lafargue, C.; Culligan, K. Experiences with autonomy: Learners' voices on language learning. *Can. Mod. Lang. Rev./La Rev. Can. Des Lang. Vivantes* **2013**, *69*, 462–486. [CrossRef]
29. French, L.M.; Beaulieu, S. Can beginner L2 learners handle explicit instruction about language variation? A proof-of-concept study of French negation. *Lang. Aware.* **2020**, *29*, 272–285. [CrossRef]
30. van Compernolle, R.A.; Williams, L. Teaching, Learning, and Developing L2 French Sociolinguistic Competence: A Sociocultural Perspective. *Appl. Linguist.* **2012**, *33*, 184–205. [CrossRef]
31. van Compernolle, R.A. Concept appropriation and the emergence of L2 sociostylistic variation. *Lang. Teach. Res.* **2013**, *17*, 343–362. [CrossRef]
32. Stigers, E. Television Subtitles for Stylistic Awareness: Developing and Implementing Media-Based Classroom Activities. *Fr. Rev.* **2021**, *95*, 19–42. [CrossRef]

education sciences

MDPI

*Article*

# Language Learning Investment in Higher Education: Validation and Implementation of a Likert-Scale Questionnaire in the Context of Compulsory EFL Learning

**Leonor Dauzón-Ledesma * and Jesús Izquierdo**

División Académica de Educación y Artes, Universidad Juárez Autónoma de Tabasco, Ave Universidad S/N, Villahermosa 86040, Tabasco, Mexico; jesus.izquierdo@mail.mcgill.ca
* Correspondence: leonordale@hotmail.com

**Abstract:** Second language learning investment relates to the willingness and effort of learners to develop language competencies which will give them a good return in terms of personal or professional benefits. Research has often explored learning investment through learners in the target language context or language teachers. This study, however, explores learning investment with undergraduate learners who are obligated to learn English as a foreign language, regardless of their future profession. To this end, a Likert-scale questionnaire was first designed to examine four investment dimensions which have been identified in previous qualitative research: motivation, necessity, engagement and agency. For validity and reliability purposes, the questionnaire was administered to six second language research professors and 41 students who completed three compulsory English courses in a BA in Inclusive Education. Content, construct and convergent validity procedures were implemented to test the investment dimensions. Regarding reliability, equivalent forms were used to check the stability of answers and to avoid primacy and fatigue effects. In addition, internal consistency and inter-item correlations were checked through Cronbach Alpha coefficients. After the validity and reliability procedures, the four dimensions of learning investment were explored among the language learners. The statistical analyses revealed favorable motivation and engagement results. Nonetheless, they raised some concerns regarding necessity and agency.

**Keywords:** learning investment; language learning; higher education; motivation; needs; engagement; agency; L2 quantitative research

**Citation:** Dauzón-Ledesma, L.; Izquierdo, J. Language Learning Investment in Higher Education: Validation and Implementation of a Likert-Scale Questionnaire in the Context of Compulsory EFL Learning. *Educ. Sci.* **2023**, *13*, 370. https://doi.org/10.3390/educsci13040370

Academic Editors: Martin Howard and Doris L. Watson

Received: 17 January 2023
Revised: 15 March 2023
Accepted: 31 March 2023
Published: 4 April 2023



## 1. Introduction

In the field of second language acquisition, investment is a construct put forward by Norton in 1995. Based on Bourdieu's (1987) [1] ideas about cultural, linguistic and economic capital, Norton explains that students invest in language learning to develop competencies that can give them, for instance, peer recognition, job opportunities or economic benefits that will help them compensate for the effort that the learning process requires. In the case of immigrant students, they invest in learning a language which is necessary for social interaction and integration into a community outside of the classroom [2]. For these learners, language learning investment gives them opportunities to succeed in life. In educational settings, where the new language is not spoken outside of the classroom, language learning investment relates to the development of the communicative competence which can translate into economic benefits and professional recognition [3,4]. In both contexts, language learning investment then creates good return conditions. Therefore, this construct has been explored to understand its benefits during second/foreign language acquisition.

In a world where English is a global language, the exploration of language learning investment has become a necessity, since educational stakeholders have turned to policies and curricular changes that sanction the learning of this language not only in primary, secondary and vocational education but also in higher education [5–7]. These policies aim

to equip students with the linguistic skills that they need in order to join the productive forces of a country upon the completion of their university studies [8]. In the case of undergraduate students, irrespective of their future profession, compulsory English language learning may imply that the students need to satisfy requirements, such as language attainment levels, language courses and sometimes, standardized language tests [5,9].

In the context of higher education, it then becomes important to examine the learning investment of the students as they meet the curricular demands for learning the English language. To date, a handful of qualitative studies have explored this issue through individual cases in different contexts [10–12]. As indicated in the upcoming sections, these studies provide some valuable insights into the dimensions that seem to underpin language learning investment. However, they are limited in terms of the representativeness and generalizability of the results in higher education [10], which includes large and diverse populations of future professionals [5]. As argued in the manuscript, the Likert-scale questionnaire is one type of instrument that could be considered for the elicitation of representative and generalizable quantifiable data [13–18] from large student populations.

Thus, to better understand language learning investment in the context of compulsory English language learning in higher education, the purpose of this study was twofold: first, it aimed at the development of a Likert-scale questionnaire that allowed for the systematic exploration of the construct of language learning investment and its underpinning dimensions; second, through its administration in an undergraduate program in inclusive education, the study examined the language learning dimensions of investment among students who are obligated to meet compulsory English language learning demands in the foreign language learning context.

## 2. Literature Review

Investment relates to a learner's willingness to learn something which they believe could "give them a good return on that investment" [19] (p. 17), [20,21]. Qualitative research findings indicate that the construct of investment is complex and interweaves different aspects of language learning, such as motivation [22], necessity or personal needs [2,23–26], engagement [20,27,28] and agency [29]. The upcoming sections elaborate upon these dimensions and discuss research gaps that require attention in the context of higher education.

### *2.1. Language Learning Investment Dimensions*

In the study of learning investment, motivation can be defined as "an internal state that enables an action (to learn) and involves understanding the factors that cause this state" [22] (p. 5). Motivation is described as the reasons that determine an individual's behavior to achieve a goal. Motivation gives direction to intentions and actions. Learners invest in learning because "they will acquire a wider range of symbolic and material resources, which will in turn increase the value of their cultural capital and social power" [30]. Since attitudes can be considered as part of motivation [26], studies on language learning investment often use the terms interchangeably [31] or explore attitudes and motivation together [32]. The learners' motivation may yield results that generate satisfaction and detect benefits in the learning process [33,34]. When learners are motivated to learn, investment in language learning produces good outcomes that act as positive feedback. Learners then enter a positive cycle, where investment could benefit their economic capital in terms of employment opportunities. The benefit increases motivation and promotes more investment.

The second factor which is often associated to investment is the learners' necessity to use the language [3,29,35]. Learning necessities vary depending on the learners' differences, their linguistic and sociocultural backgrounds [23] and the expectations they have regarding the outcomes and personal benefits of learning a second language. In this regard, personal, professional, heritage or economic needs can have an influence on learning investment [2,23–26]. The reasons that different groups of learners have to learn a second

language vary depending on the context and factors that are involved. For instance, there are learners immersed in the second language context, such as immigrants in English-speaking communities who need the target language to communicate, work, become part of society and have the chance to speak and build relationships [33]. There are learners who are not immersed in the target language context but need the language for professional or vocational purposes [35]. Moreover, in the context of foreign language learning, the learners' needs may relate to recognition and profit [12].

Engagement and agency are two factors that are also associated to investment when the learners show interest, have initiative and set goals for their own learning [29]. According to Norton, engagement in language learning investment is related to actions [20]. Although Hiver et al. indicate that engagement is composed of different dimensions, Norton's construct of investment seems more connected to Hiver et al.'s behavioral dimension of engagement which implies action and voluntary active involvement on a task [27]. In Hiver's words, engagement "refers to the amount (quantity) and type (quality) of learners' active participation and involvement in a language learning task or activity" [27] (p. 2). The behavioral aspect of engagement has also been acknowledged by other authors. For instance, Mercer [28] (p. 645) refers to Skinner et al. (2009, p. 225), who describe engagement as "energized, directed, and sustained actions." Moreover, Angelovska et al. indicate that engagement implies behaviors that show effort and action while achieving goals [14]. Within this perspective of engagement, if there is no engagement during the learning process, there will be no achievement of goals, and the development of the learners' competencies will not occur. Since motivation and engagement relate to actions, Artamonova [31] highlights that motivation is limited to the intention of doing something. Nonetheless, engagement involves the realization of an action.

Agency is another concept that has been associated with investment. Harrison et al. [29] (p. 4) define it as "actions where the students contribute actively to shape their own learning thereby enhancing their investment in the process." Agency is closely related to investment since learners expect to receive something good in return for the effort and time that they invest in chosen activities with a specific goal. Agency can be identified through will but also, above all, through the determination and perseverance that the learners have in order to achieve goals which produce good results at the end of a learning process [29]. In other words, the learners take the initiative of deciding how to undertake the learning process and regulate their achievements through the autonomous selection of learning activities and time organization [36]. As explained by Naderpour [37], engagement can be considered the first step of agency while taking action and perseverance would be the last step of agency. It can be said that, during language learning investment, motivation, necessity, engagement and agency are interrelated and support the learning process and achievement of goals. Nonetheless, agency implies greater efforts that are mediated through initiative, determination and perseverance.

### *2.2. Second Language Learning Investment in Higher Education*

Research has provided some initial evidence on the aforementioned factors as important dimensions of language learning investment among different groups of students. One group includes learners who are immersed in the language context and invest in language learning to interact and communicate with others in real life. Due to language learning investment, they can become part of the community where they live and learn the culture [22,25]. Additionally, due to their desire to be part of the society which uses the target language, through language learning investment, the learners claim their right to speak and be recognized in the second language community. While the learners invest in language learning, they acquire resources that increase their cultural capital and social power. Additionally, changes in their identity may occur as they become part of the community [30].

A second group of learners includes those that are not immersed in the target language context but require the language for professional purposes. Within this group, a population

that has received attention includes professionals in the area of language teaching [14,38,39], particularly teachers of English as a foreign language (EFL). Research with pre-service [11,40] and in-service teachers of English [41] has revealed that language learning investment constitutes a valuable asset because of the need to master the language in the profession [17]. Language educators will teach the language as the object of study and as the means of communication. Moreover, for language education professionals, language learning investment promotes a sense of belongingness [41]. For English language teachers who are not immersed in the target language context, language learning investment contributes to the process of second language identity construction. Motivation and professional needs also constitute the driving force for this group because meaningful interests can be attached to learning investment. Their desire to succeed in language learning for professional development may instantiate engagement and agency when the activities are significant for the accomplishment of previously established language learning goals [42].

Educational stakeholders and institutions are promoting the learning of English in higher education [4,24] and the establishment of EFL requirements [9]. This policy aims to enhance the success and level of competitiveness of future professionals [4,24,35]. Within this context, attention should be paid to the factors that contribute to learning investment among learner populations that will require English language competencies in professions outside of language teaching. To date, qualitative insights into what people from different higher education programs think about EFL learning come, for instance, from the work of Diep and Hieu, who used questionnaires and interviews with students in Engineering and Technology, Economic and Business Administration, Health Science and Social Sciences and Humanities [9]. In this study, the participants acknowledged the necessity of learning English during their education and recognized that personal motivation and attitudes affect the quality of language learning [9]. Although the participants came from different programs, they indicated that English constitutes an add-on qualification that can translate into better job opportunities and the chance to communicate with colleagues for business or research. However, the learning environment, the curriculum and the quality of teaching may have an impact on language learning outcomes.

In a different study, Lacka-Badura elicited data from students enrolled in different areas of studies such as Management, Tourism, Economics, Finance and Accounting and International Economic Relations, among others. As the participants had some previous English language learning experience, the author indicated the students' needs and expectations which should be considered to enhance learners' motivation and engagement [35]. Across different professions, learners see then that the learning of an additional language can pave the road for a good future. Nonetheless, as mentioned by Amorati [24], their vision might lay the foundation of language learning necessity, but it is not enough to understand whether the individuals actually engage in and become agents of their own learning. The aforementioned findings from previous qualitative research in higher education provide some insights into the dimensions that seem to contribute to language learning investment, particularly among future language teachers. Among other populations of learners in higher education, the evidence from individual cases instantiates the necessity to learn the language and the benefits the students believe that they will have in their future profession. These findings are informative but build upon individual experiences. Therefore, questions arise about the dimensions of investment [17,27,29] that may come into play as higher education students encounter compulsory EFL learning. Among these learners, there is a need to explore language learning investment vis-à-vis the state-mandated language learning opportunities they have [9,10].

In addition to the individual case nature of the previous qualitative studies, the absence of generalizable results relates to the data collection instruments that have been used, such as interviews, observations and life stories [11,42,43]. In this research, the instruments have not led to a quantifiable exploration and identification of the dimensions that instantiate language learning investment, as these instruments can be administered to a small number of cases only [13]. Quantifiable data are therefore needed because they

provide representative results from large samples [13,44] and help us determine whether the results hold internal and external validity; that is, qualitative data illustrate whether the results are context-dependent or generalizable. Moreover, quantifiable data can help us understand how the various dimensions of the construct interact and become operative during compulsory language learning.

## 3. Materials and Methods

In light of the aforementioned issues, this study sets two research objectives: First, to design an instrument that allows for the quantifiable exploration of the construct of language learning investment in higher education. Second, it explores the four dimensions of the construct (i.e., motivation, necessity, engagement and agency) among higher education students who are obligated to meet EFL learning demands. The following research question is derived from these objectives. What are the dimensions that underpin the language learning investment of higher education students who comply with compulsory English language education in the foreign language context?

To achieve the study objectives, a two-phase quantitative study with a descriptive design was conducted. During the first phase of the study, a Likert-scale questionnaire was designed. To this end, the various dimensions of the questionnaire were operationalized [13,18] and the instrument was subject to various validity and reliability procedures [13,45]. During the second phase of the study, the participants' answers were analyzed [46,47] to explore the various dimensions of language learning investment.

### *3.1. Participants*

Due to the objectives of the current study, the participants of the study were selected using convenience sampling. While this procedure is non-probabilistic, it allowed for the selection of participants whose profiles would align with the nature and phases of the research.

One group of participants consisted of six research professors who acted as experts for the purpose of content validation during the first phase of the study. This group participated as jurors during the instrument validation phase only. Five of them were affiliated to universities in Mexico and were native speakers of Mexican Spanish with international certifications that attest their EFL proficiency. The sixth professor was a Canadian English native speaker with knowledge of Mexican Spanish and was affiliated to a Canadian university. They all have published research on the learning of English in public education and have taught English language teachers in undergraduate and graduate programs. The answers from the professors were used in the first phase of the study only.

The second group of participants consisted of 41 higher education students in a BA in Inclusive Education in a public teacher training school in the southeast of Mexico. The participants have already completed two years of teacher training and three compulsory EFL courses in their BA program. Their responses were used during the first and second phases of the study. They were selected on the bases of availability and accessibility criteria [13,48]. They filled out a sociodemographic survey, where they provided information about age, gender, years of education and linguistic background, such as languages spoken at home. This was a homogeneous group of participants who use Spanish to communicate on a daily basis. The parents of two participants spoke a native language, and the parents of five learners had learned English. Prior to the data collection process, the project was presented to the authorities of the higher education institution who granted us access to the teachers and students through verbal and written consent. The students were informed of the research; confidentiality and anonymity were assured; their right of leaving the project was explained and consent for the use of the data was obtained. The students' answers were used in the first and second phases of the study.

The student sample included 6 male and 35 female students. They were finishing their second year of college and were at the end of the fourth compulsory English course. In their

program, they completed six hours of English lectures per week. Most of the participants were between 17 and 20 years old; 17% of the participants were between 21 and 24 years old and 5% were 25 years old or older. Contrary to what educational policies dictate for elementary education, learners acknowledged receiving EFL instruction during middle school and high school only. Two participants informed they had additionally attended EFL courses in a private language institute. Regarding the learning of English, their parents supported the idea of compulsory EFL education. The students (95.1%) had never travelled to an English-speaking country, although some of them (48.7%) have relatives living abroad. Table 1 presents the participants' reasons to learn English.

**Table 1.** Reasons to learn English.

| | Yes | No |
|---|---|---|
| To comply with the curriculum | 100% | |
| To follow my parents' decisions | 26.2% | 73.8% |
| To travel | 45.2% | 54.8% |
| To study abroad | 35.7% | 64.3% |
| To communicate with people | 76.2% | 23.8% |
| To understand songs, movies and videogames | 85.7% | 14.3% |

### *3.2. Language Learning Investment Questionnaire*

To explore the construct of language learning investment, the Likert-scale questionnaire focused on the four dimensions that were covered in the literature review (motivation, necessities, engagement and agency), although investment has been sporadically associated with commitment [20], effort, values [18] and self-efficacy [25]. For the exploration of these language learning dimensions, the use of an investment Likert-scale questionnaire was considered. Scale questionnaires allow social science researchers to explore constructs that rely on the measurement of opinions, attitudes, behaviors and perceptions [18,44,49,50]. To this end, the participants are presented with a series of stimuli that cause a reaction in the informant who then identifies a degree of response through the selection of nominal values in an ordinal scale [18]. The elicited data are assigned a numerical value that can be statistically analyzed for possible generalizations and replication. Nonetheless, the design of such an instrument requires adherence to theoretical and methodological principles of quantitative research. These principles should permeate the conceptualization, validity and reliability procedures of the instrument and provide sufficient information for the controllability and replicability of the research in other contexts [13,45], as the following sections depict.

#### 3.2.1. Questionnaire Dimensions

Based on the literature review, the questionnaire was conceptualized to explore motivation, necessities, engagement and agency, since these dimensions of learning investment often emerge in the qualitative empirical evidence. The first dimension, motivation, is defined as an internal state that gives the reason for a specific behavior and direction to an action [22,35]. The second dimension refers to necessities that are related to the personal or professional interest of a person [23]. The third dimension is engagement and implies action or active participation and involvement in a learning task [27]. The fourth dimension is agency, which refers to the actions that imply a personal initiative to shape learning [29]. The expanded definitions of these dimensions were provided in the literature review and laid the basis for the creation of the items in each questionnaire section.

#### 3.2.2. Items

Based on the quantitative research literature, three central principles for the design of the questionnaire items were observed: unidimensionality, univocality and semantic direction [49,51]. Based on these criteria, a set of items was developed for each dimension;

each item focused on one element of the dimension of interest, and each item presented a positive statement for the dimension's element.

Some items were adapted from other instruments [15,16,31,32] and others were created on the basis of the construct and dimensions of interest. One of the adaptations related to the language in the original questionnaires. It was decided to use the participants' mother tongue (i.e., Spanish) to avoid misunderstandings. Another change was the rewording of the items to avoid confusion, ambiguity or negativity. The initial version of the questionnaire included 51 items that were distributed across the dimensions of motivation ($n$ = 11), necessities ($n$ = 9), engagement ($n$ = 20) and agency ($n$ = 11).

#### 3.2.3. Scale

In addition to the conceptualization of the items, an aspect that requires attention is the scale through which the participants will express their opinions. This implies the use of a set of gradually interrelated answer options which go from a positive to a negative stand, or vice versa [13,18]. Regarding the best number of answer choices in the scale, researchers have not come to an agreement. For instance, some have used from four to eleven points. Nonetheless, the seminal work by Guy and Norvell (1977) [52] indicates that reliability is independent of the number of points in the scale.

While some authors indicate that, among the answer choices, the participants should be given the opportunity to remain neutral [13], others indicate that this opportunity should not be provided [51]; if the neutral point is omitted, then the participant is forced to take a position with respect to the stimulus that is presented [13,51,53]. When this happens, the participants demonstrate sensitivity by not using extreme responses. Therefore, most of the responses are middle-range to compensate for the missing point [52]. Nonetheless, the middle-range responses provide an indication of the positive or negative standing of the participant in reference to the stimuli. In light of these considerations, a small number of answer choices was considered. Moreover, since the items were conceptualized to elicit the participants' level of agreement with the items, the questionnaire included a four-point agreement scale that went from a negative to a positive stand: totally disagree, partially disagree, partially agree, totally agree.

### *3.3. Scoring Procedures*

Once the scale was established, a numerical value was assigned to the answer choices. This was carried out following the quantitative principle that values are assigned following a continuum, where the extreme negative answer holds the lowest value and the extreme positive answer holds the highest [13,18,49]. The numerical values allow for the numerical treatment of opinion-related data. To collect responses in the questionnaire, a four-point Likert scale was used.

There are different opinions about the numerical nature of the scale values and their treatment. Ordinal scales do not guarantee equal intervals of measurement, while interval scales work with equal intervals. Some researchers consider that the values are ordinal and should be treated as non-numerical because, even if it is possible to determine the direction of the difference between the values, it is not possible to determine the size of the difference [54]; this means there is not an "equal-sized gradation between the points" [13] (p. 481). Other researchers suggest the scale could be treated as an interval because of the numerical properties it has when the answer options are assigned values. As explained by Larson-Hall (2016) [47], the treatment depends on the kind of variable, whether it is categorial or continuous, and the way the researcher decides to manage it.

When the numerical values across the questionnaire items are conceived as an interval scale, they can be added up. A general score is obtained and becomes representative of the opinion, perception or attitude that is being measured [49]. This procedure gives the scale an additive property [49]. Finally, these properties allow researchers to use statistical procedures whereby analyses for possible group or dimensional comparisons can be run [13,46]. Based on these principles, it was decided to treat the numerical values

as an interval scale with additive properties. Thus, each nominal choice in the scale was assigned a numerical value (1 = totally disagree, 2 = partially disagree, 3 = partially agree, 4 = totally agree), and the values of the items in each dimension were added up in order to obtain a dimension score.

### *3.4. Validity*

For Cohen et al. (2018) [13], to validate an instrument is basically to prove that it measures what it intends to measure, that it represents the theory, concepts or conclusions it intends to explain. While there are different forms of validity, our instrument was subject to content and construct validity.

Content validity refers to the topic, domain or concepts that should be covered in an instrument. The relevance of the content can be evaluated through professional judgement [13]. This kind of validity ensures the coverage and the relevance of an instrument. To carry out this type of validation in our study, a committee of six experts (i.e., the research professors) in the topic reviewed and evaluated the content of the instrument. This helped us avoid bias that might happen when a single researcher revises the items [13] (p. 262).

Construct validity is fundamental because it refers to the construct itself or its definition and not to methodological factors which operationalize it. It is necessary to have a clear and warranted theoretical construction of the addressed issue. Cohen et al. (2018) [13] cite other authors who explain that construct validity can be addressed by different techniques. When different methods yield a high inter-correlation for the same construct, they are convergent. Thus, construct validity was operationalized through convergent validity. Convergent validity refers to elements or factors that are related and are consistent with each other. Convergent validity is proven when the relation between factors that was previously assumed is verified by running a test to find the appropriate indicators. To this end, correlation analyses were used in this study because they could test the relationship among the dimensions of the construct [18].

### *3.5. Reliability*

The reliability of the instrument [46,55] was examined through equivalence of forms and internal consistency. Equivalence is a concept related to reliability that implies the use of parallel forms to gather data. In this case, reliability is demonstrated when two forms of the same instrument show consistent results through parametric tests for normally distributed data or non-parametric tests for non-normally distributed data. In the context of this study, the use of equivalent forms was operationalized through the use of the same questionnaire items, but the items were presented in reverse order to avoid primacy and fatigue effects [13]. Thus, two versions of the test were used: Version 1 and Version 2.

Internal consistency examines if there is homogeneity in the items in a questionnaire [18]. To test the internal consistency of a questionnaire, it is necessary to check the Cronbach Alpha coefficient [46,49]. Its value can range from zero to one. The expected acceptable value for the Cronbach Alpha coefficient is 0.70. The coefficient demonstrates the cohesiveness of the items included in the questionnaire. In addition to the Cronbach Alpha, the inter-correlation coefficient of each item can also be used to decide which items contribute to the cohesiveness and should therefore be retained. This is accomplished by retaining items that yield an inter-item correlation above 0.3 [46]. This procedure helps researchers reduce the number of questionnaire items.

## 4. Analysis Procedures and Results

In this section, the validity and reliability results are first presented. These results are presented following the validity and reliability steps, as they were undertaken: (1) Content Validity, (2) Stability of Answers, (3) Internal Consistency and (4) Construct Validity. Then, using the final version of the questionnaire, the data from the participants are analyzed to examine the dimensions of language learning investment in the context of compulsory EFL education.

### 4.1. Content Validity

To check content validity, the experts were asked to determine whether the items represented the construct that was being measured in each section [46]. To this end, the jurors rated first the congruence between the items with the construct and their corresponding dimension. Then, they were asked to assess the comprehensibility of the items. The jurors expressed their opinions by rating the items, writing comments on individual items, providing feedback related to wording and presenting suggestions to ensure the clarity of the items.

In order to examine content validity, the jurors' answers were treated using descriptive statistics. For each item, two agreement ratios were computed. One ratio was obtained to identify the congruence between the item and the construct. The other ratio was computed to identify the congruence between the item and its comprehensibility. To obtain these ratios, the number of experts that expressed agreement was divided by the total number experts. For instance, item 1 in Section 1 elicited agreement from five experts during the examination of the item and its dimension. This item, therefore, yielded an agreement ratio of 83.3% (i.e., $5/6 \times 100$).

If the item obtained 100% agreement, it was retained without modifications; if the item achieved between 50 and 99% agreement, it was retained with modifications. If the item obtained a ratio of agreement below 50%, it was excluded. From the original list of 51 items, 6 items obtained 50% agreement in congruence with the construct. These items were revised, and the wording was changed to make them clearer and more congruent with the construct. Therefore, the original number of items per section was not affected: motivation ($n$ = 11), necessities ($n$ = 9), engagement ($n$ = 20) and agency ($n$ = 11).

### 4.2. Reliability: Stability of Answers

To verify the stability of answers, the data collected through the parallel forms of the questionnaire were examined using between-group comparisons. To this end, first, the normal distribution of the data collected for each item in both versions of the questionnaire was tested using the Kolmogorov–Smirnov test. In this test, normality is assumed when the significance value is greater than the alpha level of 0.05. Based on this criterion, the analyses of the individual items indicated that only two items obtained a $p \geq 0.05$: item 3 in Section 1 version 2 and item 10 in Section 4 version 2. In the analyses of the dimension scores, in version 1, only the score of dimension 2 (i.e., necessities) achieved normality. In version 2, the normality distribution was observed in the dimension scores of motivation (dimension 1), engagement (dimension 3) and agency (dimension 4).

Due to the absence of a consistent normal distribution in the questionnaire data, non-parametric analyses were used to test whether the participants provided similar answers between versions and, therefore, could be considered reliable. As the data came from independent samples, non-parametric Kruskal–Wallis tests were run to check answer differences between the two versions. The results showed a difference in the answers of the questionnaire item 14 from Section 3 (engagement), H(1) = 4.87, $p$ = 0.027 and item 2 from Section 4 (agency), H(1) = 4.49, $p$ = 0.034. Since these items exhibited unreliable answer patterns, they were excluded from the final version of the scale questionnaire. Thereafter, 49 items were retained. As there was not a significant difference between questionnaire versions in the remaining items, the answers from both versions were pooled in the upcoming analyses.

### 4.3. Reliability: Internal Consistency

After checking the stability of answers, the Cronbach Alpha and inter-item correlation coefficient were verified. To claim reliability with the Cronbach Alpha test, the expected value for the internal consistency value should be greater than 0.7, and for the inter-item correlation, it should be greater than 0.3. As the items were constructed and initially grouped into dimensions, independent Cronbach analyses for each dimension were run. The results for dimension 1 revealed a reliability coefficient of 0.807 and that items 2 and 5

exhibited a corrected item correlation below 0.3. Based on these results, these items were excluded, and new reliability analyses were run considering only the results from items 1, 3, 4, 6, 7, 8, 9, 10 and 11 of dimension 1. The new results yielded a higher reliability coefficient of 0.881.

For dimension 2, the initial Cronbach coefficient was 0.772. In this dimension, only item 3 exhibited a corrected correlation below 0.3. After the exclusion of this item, the statistical test was rerun including items 1, 2, 4, 5, 6, 7, 8 and 9 of dimension 2. This time, the reliability coefficient for dimension 2 increased to 0.779.

Dimension 3 included 20 items. The initial correlation coefficient was 0.747 and items 7, 9, 10, 11, 12, 17 and 18 showed a corrected correlation coefficient below 0.3. After the exclusion of these items, a Cronbach analysis was run including items 1, 2, 3, 4, 5, 6, 8, 13, 15, 16, 19 and 20 of dimension 3. The second analysis yielded a higher coefficient of 0.853.

The analyses for dimension 4 with the initial 10 items exhibited a correlation coefficient of 0.670 and a correlation coefficient below 0.3 for items 4, 6 and 10. After the exclusion of these items, the test was run again with items 1, 3, 5, 7, 8, 9 and 11 of dimension 4. The final reliability coefficient for dimension 4 was 0.729.

After the internal consistency results and item exclusion procedures, the four sections were considered reliable, as they obtained coefficient values greater than 0.7 and inter-item correlations above 0.3. In sum, from the initial number of 51 items that were subject to reliability analyses, 36 items were retained for the final version.

### *4.4. Construct Validity*

Construct validity allows the researcher to prove that a scale questionnaire shows correlation between dimensions and the item clusters within each dimension in the questionnaire [13]. Construct validity could be checked through correlation or factor analysis. Since the items were conceived independently for each of the dimensions of interest and the independent reliability analyses for each dimension showed high inter-item correlation coefficients [46], correlation analyses were used to achieve convergent validity and thereby prove the relation among the dimensions of the questionnaire [13] (p. 258).

The correlation analyses were run using the scores that result from the addition of the items that remained in each dimension. Due to the skewness of data (i.e., the absence of a normality) in the majority of the items and dimension scores, one-tailed Spearman tests were run with the four dimensions of the questionnaire. The results revealed a significant correlation between all dimensions with a $p$ value of 0.01. The analysis of the correlation strength revealed a moderate correlation among motivation, engagement and agency, based on Hinkle et al.'s (2003) [56] interpretation of the correlation coefficient. However, dimension 2, necessity, showed a weak correlation with engagement (rho $\rho = 0.434$) and agency (rho $\rho = 0.486$), and even weaker correlation with motivation (rho $\rho = 0.179$). In sum, a stronger correlation is evident among three of the language learning investment questionnaire dimensions: motivation, engagement and agency.

### *4.5. Language Learning Investment Dimension Results*

As mentioned at the beginning of this paper, the second research objective was to explore the motivation, necessity, engagement and agency in the language learning investment of higher education students who are obligated to learn English in a BA program outside of the L2 teaching profession.

In the upcoming sections, for each item in the questionnaire dimensions, the median and the mode are provided from Tables 2–5, based on the argument that these central tendency values best portray answer patterns with ordinal scale data which do not meet a normal distribution [13]. Then, in these tables, the distribution of the participants across the scale answer choices is provided in percentages. In the interpretation paragraphs, the percentages for partial and total agreement are pooled; the same procedure was used with the results for partial or total disagreement.

**Table 2.** Answer Analysis Results for Motivation.

| Section 1: Motivation | Mean | Median | Mode | SD | Percentages | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Totally Disagree | Partially Disagree | Partially Agree | Totally Agree |
| I am interested in having the materials prepared for the class. | 3.68 | 4.00 | 4 | 0.521 | 0.00 | 2.44 | 26.86 | 70.73 |
| I am very perseverant in completing my English class activities. | 3.07 | 3.00 | 3 | 0.608 | 0.00 | 14.63 | 63.41 | 21.95 |
| I enjoy the time I spend on my English activities. | 2.46 | 3.00 | 3 | 1.027 | 21.95 | 26.83 | 34.15 | 17.07 |
| I enjoy solving the English exercises that the teacher gives us. | 3.00 | 3.00 | 3 | 0.949 | 12.20 | 7.32 | 48.78 | 31.71 |
| I think speaking English in front of my classmates makes me nervous. | 3.02 | 3.00 | 4 | 1.129 | 17.07 | 9.76 | 26.83 | 46.34 |
| I recognize that solving the English activities involves an effort that I am willing to make. | 3.49 | 4.00 | 4 | 0.779 | 2.44 | 9.76 | 24.39 | 63.41 |
| I am very enthusiastic about learning English. | 3.15 | 3.00 | 3 | 0.937 | 9.76 | 7.32 | 41.46 | 41.46 |
| I enjoy doing English activities. | 3.05 | 3.00 | 3 | 0.865 | 7.32 | 12.20 | 48.78 | 31.71 |
| I am proud to be able to conclude the class activities successfully. | 3.68 | 4.00 | 4 | 0.650 | 2.44 | 2.44 | 19.51 | 75.61 |
| I am interested in studying English more than other subjects. | 2.71 | 3.00 | 3 | 0.873 | 9.76 | 26.83 | 46.34 | 17.07 |
| I believe that the investment of time and money to learn English is well worth it. | 3.54 | 4.00 | 4 | 0.636 | 0.00 | 7.32 | 31.71 | 60.98 |

**Table 3.** Answer Analysis Results for Necessities.

| Section 2: Necessities | Mean | Median | Mode | SD | Percentages | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Totally Disagree | Partially Disagree | Partially Agree | Totally Agree |
| I need English to: communicate with other people. | 3.83 | 4.00 | 4 | 0.381 | 0.00 | 0.00 | 17.07 | 82.93 |
| interact with people from other cultures. | 3.68 | 4.00 | 4 | 0.521 | 0.00 | 2.44 | 26.83 | 70.73 |
| be part of communities from other countries. | 3.73 | 4.00 | 4 | 0.549 | 0.00 | 4.88 | 17.07 | 78.05 |
| obtain a scholarship. | 3.71 | 4.00 | 4 | 0.680 | 2.44 | 4.88 | 12.20 | 80.49 |
| have access to updated information. | 3.63 | 4.00 | 4 | 0.536 | 0.00 | 2.44 | 31.71 | 65.85 |
| get a well-paid job. | 3.63 | 4.00 | 4 | 0.662 | 2.44 | 2.44 | 24.39 | 70.73 |
| get a job abroad. | 3.71 | 4.00 | 4 | 0.602 | 2.44 | 0.00 | 21.95 | 75.61 |
| access technology. | 3.12 | 3.00 | 3 | 0.900 | 9.76 | 4.88 | 48.78 | 36.59 |
| earn more money by demonstrating more competencies than others. | 3.39 | 4.00 | 4 | 0.802 | 4.88 | 4.88 | 36.59 | 53.66 |

**Table 4.** Answer Analysis Results for Engagement.

| Section 3: Engagement | Mean | Median | Mode | SD | Percentages | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Totally Disagree | Partially Disagree | Partially Agree | Totally Agree |
| Attending English classes regularly is important to me. | 3.63 | 4.00 | 4 | 0.623 | 0.00 | 7.32 | 21.95 | 70.73 |
| Having the material that I need for the class is a priority for me. | 3.44 | 4.00 | 4 | 0.673 | 0.00 | 9.76 | 36.59 | 53.66 |
| Doing well on the work that is assigned to me during class is a worthwhile endeavor. | 3.68 | 4.00 | 4 | 0.521 | 0.00 | 2.44 | 26.83 | 70.73 |
| Doing the homework is useful for me to review what I covered in class. | 3.59 | 4.00 | 4 | 0.591 | 0.00 | 4.88 | 31.71 | 63.44 |
| Paying attention to the teacher's explanations is necessary for me. | 3.88 | 4.00 | 4 | 0.331 | 0.00 | 0.00 | 12.20 | 87.80 |
| Doing the written activities that are assigned by the teacher is useful to me. | 3.66 | 4.00 | 4 | 0.617 | 0.00 | 7.32 | 19.51 | 73.13 |
| Participating in the assigned speaking activities increases my speaking confidence. | 3.39 | 4.00 | 4 | 0.802 | 4.88 | 4.88 | 36.59 | 53.66 |

**Table 4.** *Cont.*

| Section 3: Engagement | Mean | Median | Mode | SD | Percentages | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Totally Disagree | Partially Disagree | Partially Agree | Totally Agree |
| Taking notes during class makes me feel confident. | 3.63 | 4.00 | 4 | 0.733 | 2.44 | 7.32 | 14.63 | 75.61 |
| Listening attentively when my classmates participate is important to me. | 3.34 | 3.00 | 3 | 0.656 | 0.00 | 9.76 | 46.34 | 43.90 |
| Copying the activities from my peers implies less effort to me. | 3.05 | 3.00 | 4 | 1.024 | 9.76 | 19.51 | 26.83 | 43.90 |
| Waiting for others to respond to the teacher's requests prevents me from stressing out. | 2.05 | 2.00 | 1 | 1.024 | 39.02 | 26.83 | 24.39 | 9.76 |
| I study the materials the teacher gives me, even if they are not of interest to me. | 2.78 | 3.00 | 3 | 0.936 | 9.76 | 26.83 | 39.02 | 24.39 |
| I make an effort to complete the reading exercises, even if they are difficult for me. | 3.49 | 4.00 | 4 | 0.675 | 2.44 | 2.44 | 39.02 | 56.10 |
| I do all the writing exercises, even if I have to spend a lot of time on them. | 3.24 | 3.00 | 4 | 0.799 | 2.44 | 14.63 | 39.02 | 43.90 |
| I make an effort to speak English during class even if I don't feel confident. | 3.12 | 3.00 | 4 | 0.927 | 7.32 | 14.63 | 36.59 | 41.46 |
| I do my best to understand what others say, even if it is difficult. | 3.63 | 4.00 | 4 | 0.536 | 0.00 | 2.44 | 31.71 | 65.85 |
| I speak English with my classmates outside of the class, even if I don't feel confident. | 1.66 | 1.00 | 1 | 0.911 | 56.10 | 29.27 | 7.32 | 7.32 |
| I speak English with the teacher outside of the class, even if it's hard for me. | 1.76 | 2.00 | 1 | 0.860 | 48.78 | 29.27 | 19.51 | 2.44 |
| I complete the English classroom activities, even if I feel uncomfortable with them. | 2.98 | 3.00 | 4 | 1.060 | 12.20 | 19.51 | 26.83 | 41.46 |
| I study just enough to pass the exams. | 3.27 | 3.00 | 4 | 0.775 | 2.44 | 12.20 | 41.46 | 43.90 |

**Table 5.** Answer Distribution Results for Agency.

| Section 4: Agency | Mean | Median | Mode | SD | Percentages | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Totally Disagree | Partially Disagree | Partially Agree | Totally Agree |
| To improve my English: I search for videos on the Internet. | 2.71 | 3.00 | 4 | 1.209 | 24.39 | 17.07 | 21.95 | 36.59 |
| I look for readings in English (internet, magazines or books) to complement what I see in English class. | 2.27 | 2.00 | 3 | 1.025 | 29.77 | 26.83 | 31.71 | 12.20 |
| I watch movies with English subtitles to improve my listening comprehension. | 2.98 | 3.00 | 4 | 1.107 | 17.07 | 9.76 | 31.71 | 41.46 |
| I participate in online communities of videogame players to practice my English. | 1.27 | 1.00 | 1 | 0.708 | 85.37 | 4.88 | 7.32 | 2.44 |
| In my free time, I listen to English songs to improve my English. | 3.10 | 3.00 | 4 | 1.114 | 17.17 | 4.88 | 29.27 | 48.78 |
| I take English classes at other institutions to enhance my English language learning. | 1.24 | 1.00 | 1 | 0.663 | 85.37 | 7.32 | 4.88 | 2.44 |
| Outside of the classroom, I use applications on my cell phone with English games to improve my vocabulary. | 1.98 | 2.00 | 1 | 1.012 | 43.90 | 21.95 | 26.83 | 7.32 |
| If I use apps on mobile devices, I can improve my English grammar. | 2.85 | 3.00 | 4 | 1.174 | 19.51 | 17.07 | 21.95 | 41.46 |
| If I am involved in my learning, I will increase my proficiency level. | 3.66 | 4.00 | 4 | 0.617 | 0.00 | 7.32 | 19.51 | 73.17 |
| I learn more independently if there is virtual interaction with other people. | 2.56 | 3.00 | 2 | 1.074 | 19.51 | 29.27 | 28.63 | 24.39 |
| My confidence increases as my level of performance in communicating in English improves. | 3.37 | 4.00 | 4 | 0.915 | 7.32 | 7.32 | 26.83 | 58.54 |

The median and mode of the items from Tables 2–5 indicate that, with a few exceptions in the dimension of agency, the participants show language learning investment in the

context of compulsory English language education in the foreign language context. The percentage distribution provides a finer picture of the areas that instantiate their investment.

In the dimension of motivation, 92.69% of the participants consider that it is worth investing time and money to learn English (see Table 2). The results revealed that 87.8% agreed on the fact that it is an effort they are willing to make and 95.12% feel proud when they are capable of finishing the tasks, but only 51.22% reported they enjoy the time they spend learning English.

In relation to the necessities the participants reported towards learning English, Table 3 shows that the largest numbers are for communication and to obtain a scholarship (100% and 92.69%, respectively) followed by having job opportunities abroad, to interact with people from other cultures and to increase the possibilities of a well-paid job. Although higher education students are highly exposed to technology, only 36.59% strongly agreed that English is necessary to use technology; however, 97.56% acknowledged that it is necessary to have access to updated information related to their studies.

As for learning engagement, Table 4 shows positive patterns when the participants reported attending classes (92.68%), having the materials for class (90.25%), paying attention (100%), taking notes (90.24%) and working at home (95.15%), because they regard these activities as important to succeed. Notwithstanding, 78.05% reported they try to speak English in class although they do not feel confident and 97.56% make an effort to understand others.

As for the fourth dimension of investment, agency (see Table 5), 73.17% of the students are convinced that they can improve their learning if they get involved in the process and 58.54% recognized they feel confident when they can communicate in English. What they prefer to do by themselves in order to increase their exposure to the target language is watching movies (73.17%) and listening to music (78.05%). However, 85.37% of the learners do not attend English classes in other institutions; 43.90% do not use mobile phones to improve or practice what they have learned and 85.37% answered they are not interested in online interaction with videogame players to practice their English.

## 5. Discussion

The compulsory implementation of EFL learning across higher education programs calls for research that explores how the students perceive the normalized language requirements and how they react. To explore the construct of language learning investment within this context, the first objective of this study was to design an instrument that allows for the quantifiable examination of the construct of language learning investment in higher education. To this end, a Likert-scale questionnaire was designed following the principles of the measurement theory [49,51] in order to obtain generalizable results. One strength of the questionnaire relates to the inclusion of four dimensions of the construct that have emerged from previous research. While the items are associated with these dimensions of language learning investment, they are not constrained to a particular higher education program. This is because the interest of the present study was directed to the general construct of language learning investment and the dimensions (motivation, necessity, engagement and agency) that may characterize investment among students who are not related to language teaching. Therefore, the operationalization of the dimensions allows researchers to go beyond single cases of English learners and makes the instrument suitable across different higher education programs where EFL learning has become compulsory.

In terms of instrument design, this study contributes to the scientific literature with a quantitative instrument for the exploration of the construct of language learning investment. During its design, the content, construct and convergent validity [13,18] and different reliability procedures instantiate the soundness of the final version of the instrument. The initial version of the instrument included 51 items, and as the validity and reliability of the instrument were examined, almost 30% of the original items were discarded. The final version of the questionnaire included 36 items. During the validation phase, the procedures led to the reformulation or exclusion of various ambiguous items. Moreover, the experts

and the participants commented on the representativeness and appropriateness of the items in the questionnaire [13]. During the reliability analysis phase, the verification of stability of answers, internal consistency and inter-item correlation [46,55] eventually led to the achievement of satisfactory coefficients. In comparison to previous quantitative research that has explored the construct of language learning investment, e.g., [31], the design of our instrument included a wider variety of validity and reliability procedures, yielded higher item exclusion indices and tapped into a larger number of dimensions. The validity and reliability results suggest that homogeneous information can be obtained from a group of participants and this could instantiate the internal and external validity of the results.

The second objective of the study was to explore the dimensions of language learning investment among higher education students who are obligated to meet EFL learning demands. The construct of language learning investment has been explored in different contexts for more than 20 years, and it is established in the field of second language acquisition [19–22]. As it has often been explored among learners in the target language context and language teachers, the current study further expands our understanding of the construct in a very different context and with a different clientele. At this point, there are some theoretical contributions of the study which are worth mentioning. First, the study explores four dimensions that are known to contribute to the construct of language learning investment presented by Norton and have been independently studied: motivation [22,31], necessities [23,33], engagement [14,27] and agency [29,37]. This exploration was achieved following theoretical and methodological procedures that favored the creation of a valid and reliable questionnaire. The results show that, during compulsory EFL learning, our participants realize the importance of EFL learning as they complete their university studies. Moreover, they showed motivation and considered EFL learning to be worth the effort, as Artamonova and Norton have also documented [31,33,34]. Regarding the dimension of necessities, this group of learners acknowledge that knowing the language can nurture their professional development and future working life.

The results indicate that this group of higher education students are motivated and show engagement in their learning. Nevertheless, this engagement is related to the compulsory nature of the classroom activities, and many participants do not enjoy the time they need to spend on EFL learning, expressing a lack of confidence using the language. The results also raise questions about some areas of agency, where students need to show determination and initiative. In this regard, for instance, the students need to envision the potential use of technology to enhance their language learning experience outside of the classroom. Furthermore, they need to work on becoming autonomous learners or expanding their current language learning experience through different kinds of materials [36,57,58]. Thereafter, their level of agency could benefit from greater efforts, initiative, determination and perseverance. This finding substantiates the claim that necessity and engagement are not sufficient for language students to actually become agents of their own learning [24]. This finding also corroborates qualitative insights into the utilitarian value of EFL learning [9]. During compulsory EFL learning in higher education, then, language learning investment may prompt positive mental states (motivation and necessities), but poses greater demands for the implementation of independent and voluntary actions that require perseverance and initiative.

There are methodological limitations that deserve to be acknowledged. One of the objectives of the study was to create a valid questionnaire that could provide generalizable results with respect to the construct of language learning investment in higher education. Regarding validity, it was possible to achieve internal validity, checking that the items represent the construct and there was a relationship between the dimensions. While the results hold representativeness for the language investment condition of our participants, it is not possible to achieve external validity due to the sample size and the lack of diversity of undergraduate programs [13,18]. From the quantitative point of view, to achieve generalizability, the questionnaire should be administered to students from different disciplinary

areas. Results' patterns can be checked across disciplinary areas to explore similarities and differences among the participating student clienteles.

Another methodological aspect that requires attention is the small number of scale-answer choices in the Likert-scale questionnaire. In the current study, the participants were presented with a four-point scale. The exclusion of the neutral point helped to reduce ambiguity in the answers [13,51,52]. Moreover, the use of a small number of points decreased the cognitive load of decision making [13]. In the answer choices, the number of positive and negative options in the scale was balanced. However, these positive aspects of the scale might explain the skewness of the data and thus the absence of normality. Future research should explore whether this skewness remains in the presence of a neutral point or a larger number of scale-answer choices [52]. Studies that attend to these methodological suggestions could provide a wider understanding of language learning investment during compulsory EFL learning in higher education. Our findings and those from future research could inform policy makers and teachers about the potential changes in the educational policies that sanction the learning of English for different disciplines.

## 6. Conclusions

The construct of language learning investment has been explored in qualitative research [10–12] with individual cases or small groups. Moreover, researchers have often collected data from people immersed in the target language context [20–22,33] or pre-service and in-service language teachers [38,39]. Previous qualitative studies have provided some insights into different dimensions that contribute to the construct, such as engagement [14,16], motivation or attitudes [22,31], necessities [21,34] and agency [29]. Nevertheless, there are students in the foreign language context who do not meet these learning conditions, as they are obligated to comply with compulsory EFL education regardless of their future profession. The small number of qualitative studies that have been conducted with this type of learner clientele indicate that learners often associate language learning investment with the necessity to learn English as a foreign language for personal or professional purposes. This paper, then, contributes with a quantitative instrument which explores the construct of language learning investment and some of its dimensions. The quantitative approach helped us to come up with a valid and reliable instrument to explore four dimensions (i.e., motivation, necessity, engagement and agency) that are known to contribute to learning investment. In this study, the group of participants in higher education who are compelled to complete EFL education demonstrated their understanding towards the importance of learning the target language, as in previous studies. However, our findings indicate that there might be contextual and individual factors in their specific area of studies that diminish EFL learning investment. In light of these findings, it is a matter of interest for future studies to determine if a relation exists between the area of professionalization, the context and the dimensions of language learning investment. To address this issue, for instance, mixed methods research can be conducted through the administration of follow-up interviews with learners who exhibit extreme quantitative results in the Likert-scale questionnaire. To further test the representativeness of our findings, in future research, our instrument could be used with students from different areas of professionalization; the results could help researchers identify whether the findings are not context-dependent and favor generalizability among higher education students. The results from this research would be useful for stakeholders to implement strategies and actions that can nurture learning investment during compulsory EFL education.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13040370/s1.

**Author Contributions:** Both authors equally contributed during the conceptualization and realization of the study, as well as during 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 the Escuela Normal de Educación Especial Graciela Pintado de Madrazo (Approval letter number ENEEGPM/071/2021 dated 3 August 2021).

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

**Data Availability Statement:** We are submitting a supplementary file with the data and instruments that support our research.

**Acknowledgments:** We are grateful to the participants and the local institution for their support during the realization of the project. We regret that their names cannot be openly acknowledged due to anonymity and ethical concerns.

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

## References

1. Bourdieu, P. Los tres estados del capital cultural. *Sociológica* **1987**, *2*, 11–17.
2. Juexuan, L.; Lifen, H.; Qi, S. LOTE (Languages Other Than English) learners' investment in learning languages. *Círculo De Lingüística Apl. A La Comun.* **2020**, *84*, 55–64. [CrossRef]
3. Bergson-Shilcock, A. Amplifying Impact—Combining Investment in English Language Skills and Digital Learning. *National Skills Coalition.* 25 June 2020. Available online: https://nationalskillscoalition.org/resource/publications/amplifying-impact/ (accessed on 5 September 2022).
4. Vu, N.; Do, T. Developing capital in language learning: A mixed-method study on Vietnamese English learners. *J. Ethn. Cult. Stud.* **2021**, *8*, 17–39. [CrossRef] [PubMed]
5. Arias, A.; Izquierdo, J. Language attention in content-based instruction: The case of language instructors teaching content in a foreign language in Mexican higher education. *J. Immers. Content-Based Lang. Educ.* **2015**, *3*, 194–217. [CrossRef]
6. Izquierdo, J.; Simard, D.; Garza, M. Multimedia instruction & language learning attitudes: A study with university students. *Rev. Electrónica De Investig. Educ.* **2015**, *17*, 101–115.
7. Izquierdo, J.; Aquino Zúñiga, S.; García Martínez, V. Foreign language education in rural schools: Struggles and initiatives among generalist teachers teaching English in Mexico. *Stud. Second. Lang. Learn. Teach.* **2021**, *11*, 133–156. [CrossRef]
8. Shapiro, H.; Haar, J.; Bayer, I. *Background Paper on Innovation and Education*; Shapiro, H., Boekholt , P., Eds.; 2007; pp. 1–61. Available online: https://www.researchgate.net/publication/267955018_Background_Paper_on_Innovation_and_Education (accessed on 11 July 2021).
9. Diep, L.; Hieu, V. Examining quality of English language learning of university students in Vietnam: The moderating role of competition factor. *Eurasian J. Educ. Res.* **2021**, *95*, 55–79. [CrossRef]
10. Teng, M. Learner identity and learner's investment in EFL learning: A multiple case study. *Iran. J. Lang. Teach. Res.* **2019**, *7*, 43–60. [CrossRef]
11. Tülüce, H. Resilience in foreign language learning: A study on pre-service teachers' language learning experiences. *Int. Online J. Educ. Teach.* **2018**, *5*, 895–908.
12. Yang, S.; Shu, D.; Yin, H. Frustration drives me to grow: Unraveling EFL teachers' emotional trajectory interacting with identity development. *Teach. Teach. Educ.* **2021**, *105*, 103420. [CrossRef]
13. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*, 8th ed.; Routledge: New York, NY, USA, 2018.
14. Angelovska, T.; Mercer, S.; Talbot, K. Personality traits as predictors of language learner engagement. *CercleS* **2021**, *11*, 285–310. [CrossRef]
15. De la Cruz, M.; Resendiz, C.J.; Romero, P.A.; Domínguez, A.G. Adaptation and validation of the student version of the UTRECHT work engagement scale in Mexico. *Psicol. Iberoam.* **2017**, *25*. Available online: https://psicologiaiberoamericana.ibero.mx/index.php/psicologia/article/download/104/275?inline=1 (accessed on 5 November 2022).
16. Soltanian, N.; Ghapanchi, Z.; Razaei, S.; Pishghadam, R. Quantifying Investment in language learning: Model and questionnaire development and validation in the Iranian context. *Issues Lang. Teach.* **2018**, *7*, 25–56. [CrossRef]
17. Yesilcinar, S. Motivational strategies in language learning: Student-teachers' perceptions and views. *Mus Alparslan Univ. Fac. Educ. J.* **2021**, *1*, 41–56.
18. Dörnyei, Z. *Questionnaires in Second Language Research*, 2nd ed.; Taguchi, T., Ed.; Routledge: New York, NY, USA, 2010.
19. Norton Peirce, B. Social identity, investment and language learning. *TESOL Q.* **1995**, *29*, 9–31. [CrossRef]
20. Norton, B.; Nicolaides, C.; Mira, C. Identity and investment in language education: An interview with Bonny Norton. *Calidoscópio* **2020**, *18*, 767–775. [CrossRef]
21. Norton, B.; Toohey, K. Identity, language learning and social change. *Lang. Teach.* **2011**, *44*, 412–446. [CrossRef]
22. Darvin, R.; Norton, B. Investment and motivation in language learning: What's the difference? *Lang. Teach.* **2021**, *56*, 29–40. [CrossRef]
23. Ali, F. Identity and investment in language learning: A case study of heritage Spanish speakers. *Span. Context* **2020**, *18*, 430–458. [CrossRef]
24. Amorati, R. Making languages marketable: An analysis of a promotional brochure for prospective students in an Australian tertiary institution. *Lang. Learn. High. Educ.* **2018**, *8*, 313–331. [CrossRef]

25. Carlioni, G.; Sisti, F. Language learning and inclusion in Italy. *Lang. Learn. High. Educ.* **2019**, *9*, 97–115. [CrossRef]
26. Lightbown, P.; Spada, N. *How Languages Are Learned*, 3rd ed.; Oxford University Press: Oxford, UK, 2006.
27. Hiver, P.; Al-Hoorie, A.; Vitta, J.; Wu, J. Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Lang. Teach. Res.* **2021**, *0*, 13621688211001289. [CrossRef]
28. Mercer, S. Language learner engagement: Setting the scene. In *Second Handbook of English Language Teaching*; Gao, X., Ed.; Springer: Sydney, Australia, 2019; pp. 643–660.
29. Harrison, M.; Uusipaikka, M.; Karinen, A.; Räsänen, T.; Raitala, D.; Ellonen, R.; Tuomela, O. Bridging passion and profession: Supporting agency and investment in multilingual university writers. *Lang. Learn. High. Educ.* **2013**, *3*, 1–25. [CrossRef]
30. Darvin, R.; Norton, B. Identity and a model of investment in applied linguistics. *Annu. Rev. Appl. Linguist.* **2015**, *35*, 36–56. [CrossRef]
31. Artamonova, T. L2 learners' language attitudes and their assessment. *Foreign Lang. Ann.* **2020**, *53*, 807–826. [CrossRef]
32. Taylor, F.; Marsden, E. Perceptions, attitudes and choosing to study foreign languages in England: An experimental intervention. *Mod. Lang. J.* **2014**, *98*, 902–920. [CrossRef]
33. Norton, B. *Identity and Language Learning. Extending the Conversation*, 2nd ed.; Multilingual Matters: Toronto, ON, Canada, 2013.
34. Norton, B. Identity and language learning: A 2019 retrospective account. *Can. Mod. Lang. Rev.* **2019**, *75*, 299–307. [CrossRef]
35. Lacka-Badura, J. Acquiring business knowledge through business English reading materials: Pre-experience students' perspective. *Lang. Learn. High. Educ.* **2021**, *11*, 433–451. [CrossRef]
36. Little, D. Plurilingualism, learner autonomy and constructive alignment: A vision for university language centres in the 21st century. *Lang. Learn. High. Educ.* **2020**, *10*, 271–286. [CrossRef]
37. Naderpour, Z. A contribution of fanfiction writing to improve voice and fluency of learners of Japanese from an agentic perspective. *System* **2022**, *105*, 102723. [CrossRef]
38. Kim, H. Learner investment, identity, and resistance to second language pragmatic norms. *System* **2014**, *45*, 92–102. [CrossRef]
39. Tan, C.P.; Van der Molen, H.; Schmidt, H.G. A measure of professional identity development for professional education. *Stud. High. Educ.* **2017**, *42*, 1504–1519. [CrossRef]
40. Turgut, G. Examining the impact of constructivist and reflective ELT courses on pre-service teachers' beliefs about language learning and teaching: A longitudinal study. *Int. J. Educ. Teach.* **2021**, *8*, 2097–2116.
41. Villegas-Torres, P.; Mora-Pablo, I. The role of language in the identity formation in transnational EFL teachers. *HOW* **2018**, *25*, 11–27. [CrossRef]
42. Hajar, A. Identity, investment and language learning strategies of two Syrian students in Syria and Britain. *Lang. Cult. Curric.* **2017**, *30*, 250–264. [CrossRef]
43. Mora, A.; Trejo, P.; Roux, R. The complexities of being and becoming language teachers: Issues of identity and investment. *Lang. Intercult. Commun.* **2016**, *16*, 182–198. [CrossRef]
44. Dörnyei, Z. *Research Methods in Applied Linguistics. Quantitative, Qualitative, and Mixed Methodologies*; Oxford University Press: New York, NY, USA, 2007.
45. Muijs, D. *Doing Quantitative Research in Education with SPSS*; Sage Publications: Thousand Oaks, CA, USA, 2011.
46. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE: Thousand Oaks, CA, USA, 2013.
47. Larson-Hall, J. *A Guide to Doing Statistics in Second Language Research Using SPSS*; Routledge: New York, NY, USA, 2010.
48. Creswell, J.; Plano, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; SAGE: Thousand Oaks, CA, USA, 2018.
49. Fabila, A.; Minami, H.; Izquierdo, J. La escala de Likert en la evaluación docente. Acercamiento a sus caractarísticas y principios metodológicos. *Perspect. Docentes* **2013**, *50*, 31–40.
50. Iwaniec, J. Questionnaires. Implications for effective implementations. In *The Routledge Handbook of Applied Linguistics*; McKinley, J., Rose, H., Eds.; Routledge: London, UK, 2020; pp. 324–335.
51. Hodge, D.; Gillespie, D. Phrase completion scales. *J. Soc. Serv. Res.* **2007**, *33*, 1–12. [CrossRef]
52. Guy, R.; Norvell, M. The neutral point on a Likert scale. *J. Psychol.* **1977**, *95*, 199–204. [CrossRef]
53. Hernández, M.; Izquierdo, J. Cambios curriculares y enseñanza del inglés: Cuestionario de percepción docente. *Sinéctica. Rev. Electrónica De Educ.* **2020**, *54*, 1–22. [CrossRef]
54. Gravetter, F.; Wallnau, L.; Forzano, L.A.; Witnauer, J. *Essentials of Statistics for the Behavioral Sciences*, 10th ed.; CENGAGE: Boston, MA, USA, 2021.
55. Boyle, J.; Fisher, S. *Educational Testing. A Competence Based Approach*; BPS Blackwell: Malden, MA, USA, 2007.
56. Hinkle, D.E.; Wiersma, W.; Jurs, S.G. *Applied Statistics for the Behavioral Sciences*; Houghton Mifflin Company: Boston, MA, USA, 2003.
57. Tang, E.; Chung, E.; Li, E.; Yeung, S. Independent vocabulary building experience of Hong Kong university students: A pedagogical perspective. *IAFOR J. Educ.* **2016**, *4*, 13–29. [CrossRef]
58. Wang, H.C.; Chen, C.W.Y. Learning English from YouTubers: English L2 learners' self-regulated language learning on YouTube. *Innov. Lang. Learn. Teach.* **2020**, *14*, 333–346. [CrossRef]

education sciences

MDPI

*Article*

# Is Flow Possible in the Emergency Remote Teaching Foreign Language Classroom?

**Jean-Marc Dewaele *, Alfaf Albakistani and Iman Kamal Ahmed**

Department of Applied Linguistics, Birkbeck University of London, London WC1H 0PD, UK; aalbak02@mail.bbk.ac.uk (A.A.); iahmed14@student.bbk.ac.uk (I.K.A.)
* Correspondence: j.dewaele@bbk.ac.uk

**Abstract:** The present study focuses on the experience of flow among 168 Arab and Kurdish English Foreign Language (EFL) learners in both in-person and emergency remote teaching (ERT) classes. Statistical analyses of questionnaire data revealed that learners did experience flow in their ERT classes but for a significantly shorter time than in the pre-pandemic in-person classes. Those who experienced flow in in-person classes were also more likely to experience it in ERT classes. In the in-person classes, the proportion of time in flow was linked to age, self-rated proficiency, attitudes toward English, attitudes toward the teacher, and the teacher frequency of use of English. In contrast, in ERT classes, the proportion of time in flow was only linked to attitude toward the teacher. This is interpreted as evidence that the ERT does not just cause physical and social isolation but also mental isolation.

**Keywords:** flow; emergency remote teaching; in-person teaching; English Foreign Language

**Citation:** Dewaele, J.-M.; Albakistani, A.; Kamal Ahmed, I. Is Flow Possible in the Emergency Remote Teaching Foreign Language Classroom? *Educ. Sci.* **2022**, *12*, 444. https://doi.org/10.3390/educsci12070444

Academic Editors: Martin Howard and Graham Crookes

Received: 23 April 2022
Accepted: 21 June 2022
Published: 28 June 2022



## 1. Introduction

When the COVID-19 pandemic forced everybody to stay home in early 2020 and much of the following year, students found themselves connected to their teachers and peers only through a tenuous electronic link. Students living in areas with regular power outages, such as Iraqi Kurdistan, found themselves at the mercy of electricity companies, as their screens could go black in mid-sentence and a candle had to be lit, hoping that power would return before the end of class.

The rushed, last-minute switch to emergency remote teaching (ERT) made it very different from the well-established delivery of carefully prepared online teaching courses. The whole education sector was instead forced into "a temporary shift of instructional delivery to an alternate delivery mode due to crisis circumstances" [1]. The sudden switch to ERT, in a period of heightened anxiety, grief, and social isolation, affected everybody in the education system [2,3]. There was also a worry that ERT would be insufficient and that students would learn less [4]. The fear was that the sole reliance on online teaching would dampen learners' positive emotions and exacerbate their negative emotions, affecting their learning and coping strategies [5]. Recent research on learner emotions has indeed shown that positive emotions are moderate positive predictors of foreign language (FL) performance, while negative emotions are negative predictors [6,7]. Learners' emotions have been found to be dulled in ERT classes compared to in-person classes. It seems that ERT classes generate much less enjoyment, a little less anxiety, and much more boredom among FL learners [8,9]. Nobody, to our knowledge, has investigated whether learners' flow experience is equally affected by the learning environment.

The concept of flow was introduced by Csíkszentmihályi in 1990, who described it as an optimal psychophysical state that can emerge when the demands of a situation or a task match the skills of the person performing the task. It generates a consciousness that is "harmoniously ordered" where thoughts, actions, and emotions become well-coordinated ([10], p. 6). Research has shown that the emergence of flow in the FL is gradual and increases as learners

become more advanced and more proficient. Experiencing a state of flow is exhilarating and has positive longer-term effects on motivation [11–15] The present study will focus more specifically on the experience of flow in ERT and in-person FL classrooms.

## 2. Literature Review

The literature review is organized into three sections. We start by briefly defining the concept of flow as our dependent variable, and we refer to some key texts in positive psychology. After that, we sketch how positive psychology energized research on learner emotions in applied linguistics and how it led to a more holistic perspective on both positive and negative learner emotions, including research on flow in the foreign language classroom. The final section reports studies that considered the effect of emergency remote teaching contexts on FL Enjoyment (FLE), FL classroom anxiety (FLCA) and FL boredom (FLB).

### *2.1. Flow in Positive Psychology*

Csíkszentmihályi [10] described flow as follows:

> *Concentration is so intense that there is no attention left over to think about anything irrelevant, or to worry about problems. Self-consciousness disappears, and the sense of time becomes distorted. An activity that produces such experiences is so gratifying that people are willing to do it for its own sake, with little concern for what they will get out of it, even when it is difficult, or dangerous.* (p. 71)

Nakamura and Csíkszentmihályi [16] distinguished nine components of flow: (1) challenge-skills balance, (2) clear goals, (3) unambiguous feedback, (4) merging action and awareness, (5) total concentration, (6) feelings of control, (7) transformation of time, (8) loss of self-consciousness, and (9) an autotelic or intrinsically motivating activity.

The first component is undoubtedly the fundamental one: an optimal experience can only arise when the challenge is just right: not overly difficult but not too easy either [16]. People in a state of flow "report feeling more active, alert, concentrated, happy, satisfied, and creative" ([17], p. 816). Being in a state of flow is a gratifying and highly motivating experience [18]. Flow is more likely to emerge in group activities than in solitary activities [19].

### *2.2. Learners' emotions in Positive Psychology*

The introduction by MacIntyre and Gregersen [8] of positive psychology to the field of applied linguistics in 2012 has led to a surge of interest in both positive and negative learner emotions (see [3,20]).

Dewaele and MacIntyre [21,22] were inspired by Csikszentmihalyi's flow theory [10] in developing the concept of foreign language enjoyment which Dewaele and MacIntyre [22] defined as "a complex emotion, capturing interacting dimensions of challenge and perceived ability that can reflect the human drive for success in the face of difficult tasks" (p. 216). They sought to understand its relationship with foreign language classroom anxiety [23].

Dewaele and MacIntyre [21] used an online questionnaire to obtain quantitative and qualitative data from 1746 multilingual FL learners from around the world. Statistical analyses revealed that levels of FLE and FLCA were linked to age, gender, type of institution (universities vs. secondary schools), degree of multilingualism, level of mastery and proficiency in the FL, and wider geographical background (Western vs. Asian). Further research showed that sources of FLE were more likely to be learner-external (mainly the teacher) and to some extent also personality traits such as trait emotional intelligence and cultural empathy. In contrast, FLCA was more strongly linked to learner-internal variables (personality traits such as neuroticism) [9,24]. Analysis of descriptions of enjoyable episodes by participants in Dewaele and MacIntyre [21] showed that FLE was linked to specific classroom activities that allowed a degree of autonomy, peer recognition, a realization of progress, teacher recognition and teacher skills. In a follow-up study using the same

database, Dewaele and MacIntyre [12] compared the proportion of time in a state of flow among EFL learners and FL learners of languages other than English (LOTE). LOTE learners reported spending a significantly higher proportion of class time in a state of flow than the EFL group, but the effect size was small. This was interpreted as evidence of stronger emotional involvement in LOTE classes, where learners often knew English already. Dewaele and MacIntyre's [13] mixed-methods study investigated the relationship between FLE, FLCA, and proportion of class time in flow among 1044 FL learners. The authors found a significant positive correlation between FLE and flow ($r = 0.62$) and a significant negative correlation between FLCA and flow ($r = -0.27$). Flow turned out to be an emergent phenomenon, with beginners reporting lower levels of flow than more advanced learners. Analyses of the enjoyable episodes confirmed these patterns, with advanced learners describing enjoyable experiences that were much more intense, more frequent, and of longer duration than those described by beginners and low-intermediate learners.

Egbert [25] was the first study on flow in second language acquisition. She used seven specific language-learning tasks in a high school Spanish class with 13 Anglophone students. She found that causality in flow and performance is bi-directional: an appropriate balance between challenge and skill can support flow which can lead to improved performance. The challenge–skill balance emerged as a crucial aspect of flow experiences of 36 Japanese learners of English in [26,27]. Researchers have also investigated different types of flow. Czimmermann and Piniel [28] were surprised to find that task flow and general classroom flow of their 85 Hungarian EFL students were only moderately positively correlated. Anxiety, boredom and apathy were found to be significantly negatively correlated with task-specific flow. The importance of the presence of the group was highlighted in Rubio [29]. He found that his 29 Spanish EFL learners were more likely to experience flow in group-work tasks that encouraged learner agency and a degree of autonomy in performing the task. Following a similar avenue, Liu and Song [30] found that flow antecedents (skill, challenge, and clear goals) determined flow experience among their 235 Chinese EFL students engaged in challenging online learning activities. Beneficial long-term effects of flow on motivation were found in Piniel and Albert's [14] study on 214 Hungarian EFL students suggesting that they can be mutually reinforcing [15].

In an attempt to create a better conceptualization and measurement of flow and anti-flow in a blended learning environment, Wang and Huang [31] developed a new instrument, the Foreign Language Flow Scale. The authors collected data from 661 Chinese EFL learners. Their analyses showed that FL learning flow is a three-dimensional construct involving FLE, FLCA, and FLB. They found that FLE is a core component of flow and is a predictor for FL achievement. FLB and FLCA were linked to negative flow.

Adopting an innovative neurophysiological approach, Nozawa et al. [32] discovered a significant link between inter-brain synchronization and interpersonal similarity of flow state dynamics during collaborative learning process by 56 Japanese EFL learners. Students who had been working in the same group showed significantly more inter-brain synchronization than members of other groups.

### *2.3. Learner Emotion Research in Emergency Remote Teaching Contexts*

Researchers have looked into the sources of FL learners' emotions in in-person and ERT classes as a consequence of the COVID-19 pandemic.

Li and Dewaele [33] found that their 348 Chinese EFL learners' foreign language boredom (FLB) was linked to the perceived meaningfulness of engaging in ERT courses and their degree of confidence. Levels of FLB were higher in the ERT course compared to the in-person course, but the causes of FLB were broadly similar to those in in-person classes [34]. Students felt that the ERT course was too time-consuming, meaningless, too socially detached, and resulted in low test results.

Combining an online questionnaire ($n = 52$) and interviews ($n = 16$), Kohnke, Zou, and Zhang [35] looked at learner emotions, self-regulated strategies, and perceived difficulties of Chinese FL students in an ERT setting. The authors found that students enjoyed attending

the ERT classes and felt that they were able to develop their L2 skills online. However, the amount of time needed and the workload, combined with the lack of experience in ERT, caused them to experience stress, doubt, and loneliness. They regretted the absence of group work and social interaction and developed coping strategies to interact with their peers.

Adopting a similar approach, Resnik and Dewaele [9] found that 510 European EFL students experienced both significantly less FLE and FLCA in ERT classes compared to in-person classes. The increased teacher-centeredness of ERT classes limited the number of opportunities to engage with peers, which led to increased boredom. Enjoyment suffered in the ERT classes because of reduced group solidarity, little laughter, and more detached relationships with teachers and peers. The lower anxiety in ERT classes was linked to the ability to turn the camera off and to avoid participation.

Following this path, Resnik, Dewaele, and Knechtelsdorfer [36] explored differences in FLE in ERT and in-person EFL classes among 437 university EFL learners. FLE levels were found to be significantly lower in ERT classes than in in-person classes. Interviews revealed that the sources of FLE differed in both conditions: interaction with teachers and peers, group solidarity, and a fun atmosphere drove FLE in in-person classes; while the home comfort and convenience of not having to travel combined with increased autonomy were mentioned as sources of FLE in ERT classes. They did point out that social interactions with peers and teachers were possible in the ERT context too.

Resnik, Dewaele, and Knechtelsdorfer [37] used the same dataset to focus on FLCA. Overall, FLCA was found to be lower in the ERT context. Interviews revealed that the sources of FLCA in ERT classes were different from the sources of FLCA in regular classes. Being forced to contribute to class discussions was the most frequently mentioned cause of FLCA in in-person classes, while technological and connection issues were the main source of FLCA in ERT, followed by uncertainty about requirements for assignments and worries about the resulting grade.

Pursuing a similar approach, the mixed-method study by Maican and Cocarada [38] focused on FLE and FLCA among 207 Romanian FL university students during the pandemic. Participants were found to appreciate online resources offered in their ERT classes but complained about time-consuming tasks in ERT and resented the physical separation from peers and teachers. They had developed coping behaviors using online resources, which led to higher FLE.

A more in-depth analysis of the reasons for FLB was presented in Pawlak et al.'s [39] qualitative study that included 34 teachers and 256 Iranian EFL students. It showed that a majority of both teachers and students reported that in-person classes were less boring than online classes due to their lecture-type nature. Ineffective coping strategies such as playing games or disconnecting left students bored in the online mode.

Some researchers have also looked at the fluctuation of emotions during ERT classes. A small-scale longitudinal study by Sun and Zhang [40] mapped out the emotional trajectory of 11 Chinese EFL students during ERT classes. Their emotions fluctuated from feeling very anxious at the start and end of the ERT course, with a calm period in the middle. The absence of cooperation with peers during tasks weighed on students.

Finally, in a mixed-methods study based on the same database as the current one, Dewaele, Albakastani and Kamal Ahmed study [41] compared levels of FLE, FLCA and FLB among Arab and Kurdish EFL learners in both in-person and ERT classes. Levels of FLE and FLCA were significantly higher in in-person classes than in ERT classes, while FLB was higher in ERT classes. Qualitative data revealed that learners felt more isolated, disengaged, and distracted in the ERT context. However, some felt that ERT did allow relationship-building, lowered their FLCA, and encouraged them to develop new coping strategies. FLB resulted in a lack of exciting social interactions and monotonous delivery by the teacher. In a follow-up study, Dewaele, MacIntyre, Albakastani and Kamal Ahmed [42] found that FLE was a significant positive predictor of flow while FLB was a significant negative predictor. FLCA did not predict any unique variance in flow.

This literature review has shown that the concept of flow is well established in applied linguistic research, as well as the sources of variation in FLE, FLCA, FLB, and flow. A clear network of relationships has been identified between emotions and flow in in-person FL classes. Given that the pandemic started only two years ago, research is only just emerging about the effect of ERT on FL learner emotions. So far, to our knowledge, no research has been published on the effect of ERT on flow. The present study thus aims to fill this gap. We have also decided to focus on EFL learners from Arab countries, as this is a population that is underrepresented in FL emotion research so far.

Based on the abovementioned literature and the gap it showed, the current study aims to investigate the following research questions:

1. Do students spend a larger proportion of time in flow in in-person than in ERT classes?
2. Are students who spend a larger proportion of time in flow in in-person classes also more likely to experience flow in ERT classes?
3. Are learner-internal and learner-external variables similarly linked to proportion of time in flow in both conditions?

## 3. Methodology

### *3.1. Participants*

Participants were 168 learners (111 females, 53 males, 4 participants did not respond) studying EFL in Arab countries and the Kurdistan region in Iraq. Participants were Moroccan ($n$ = 58), Iraqi Kurds ($n$ = 53), Saudi ($n$ = 46) with smaller numbers of twelve other nationalities. Participants were aged between 16 and 38 years ($M$ = 20, $SD$ = 3). The majority of the respondents were studying at universities ($n$ = 152), while a few were school students ($n$ = 15). Most of the participants studied English in their local countries, and all had moved to ERT due to the pandemic. English proficiency levels varied, with 5 describing themselves as advanced beginners, 19 participants as low intermediates, 65 participants as intermediate, 56 as high intermediates, and 23 as advanced. Most participants reported knowing two languages ($n$ = 64), fewer spoke three ($n$ = 49), four ($n$ = 45) and five to seven languages ($n$ = 10).

Participants' L1 included Arabic ($n$ = 93), Kurdish ($n$ = 39), Amazigh ($n$ = 17), Turkmen ($n$ = 6), English ($n$ = 4), French ($n$ = 3), and other languages. All spoke Arabic and English.

### *3.2. The Instruments*

The data was collected once through an online questionnaire while all the teaching happened remotely. The first section included demographic questions and a language profile, as reported above. Self-reported proficiency scores were converted in a 5-point Likert scale. They also filled out items on their attitude toward English, their attitude toward the English teacher, and the teacher's frequency of use of English in class. They reported their latest test results (in %). All these variables have been found to be linked to learner emotions [43]. The mean scores and standard deviation are reported in Table 1.

**Table 1.** Means and SD of the independent variables.

| Variable | Mean | Standard Deviation |
|---|---|---|
| English Proficiency | 3.43 | 0.96 |
| Attitude toward English | 4.20 | 0.87 |
| Attitude toward English teacher | 4.11 | 1.00 |
| Teacher's frequency of English Use | 4.64 | 0.78 |
| Test Result | 82.67 | 17.50 |

The participants were then asked to rate their emotions (not included in the present study) and the time spent in a state of flow, both pre-pandemic in their in-person classes and in their ERT setting. They filled out items extracted from Larson and Csíkszentmihályi's [44] *Experience Sampling Method* and used in Dewaele and MacIntyre [12,13]. The items were

preceded by the question: "What percentage of EFL class time does this apply to you (ranging from 0%-never to 100%-always)? During my EFL class . . . ". After that, they were presented with the following four items:

(1) I lose sense of time . . . %
(2) I'm totally absorbed . . . %
(3) I feel fulfilled . . . %
(4) I'm happy . . . %

Taken together, the four items allow the calculation of the proportion of class time in a state of flow. The internal consistency is satisfactory for both In-person and ERT classes (Cronbach $\alpha = 0.843$, $N = 4$; and Cronbach $\alpha = 0.823$, $N = 4$ respectively).

### 3.3. Procedure

The questionnaire was designed using Google Forms. Snowball sampling was used: an open call was issued to Arab and Kurdish learners who were studying EFL in Arab countries during the pandemic, asking them to share the link with their friends and classmates. The survey was accessible from April to June 2021, and took between 15–20 min to complete. Anonymity of participants' responses was guaranteed, and their consent was obtained at the start of the survey. The research design received ethical approval from the authors' institution.

### 3.4. Data Analysis

The calculation of Q-Q plots suggests that values for the proportion of class time in a state of flow follow a normal distribution reasonably well, except for the extreme tails (available from the authors on request). We thus opted for parametric statistics.

## 4. Results

A paired *t*-test was used to answer the first research question. It revealed that participants reported spending a larger proportion of time in a state of flow when they attended their FL classes in-person (*Mean* = 58.16, *SD* = 28.41) rather than online (*Mean* = 46.49, *SD* = 28.00). The difference was significant ($t(167) = 4.99$, $p < 0.001$, Cohen's $d = 0.31$). According to Plonsky and Oswald ([45], p. 889) this is a small effect size. To gain a better understanding of the difference, a boxplot was created (see Figure 1).

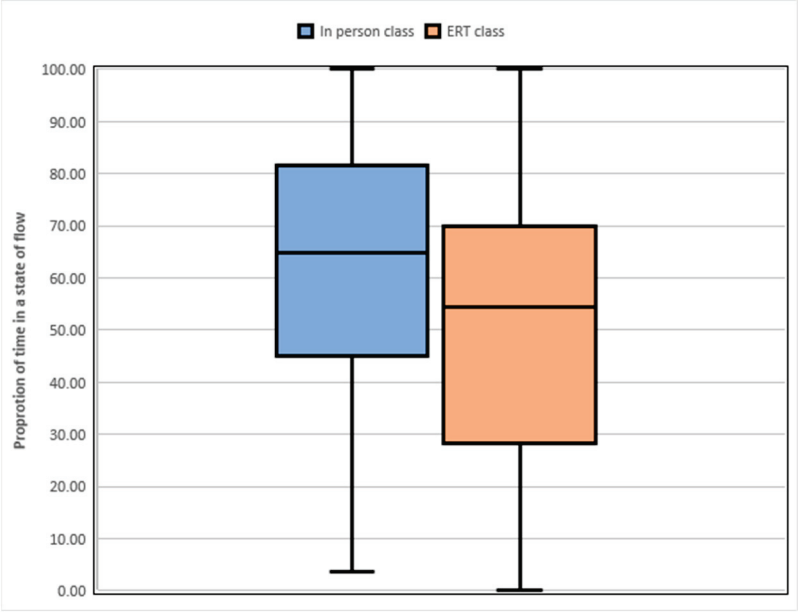



**Figure 1.** Proportion of time in flow in In-person and ERT classes.

Looking at the differences between the box plots, it is striking that the median is much higher for the in-person condition than the ERT condition, in contrast with the means that were much closer to each other. The interquartile range box is smaller the in-person condition, suggesting a slightly more limited spread. The whiskers suggest that the bottom 25% of participants were more widely spread out in the in-person condition than in the ERT condition, where the top and bottom whisker are equal in length. Thus, it seems that the dispersion was greater in the ERT condition.

A two-tailed Pearson correlation analysis was used to answer the second research question on the relationship between proportion of time in flow in in-person and ERT classes. A very strong positive correlation emerged: $r(168) = 0.681$, $p < 0.0001$. In other words, both sets of values share 46.37% of variance. According to Plonsky and Oswald [45], this represents a large effect size.

The third research considered the relationship between learner-internal and learner-external variables and the proportion of time in flow in both conditions. Firstly, an independent t-test showed no gender effect on proportion of time of flow in in-person and ERT classes ($t(162) = -0.59$, $p = ns$ and $t(162) = -1.64$, $p = ns$. Secondly, a series of two-tailed Pearson correlation analyses revealed that five independent variables were linked with proportion of time of flow in in-person classes, but only one was linked significantly to proportion of time in flow in ERT classes (see Table 2). Younger learners, learners who rated themselves as proficient, with positive attitudes toward English and their teacher, and whose teacher used English frequently were likely to spend a longer proportion of time in a state of flow in in-person classes. However, in the ERT context, only those with positive attitudes toward the English teacher were likely to spend more time in a state of flow. Comparing effect sizes in both conditions, it appears that those for in-person classes were small-to-medium, while the effect size for the only independent variable in the ERT context was small [45].

**Table 2.** Pearson correlations between independent variables and proportion of time in flow in In-person and ERT classes.

| Variable | In-Person | ERT |
|---|---|---|
| Age | −0.177 * | −0.05 |
| Number of Languages | 0.119 | 0.019 |
| English Proficiency | 0.159 * | 0.124 |
| Test Result | −0.073 | 0.002 |
| Attitude toward English | 0.226 ** | 0.048 |
| Attitude toward English teacher | 0.390 ** | 0.178 * |
| Teacher's frequency of English Use | 0.315 ** | 0.137 |

** $p < 0.01$ (2-tailed); * $p < 0.05$ (2-tailed).

## 5. Discussion

The first research question focused on the effect of teaching modality (in-person classes versus ERT classes) on the emergence of flow. Participants reported spending a significantly higher proportion of time in a state of flow in in-person classes compared to ERT classes. The effect size is small but it fits the pattern reported in previous studies [33,41,42] where levels of FLE were found to be significantly higher and levels of FLB lower in in-person classes than in ERT classes. It is possible that an ERT environment and the physical separation from peers and teachers gen46erate a weaker social interdependence, making it harder for group members to reach a state of flow [19]. The finding that our participants spent close to two thirds of their in-person FL classes in a state of flow, and over half of their ERT FL classes in a state of flow, corresponds very closely to the values reported by Dewaele and MacIntyre [13], where a similar method of measurement showed that participants reported being in a state of flow close to 60% of the class time. Researchers who used different ways of measuring flow found broadly similar values [25,29,43]. The finding that the proportion of time in a state of flow in

ERT classes is lower than in-person class is not surprising, as it fits with the patterns for learner emotions in both settings. Lower FLE, slightly lower FLCA and higher FLB in the ERT setting compared to in-person confirms a general feeling of social and emotional detachment [9,35,36,44]. Dewaele et al. [41] also found that many students were worried about the stability of the internet connection and the fear of being literally left in the dark. They also complained about their own disengagement, distraction, and feelings of isolation during their ERT classes. However, adopting a more positive view, we could rejoice that even while physically distant, learners still spent half their time in a state of flow, on average, in front of their computer screen. In other words, even in an ERT class, it is possible to reach a good challenge-balance with clear goals [16,26,27,30].

One possible explanation for the longer time in flow in the in-person classes, and the slightly lower dispersion around the median, is that being in the same room as the teacher and peers means full physical immersion in the class activities, breathing, joking, struggling, and laughing together and co-constructing the classroom activities under the watchful eye of the teacher while developing social relationships [38,40]. It also implies that in-person classes have fewer sources of distraction that can potentially break the learner's concentration compared to ERT classes. This is probably due to the teacher's physical presence in the classroom and the constant observation of all students. In other words, the teacher can immediately counter comments or behaviors that would disrupt the flow experience and impede attaining the objectives of that class. The joint classroom experience includes reactions to unexpected things that happen inside or outside the classroom, ranging from the bird that flies in through the open window to sudden commotion in the corridor. In contrast, the online class may well occupy only a small part of the computer screen, with microphone muted, loudspeaker on minimal volume and the camera turned off, as the student may simultaneously be chatting online with friends, answering phone calls, having music in the room, or a family member walking in bringing food. In addition, unexpected funny events, such as a bird suddenly perching on the computer screen, are not joint experiences that strengthen group cohesion, but rather isolated distractions that cannot benefit the student's social capital. In other words, it is slightly harder to be 'in the zone' when performing a task, and to remain in that state, when the teacher's voice and the occasional voice of peers come through a tiny loudspeaker and their faces are mere two-dimensional thumb-sized blots on the screen ([38], p. 38). The temptation to play games or to disconnect is probably also playing in the back of the minds of bored students [39]. It is also less likely that pair work is as exciting through an internet connection, and brain synchronization between learners in different locations is less likely to occur [32]. Finally, the stress linked to the pandemic, the isolation, and potentially the grief for having lost loved ones may have weighed on their emotions and prevented them from reaching a state of flow as easily as before in in-person classes.

The finding that learners who reported spending more time in a state of flow in in-person classes also spent more time in online classes strengthens the validity of the instrument and measure. It suggests that the instrument taps into something real that exists in different conditions. It could be argued that those who had experienced flow in In-person classes were best able to reach that same state again in online classes, even if it was for shorter periods of time. As was pointed out before, being able to coordinate thoughts, actions, and emotions when reaching a state of flow is not only highly satisfying; it is also motivating, addictive, and likely to lead to accelerated learning and better performance [13–15].

The answer to the third research question on the role of learner-internal and learner-external variables on the proportion of time in a state of flow could also shed light on the first research question. The finding that multiple independent variables were linked to flow in in-person classes but that only a single independent variable was linked to flow in ERT classes was unexpected. Why would age, self-rated proficiency, attitudes toward English, attitudes toward the English teacher, and the teacher's frequency of use of English frequently be linked to the flow of in-person classes, but only the attitude toward

the English teacher be (less strongly) linked to flow in ERT classes? The finding of these relationships in the in-person classes converges with the findings of the studies that looked at the effects of these variables on FLE and FLB [21,33,43]. Why then would all these variables no longer have an effect in the ERT classes? It could be argued that sitting in the classroom among peers, with the teacher standing in front of a well-filled blackboard, with iconic posters of Big Ben, the Capitol, Sydney Harbour, and Cape Town, feeds a connection with the whole wide world. Every student in the classroom is co-constructing the teaching event by their mere presence. However, sitting in one's bedroom watching the teacher's face and the teaching material on the computer screen might be sufficient to reach a state of flow in carrying out the required activities, but it will be shorter in duration and the feeling of co-construction of the event will be absent. ERT classes are more likely to be perceived as emotionally disembodied [9] and learners may feel mentally disconnected from the rest of world.

The present study is not without limitations. Firstly, a cross-sectional design meant we only had a "snapshot" of a dynamic situation as learners' emotions were likely to change over time, especially with the ERT context becoming more commonplace during the pandemic. Also, the quantitative design does not allow us to pinpoint specific causes for the patterns we observed. Only interviews with participants about their flow experiences in both conditions could allow us to throw a light on reasons why flow was more difficult to sustain in front of the computer at home. We are aware that the modest sample size does not allow us to make sweeping generalizations. The patterns extracted apply only to our population, but they seem to be broadly in line with previous research. Finally, the questionnaire did not inquire whether the participant had experience with blended learning, which may have made a difference in the transition to ERT [46].

Further research could also adopt a more granular approach, focusing on the contextual factors that could have an effect on flow in a specific classroom, including the temperature or humidity in the classroom, the time of the day, the smile (or absence of smile) on the face of the teacher, the tasks at hand, and the group cohesion at that moment in time: did students fight or laugh in the previous class? For research on flow in the ERT setting, it would be worth exploring whether certain activities were linked to a higher level of flow, whether more autonomy through breakout-room activities may have made a difference as well as the cameras being switched on or off during class.

The pedagogical implications of the present study are relatively limited as this was not an intervention study but a cross-sectional study using a correlational design that does not allow for determining causality. The historical circumstances did allow an unexpected comparison of the experience of flow in pre-pandemic in-person classes with those of pandemic ERT classes. The finding that flow can occur in both conditions will be a relief for teachers who had to adapt their teaching methods overnight in extremely challenging circumstances. The absence of relationships between background variables and proportion of time of flow in the ERT condition raises uncomfortable questions about the future. Will teachers be able to reconnect their learners in post-pandemic in-person classes with the rest of the world? Will they be able to strengthen the social bonds between their learners and themselves? Will the experience of flow be predicted by more than just the attitude toward the teacher?

## 6. Conclusions

Depending on whether one is more of an optimist or more of a pessimist, the findings of this study could be interpreted as a glass being half full or a glass being half empty. The finding that flow does occur in ERT foreign language classes-albeit for shorter timespans-is a positive finding because it means that the world-wide efforts of foreign language teachers to keep their classes going and to keep their students engaged despite the unfamiliar software and the challenging conditions were largely successful. The pessimist might acknowledge that it would have been much worse if the pandemic had hit thirty years earlier when fewer students had computers and access to the internet. Yet, the pessimist

would also point to the fact that while ERT was better than nothing, it did not quite replace the real thing, namely the hustle and bustle of an exciting classroom full of peers and a teacher where flow is very much a group experience to which everybody contributes. Lower levels of FLE and higher levels of FLB in the ERT condition [41] might also explain shorter periods in flow.

The original findings of the current study are that, firstly, the teaching condition (in-person or ERT) affects learners' time in flow. Secondly, being locked at home in front of a computer meant that learners were not just physically and socially isolated, but also mentally closed. The phrase "no human being is an island" comes to mind. Learners need to be part of a community in order to thrive, and the pandemic and the resulting ERT tested everybody's resilience as never before (and hopefully never after). Learning is harder on a small island, even with an internet connection.

**Author Contributions:** Conceptualization, J.-M.D.; methodology, J.-M.D., A.A. and I.K.A.; formal analysis, J.-M.D., A.A. and I.K.A.; investigation, J.-M.D., A.A. and I.K.A.; resources, J.-M.D., A.A. and I.K.A.; data curation, J.-M.D., A.A. and I.K.A.; writing—original draft preparation, J.-M.D., A.A. and I.K.A.; writing—review and editing, J.-M.D., A.A. and I.K.A. 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 according to the guidelines of the Declaration of Helsinki, and approved by the School of Social Sciences, History and Philosophy, Ethics Committee, Birkbeck, University of London (approval date: 15 June 2020).

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

**Data Availability Statement:** Not applicable.

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

## References

1. Hodges, C.; Moore, S.; Lockee, B.; Trust, T.; Bond, A. The Difference between Emergency Remote Teaching and Online Learning. 2020. Available online: https://er.educause.edu/%20articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning (accessed on 27 February 2022).
2. Fraschini, N.; Tao, Y. Emotions in online language learning: Exploratory findings from an ab initio Korean course. *J. Multiling. Multicult. Dev.* **2021**, *2*, 1–19. [CrossRef]
3. MacIntyre, P.D.; Gregersen, T.; Mercer, S. Setting an agenda for Positive Psychology in SLA: Theory, practice, and research. *Mod. Lang. J.* **2019**, *103*, 262–274. [CrossRef]
4. Engzell, P.; Frey, A.; Verhagen, M.D. Learning loss due to school closures during the COVID-19 pandemic. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022376118. [CrossRef]
5. Marchand, G.; Gutierrez, A.P. The role of emotion in the learning process: Comparisons between online and face-to-face learning settings. *Internet High. Educ.* **2012**, *15*, 150–160. [CrossRef]
6. Botes, E.; Dewaele, J.M.; Greiff, S. The Foreign Language Classroom Anxiety Scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *J. Psychol. Lang. Learn.* **2020**, *2*, 26–56. [CrossRef]
7. Botes, E.; Dewaele, J.M.; Greiff, S. Taking stock: A meta-analysis of the effects of Foreign Language Enjoyment. *Stud. Second. Lang. Learn. Teach.* **2022**, *12*, 205–232. [CrossRef]
8. MacIntyre, P.; Gregersen, T. Emotions that facilitate language learning: The positive-broadening power of the imagination. *Stud. Second. Lang. Learn. Teach.* **2012**, *2*, 193–213. [CrossRef]
9. Resnik, P.; Dewaele, J.-M. Learner emotions, autonomy and trait emotional intelligence in 'In-person' versus emergency remote English foreign language teaching in Europe. *Appl. Linguist. Rev.* **2021**. [CrossRef]
10. Csíkszentmihályi, M. *Flow: The Psychology of Optimal Experience*; Harper Collins: New York, NY, USA, 1990.
11. Dewaele, J.-M.; MacIntyre, P.D. El flujo en el aula de Español como lengua extranjera. In *Factores cognitivos y afectivos en la enseñanza del español como LE/L2*; Méndez Santos, M.D.C., Duñabeitia, J.A., Eds.; Thomson Reuters Aranzadi: Cizur Menor, Spain, 2021; pp. 303–330.
12. Dewaele, J.-M.; MacIntyre, P.D. Do Flow, Enjoyment and Anxiety emerge equally in English Foreign Language Classrooms as in other Foreign Language Classrooms? *Rev. Bras. De Linguística Apl.* **2022**, *22*, 156–180. [CrossRef]
13. Dewaele, J.-M.; MacIntyre, P.D. "You can't start a fire without a spark". Enjoyment and the Emergence of Flow in Foreign Language Classrooms. *Appl. Linguist. Rev.* **2022**. [CrossRef]

14. Piniel, K.; Albert, Á. L2 motivation and self-efficacy's link to language learners' flow and anti-flow experiences in the classroom. In *UZRT 2016 Empirical Studies in Applied Linguistics*; Krevelj, S.L., Geld, R., Eds.; FF Press: Zagreb, Croatia, 2017; pp. 90–103.
15. Piniel, K.; Albert, Á. *Motivation and Flow. The Palgrave Handbook of Motivation for Language Learning*; Lamb, M., Csizér, K., Henry, A., Ryan, S., Eds.; Palgrave Macmillan: Basingstoke, UK, 2019; pp. 579–597.
16. Nakamura, J.; Csíkszentmihályi, M. Flow theory and research. In *Oxford Handbook of Positive Psychology*; Lopez, S.J., Snyder, C.R., Eds.; Oxford University Press: Oxford, UK, 2009; pp. 195–206.
17. Csikszentmihalyi, M.; LeFèvre, J. Optimal experience in work and leisure. *J. Personal. Soc. Psychol.* **1989**, *56*, 815–822. [CrossRef]
18. Schüler, J. Flow experience and learning. In *Encyclopedia of the Sciences of Learning*; Seel, N.M., Ed.; Springer: Boston, MA, USA, 2012; pp. 1304–1305.
19. Walker, C.J. Experiencing flow: Is doing it together better than doing it alone? *J. Posit. Psychol.* **2010**, *5*, 3–11. [CrossRef]
20. Dewaele, J.-M.; Chen, X.; Padilla, A.M.; Lake, J. The flowering of positive psychology in foreign language teaching and acquisition research. *Front. Psychol.* **2019**, *10*, 2128. [CrossRef]
21. Dewaele, J.-M.; MacIntyre, P.D. The two faces of Janus? Anxiety and enjoyment in the Foreign Language classroom. Studies. *Second. Lang. Learn. Teach.* **2014**, *4*, 237–274. [CrossRef]
22. Dewaele, J.-M.; MacIntyre, P.D. Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of the language learner? In *Positive psychology in SLA*; MacIntyre, P.D., Gregersen, T., Mercer, S., Eds.; Multilingual Matters: Bristol, UK, 2016; pp. 215–263.
23. Horwitz, E.K.; Horwitz, M.B.; Cope, J. Foreign-Language Classroom Anxiety. *Mod. Lang. J.* **1986**, *70*, 125–132. [CrossRef]
24. Dewaele, J.-M.; MacIntyre, P.D. The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. In *Evidence-based second language pedagogy*; Sato, M., Loewen, S., Eds.; Routledge: London, UK, 2019; pp. 263–286.
25. Egbert, J. A study of flow theory in the foreign language classroom. *Mod. Lang. J.* **2003**, *87*, 499–518. [CrossRef]
26. Aubrey, S. Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Lang. Teach. Res.* **2017**, *21*, 717–734. [CrossRef]
27. Aubrey, S. Measuring flow in the EFL Classroom: Learners' perceptions of inter- and intra-cultural task-based interactions. *TESOL Q.* **2017**, *51*, 661–692. [CrossRef]
28. Czimmermann, E.; Piniel, K. Advanced language learners' experiences of flow in the Hungarian EFL classroom. In *Positive Psychology in SLA*; Gregersen, T., Mercer, S., MacIntyre, P.D., Eds.; Multilingual Matters: Bristol, UK; pp. 193–214.
29. Rubio, F.D. Optimal experiences in the foreign language classroom: Flow states in speaking tasks. *Angl. Int. J. Engl. Stud.* **2011**, *22*, 63–80.
30. Liu, H.; Song, X. Exploring "Flow" in young Chinese EFL learners' online English learning activities. *System* **2021**, *96*, 102425. [CrossRef]
31. Wang, X.; Huang, C. Conceptualization and measurement of flow in a Chinese blended English as a foreign language learning context. *Front. Psychol.* **2022**, *13*, 839267. [CrossRef] [PubMed]
32. Nozawa, T.; Kondo, M.; Yamamoto, R.; Jeong, H.; Ikeda, S.; Sakaki, K.; Miyake, Y.; Ishikawa, Y.; Kawashima, R. Prefrontal inter-brain synchronization reflects convergence and divergence of flow dynamics in collaborative learning: A pilot study. *Front. Neuroergon.* **2021**, *2*, 686596. [CrossRef]
33. Li, C.; Dewaele, J.-M. The predictive effects of trait emotional intelligence and online learning achievement perceptions on foreign language class boredom among Chinese university students. *Foreign Lang. Foreign Lang. Teach.* **2020**, *5*, 33–44.
34. Li, C. A Control–Value Theory approach to boredom in English classes among university students in China. *Mod. Lang. J.* **2021**, *105*, 317–334. [CrossRef]
35. Kohnke, L.; Zou, D.; Zhang, R. Pre-service teachers' perceptions of emotions and self-regulatory learning in Emergency Remote Learning. *Sustainability* **2021**, *13*, 7111. [CrossRef]
36. Resnik, P.; Dewaele, J.-M.; Knechtelsdorfer, E. Differences between tertiary-level learners' enjoyment in on-site and emergency remote EFL classes during the pandemic: A mixed-methods study. Unpublished paper. 2022, *online ahead of print*.
37. Resnik, P.; Dewaele, J.-M.; Knechtelsdorfer, E. Differences in foreign language anxiety in regular and online EFL classes during the pandemic: A mixed-methods study. *TESOL Q.* 2022, *online ahead of print*.
38. Maican, M.-A.; Cocoradă, E. Online foreign language learning in higher education and its correlates during the COVID-19 pandemic. *Sustainability* **2021**, *13*, 781. [CrossRef]
39. Pawlak, M.; Derakhshan, A.; Mehdizadeh, M.; Kruk, M. Boredom in online English language classes: Mediating variables and coping strategies. *Lang. Teach. Res.* **2022**. [CrossRef]
40. Sun, J.; Zhang, X. Exploring Chinese college students' emotions as they engage in online learning during a pandemic. *Asia Pac. J. Educ.* **2021**. [CrossRef]
41. Dewaele, J.-M.; Albakastani, A.; Kamal Ahmed, I. Levels of Foreign Language Enjoyment, Anxiety and Boredom in Emergency Remote Teaching and in in-person classes. *Lang. Learn. J.* 2022, *online ahead of print*.
42. Dewaele, J.-M.; MacIntyre, P.D.; Albakistani, A.; Ahmed Kamal, I. The Foreign Language learner emotions that shape the flow channel. 2022, *online ahead of print*.
43. Dewaele, J.-M.; Witney, J.; Saito, K.; Dewaele, L. Foreign language enjoyment and anxiety in the FL classroom: The effect of teacher and learner variables. *Lang. Teach. Res.* **2018**, *22*, 676–697. [CrossRef]

44. Larson, R.; Csíkszentmihályi, M. Experience Sampling Method. In *Flow and the Foundations of Positive Psychology*; Csíkszentmihályi, M., Ed.; Springer: Dordrecht, The Netherlands, 2014; pp. 21–34.
45. Plonsky, L.; Oswald, F.L. How big is 'big'? Interpreting effect sizes in L2 research. *Lang. Learn.* **2014**, *64*, 878–912. [CrossRef]
46. Ruiz, L.M.S.; Moll-López, S.; Moraño-Fernández, J.A.; Llobregat-Gómez, N. B-Learning and Technology: Enablers for University Education Resilience. An Experience Case under COVID-19 in Spain. *Sustainability* **2021**, *13*, 3532. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

*Education Sciences* Editorial Office
E-mail: education@mdpi.com
www.mdpi.com/journal/education

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-1117-5