Investigating the Effects on Language-learning Outcomes of Phonics Instruction in Foreign Language Teaching: A Systematic Review

Katina Helena Gronowski

University of Oxford

Department of Education

MSc in Applied Linguistics and Second Language Acquisition

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my supervisor, Catherine Hamilton, as she guided me throughout this entire process. I will forever cherish the opportunity to learn from her incredible knowledge and experiences.

Secondly, I wish to thank my family for their endless support. I would not have had the courage nor discipline to attend Oxford if not for their love and support. I dedicate this dissertation to my parents who have always shown me that anything is possible, and that knowledge is power. My community in Hawai'i also showed me immense support as I embarked on this great academic journey.

Furthermore, I wish to thank the friends I made throughout this journey. Oxford would not be the place that it is without all of the amazing and kind people who make up this community. I will forever cherish the friendships I made and the experiences I had.

Lastly, I wish to thank Wolfson College for providing the most amazing atmosphere to learn and prosper. I am especially grateful for the Spring cows whom I spent many afternoons sitting with in the meadows enjoying their company.

Ma ka hana 'ike, Ma ka 'ike ka mana

Through work comes knowledge, through knowledge comes power.

Abstract 7
Chapter 1: Introduction and Literature Review 8
1.1 Introduction and Rationale 8
1.1.2 Rationale 9
1.2 Phonics: Definitions and Approaches 10
1.2.1 What is phonics? 10
1.2.2 Phonological Awareness 10
1.2.3 Key Approaches 11
1.2.4 Phonics in Education- US and UK contexts 12
1.2.5 Foreign Language Learning 13
1.3 Theoretical Background 14
1.3.1 Phonics in Language Learning Contexts 14
1.4 Language- Learning Outcomes 15
1.5 Current Research in Phonics Instruction 16
1.6 Summary and Research Questions 17
Chapter 2: Methodology 18
2.1 Eligibility Criteria 18
2.2 Information Sources 21
2.3 Search Strategy 22
2.3.1 Citation chaining 24
2.4 Data Management 24
2.5 Selection Process 25
2.5.1 Initial Screening: 25
2.5.2 Quality Assurance 25
2.5.3 Full-text Screening 26
2.6 Data Collection Process 26
2.7 Risk of Bias and Trustworthiness of Individual Studies 27
2.8 Data Synthesis 28

Chapter 3: Results **29**
3.1 Study Selection **29**
3.2 Study Characteristics **32**
3.2.1 Publication Details **37**
3.2.2 Geographical Context **38**
3.2.3 Educational Context **39**
3.2.4 Sample Size **40**
3.3 Study Design **40**
3.3.1 Outcome Type **41**
3.3.2 Study Duration **42**
3.3.3 Reading Measures **43**
3.3.4 Speaking measures **48**
3.3.5 Listening Measures **50**
3.3.6 Writing Measures **51**
3.3.7 Additional measures **52**
3.4 Risk of Bias **54**
3.4.1 Quality Appraisal and Trustworthiness **54**
3.4.2 Cumulative confidence across studies **58**
Chapter 4: Discussion **61**
4.1 Geographical and Educational contexts **61**
4.2 Study Design **61**
4.2.1 Language-Learning Outcome Measures **62**
4.2.1.1 Reading Measures **62**
4.2.1.2 Speaking measures **63**
4.2.1.3 Listening measures **64**
4.2.1.4 Writing measures **64**
4.2.1.5 Addtional measures **65**
4.3 Risk of Bias, Quality Appraisal, and Trustworthiness **65**
4.4 Limitations **67**

Chapter 5: Conclusion **68**

References **69**

Appendix A: IDESR Protocol **73**

Appendix B: Bibliographic Information of the included studies **81**

Appendix C: ASReview Lab Analytics **84**

Appendix D: Blank Sample Data Extraction Form **84**

Appendix E: Completed Data Extraction Form **87**

List of Figures

Fig. 1 PRISMA 2020 flow diagram for updated systematic reviews which included searches of databases, registers and other sources (Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D., 2021). 30

Fig. 2 Number of studies by publication year and educational context. 35

Fig. 3. Geographic region. 36

Fig 4. Instructional context of studies 37

Fig 5. Sample sizes of the included studies. 38

Fig. 6 Frequency of outcome types of the studies. 40

Fig. 7 Study duration of the included studies. 41

Fig 8. Weight of evidence based on quality appraisal 55

Fig. 9 Bubble “heat” chart of Quality Appraisal categories. 56

List of Tables

Table 1: Approaches to Phonics Instruction 10

Table 2. Eligibility Criteria 18

Table 3. Information Databases 20

Table 4: Search Strategy 22
Table 5: Data Collection 25
Table 6: Judging the relative trustworthiness of research results (Gorard, 2024) 26
Table 7: Study Characteristics 30
Table 8: Study Design 38
Table 9: Studies reporting Reading Measures 41
Table 10: Speaking Measures 46
Table 11: Listening Measures 48
Table 12: Writing Measures 49
Table 13: Additional Measures 50
Table 14: Quality Appraisal and Rating 52

Abstract

This systematic review critically evaluated the totality of information regarding explicit phonics instruction in a foreign language context on language learning outcomes of reading, speaking, listening, and writing. This review focused on foreign language learning contexts because learners' exposure to the target language is often limited by classroom time and instructional opportunities. The methodology adhered to PRISMA guidelines, and the studies were selected based on inclusion criteria, emphasizing the use of explicit phonics instruction and measurable learning outcomes. A total of 29 studies, conducted across Asia, the United States, Europe, the Middle East, Africa, and South America, were included. Most studies were implemented in primary school and university settings. Overall, the included studies reported positive effects of phonics instruction on reading, speaking, listening, and writing, as well as additional measures including phonological awareness and vocabulary knowledge. However, despite the majority of studies claiming positive effects, the trustworthiness and quality appraisal revealed that the overall methodological quality of the included studies was poor. Most studies received a 'weak' or 'moderate' rating suggesting that the results of the interventions were unreliable and under reported data. Implications for this review reveal that more research needs to be conducted surrounding the effects of explicit phonics instruction in a foreign language learning context and to provide educators with trustworthy, evidence-based guidance for teaching practice.

Chapter 1: Introduction and Literature Review

This chapter highlights the role of phonics instruction in first language and foreign language studies in order to preface the argument surrounding this topic. Section 1.1 provides the introduction along with the rationale behind this research. Section 1.2 discusses definitions of phonics, key approaches involved in this area, along with information regarding educational policy surrounding phonics instruction. Section 1.3 discusses theoretical background along with phonics research within language learning contexts. Section 1.4 discusses the focal language learning outcomes that could potentially be affected by explicit phonics instruction. Section 1.5 provides information on what current research into phonics instruction. The final section, 1.6, summarizes this chapter along with identifying the gaps within the literature that drive this systematic reviews research question.

1.1 Introduction and Rationale

This systematic review synthesizes evidence from intervention research that investigates the use of explicit phonics instruction in foreign language contexts to determine its impact on the linguistic outcomes of reading, speaking, listening, and writing. Phonics has been argued to be an integral aspect to the acquisition of any language. As highlighted by the UK Government, phonics is a part of the language 'pillar' that supports linguistic and language development (Ofsted, 2021), highlighting its role in education. Phonics, or the relationship between the phoneme-grapheme correspondences that make up an alphabetic language is used in first language contexts for teaching young children how to read and write (Sitthitikul, 2014). Advocates for early readers emphasize that explicit phonics instruction is necessary for reading production within first language learning (Sitthitikul, 2014). To further highlight phonics' importance, phonological awareness, or the ability to perceive and recognize sound systems within a language have been indicators of good readers within L1 studies (Connelly et al 2001; Hogan et al., 2005; Torgerson et al. 2019). Hogan et al (2005) found that children in the first grade who displayed higher phonological awareness were able to read at a second-grade level and showed higher reading and decoding abilities than to children that scored lower in phonological awareness tests. While phonics instruction is well established in first language literacy development, its application in foreign language teaching has not been greatly explored. Despite assertions in curriculum and policy, explicit phonics instruction may not be given as

much attention within a language classroom as pedagogy might suggest (Levis, 2005). The goal of this review is to examine the evidence that exists pertaining to explicit phonics instruction in a foreign language context meaning the target language is not heard other than in foreign language lessons.

1.1.2 Rationale

Following my experience with language learners whilst being an avid language learner myself, phonics has been an area that I have found that does not appear to receive much attention in the language classroom. Language classes that I enrolled in at the university level were high paced and only focused on grammar and little to the phonological aspects of words or phrases. I also struggled with acquiring the accurate phonology of the courses I was in as they focused little on authentic conversation and speech production. As a teaching assistant and substitute teacher, my experience with English as a second language (ESL) and English as a foreign language (EFL) learners demonstrated that during reading and decoding tasks, the learners struggle with fluency and pronunciation. Upon further investigation during my years teaching, I became interested with phonics instructions as it lays the foundational blocks to reading and speaking the target language. In preparing the current investigation, the research appears to be scattered with varying opinions on how phonics instruction supports foreign language learners in their acquisition journey. This systematic review seeks to identify the totality of evidence regarding phonics instruction from intervention studies where phonics instruction is investigated in foreign language contexts. The synthesis of studies will include categorizing them by linguistic outcomes, namely listening, reading, speaking, and writing. The goal of this study is to highlight the existing research that involves explicit phonics instruction and its effects on substantive linguistic outcomes in order to assess what literature already exists and what can be reliably concluded about the effects of phonics instruction on language-learning outcomes with either adult or children learners in second or foreign language contexts. This review also seeks to highlight any gaps in research in relation to phonics instruction to suggest research for the future of phonics instruction within language acquisition.

1.2 Phonics: Definitions and Approaches

1.2.1 What is phonics?

Phonics refers to the structure of sound systems that occur within a language (Mesmer and Griffith, 2005). Furthermore, phonics is a system for encoding speech sounds into written symbols (Mesmer and Griffith, 2005). Phonics instruction refers to the explicit teaching of the systematic relationships between letters and sounds, particularly in an alphabetic writing system (Woore, 2021). This method of teaching is grounded in the principle that understanding the sound-symbol correspondences can support overall reading comprehension primarily in first language research (Ehri et al., 2001). Throughout first language studies, phonics is often used to teach children how to read along with phonological decoding (Beck and Juel 1995). Based on this notion, phonics in a foreign language context supports learners' ability to interact with and learn the basic sound systems of the target language. Explicit phonics instruction draws on the learners phonological knowledge, or the organization of sounds and their rules and patterns (Gillion, 2004). When explicit phonics instruction is adapted to the target language, this can support the learners understanding of the sound-symbol correspondence and decoding strategies that are unique to that language (Birch and Fulop, 2020).

1.2.2 Phonological Awareness

Phonological awareness is the ability to recognize, identify, and manipulate any phonological unit in a word (Woore, 2021; Goouch and Lambirth, 2007; Durgunoğlu and Öney, 1999). When paired with explicit phonics instruction, this allows learners to improve phonological awareness skills that support the learner’s ability to decode words through reading and perceive sounds within their target language (Swank and Catts, 1994). Within first language studies, explicit phonics instruction is effective when providing awareness of the sound systems within the language because children who can hear and manipulate sounds before connecting them to the alphabetic letters (Castels et al., 2018). Combining these approaches can improve reading accuracy and fluency for young learners who are learning English (Castels et al., 2018; Moats and Tolman, 2009). Moreover, these skills form a foundation for more advances literacy development, including spelling writing, and vocabulary acquisition which further emphasizes the importance of integrating phonological awareness and phonics instruction throughout the language learning process (Swank and Catts, 1994).

1.2.3 Key Approaches

Phonics is supported by different approaches which includes Grapheme-phoneme correspondence (GPC), or the relationship between the written graphemes and the sounds (phonemes), that exist within the language (Macaro, 2007; Woore, 2021). GPC is a core component within phonics instruction and is linked with reading comprehension and an improvement of L2 writing skills (Macaro, 2007). There are four key approaches that are involved in the various ways of teaching phonics explicitly, as outlined in Table 1.

Table 1: Approaches to Phonics Instruction

Title	***Focus***
Synthetic Phonics	Involves breaking down words into their individual sounds (phonemes) and then 'blending' those sounds together to form words (Wyse and Goswami, 2008; Rose, 2006; Goouch and Lambirth, 2007).
Analytical Phonics	Uses "whole words" when identifying individual sounds with their corresponding letters (Stahl et al. 1998).
Systematic Phonics	Phonics is taught through a structured and sequential approach by learning the relationship between sounds and letters in a set pattern (Ehri et al, 2001; Meyer, 2001)

Embedded Phonics	Phonics is taught through contextual learning through authentic reading and writing (Scull and Lyons, 2024).

Synthetic and systematic phonics instruction are the most dominant of these four approaches with educators and researchers. In the *Independent Review of the Teaching of Early Reading,* suggests that synthetic phonics should be the "prime" approach to teaching children how to read and that instruction needs to be systematic, meaning that it follows a clear and defined program (Rose, 2006). Systematic synthetic phonics refers to an approach where the individual sounds (phonemes) and the letters (graphemes) are explicitly taught to blend those elements together (Zsargo, 2021). Within the UK, the systematic element refers to following a set program within a school setting to where the GPCs are taught for the purpose of decoding and reading for children (Zsargo, 2021).These programs have primarily been explored through first language studies in English speaking countries such as the US and UK as well as the teaching of English in ESL and EFL contexts (Urbanová, 2017; Alba, 2021). Synthetic phonics programs such as, "Jolly Phonics" have been employed through first language studies to teach young children phonics when reading (Goouch and Lambirth, 2007). Coupling these approaches can ensure that learners are taught phonics through a specific and structured nature that is comprehensible for reading development (Zsargo, 2021).

1.2.4 Phonics in Education- US and UK contexts

Phonics instruction within education has been discussed in different educational contexts. As stated by the UK government, phonics is highly valued and is considered to be a "pillar" to language learning or a core foundational aspect (Ofsted, 2021). Outlining specific policy and promoting explicit phonics instruction influences education and language learning programs. Furthermore, within the US, while there is no direct policy outlining the support of phonics instruction the US Department of Education, they suggest that states incorporate "The Science of Reading" which is a collection of evidence based on decades of research to employ what is deemed the "best practice" in terms of reading education (Petscher et al., 2020). Explicit phonics

instruction is listed in this collection to promote reading and phonological awareness for early readers within the US but only based on first language curriculum. Explicit phonics instruction has also been debated with the "Whole Language" approach which conceptualizes reading and writing as a holistic, organic extension of oral language curriculum (Reinking et al., 2023). This approach does not focus on teaching phonics explicitly, but rather focuses on meaning of words and their sounds than breaking apart the words and understanding their phonological parts. The "Whole Language" approach also sparked what scholars have called, 'reading wars' in the early 21st century to determine best practices in reading (Treiman, 2018; Woore, 2021). Those who favored phonics instruction would go against those who favored the "Whole Language" approach. However, in 2000, the National Institute of Child Health and Human Development published their Report of the National Reading Panel and it emphasized the importance of phonics instruction and phonological awareness as there are benefits for reading. They also highlight that phonics instruction should be explicit and systematic to see clear benefits. Despite intense debates that have not been fully resolved, phonics instruction continues to gain traction in L2 studies as more research emerges from empirical studies. Based on policy makers and reports suggesting that phonics be incorporated in first language (in the US) and in language learning (in the UK), it is surprising to find little investigation into foreign language. It is also important to note that much of what has been investigated revolves around learning English as an L1 in English speaking countries or ESL or EFL learning (Huo and Wang, 2017), with scant research into the effects of phonics instruction on the acquisition of other languages. Based on these instances, there needs to be emphasis on foreign language contexts to make generalizations on how phonics instruction may assist foreign language learners across the world to develop their understanding and knowledge of the target language.

1.2.5 Foreign Language Learning

The foreign language learning context is different from first and second language contexts, due to the amount of exposure the learner is receiving (Bley-Vroman, 1990). In second language contexts, learners are exposed to the target language because it is often the dominant or community language that exists outside of the classroom. Second language learners are immersed in the language, giving them more opportunities to practice outside of educational settings (Lightbown and Spada, 2013). In foreign language contexts, the learners are receiving

language input primarily from language classes which is far less exposure than other language learning context (Lightbown and Spada, 2013). Lightbown and Spada (2013) argue that the amount of exposure a learner receives is crucial during acquisition. Based on this argument, language input is especially important for developing language skills. Explicit phonics instruction, in this sense, should involve adequate exposure and intentional teaching practices (Martínez, 2011).

1.3 Theoretical Background

Having established that explicit phonics instruction has had substantial empirical evidence throughout first language acquisition and remains important throughout policy, the use of phonics in a foreign language context has been less explored. From first language studies, empirical evidence promotes that phonics instruction supports learner relationships regarding the relationship between sounds and letters within the target language. This theoretical background emphasizes that understanding these relationships can aid foreign language learners in decoding written texts, particularly in alphabetic languages.

1.3.1 Phonics in Language Learning Contexts

During the investigation of this systematic review, there has been some research conducted in second language (L2) studies surrounding the explicit phonics instruction. Due to the typological differences in languages and varying orthography, explicit phonics instruction can support decoding when learning to read (Goouch and Lambirth, 2007). Children are exposed to sound systems before learning how to read and identify them through the orthography. However, in a foreign language context, learning a new phonological system takes explicit instruction to learn, especially if the language has many different connections between letters and sounds. (Goouch and Lambirth, 2007). Similarities between languages might make this process easier as learners may have been exposed to these sounds prior to their acquisition journey. Goouch and Lambirth (2007), emphasize that there are three key challenges that language learners need to overcome in order to successfully phonologically decode in a different language such as availability of sounds exposed in a given language (sounds vs visual symbols), consistency that varies across different languages, and granularity of sound-to-symbol mapping which consists of the number of sounds in a given orthography. Ziegler and Goswami (2006) argue that resolving these challenges

coupled with direct instruction have a positive effect on reading proficiency in any language. Whilst there are challenges in learning any given language, ensuring that learners are equipped with the necessary tools through phonics instruction could support learners of other languages.

1.4 Language- Learning Outcomes

Language-learning outcomes refer to learners' measurable performance defined in this review through listening, speaking, reading, and writing. Measurement refers to quantifiable indicators of defined constructs that can be elicited through tests, interventions, or procedures (Norris, 2006). In a foreign language context, language learning outcomes are present to measure the knowledge and application of the language learned (Norris, 2006). Outcome categories in phonics are often influenced by phonological awareness and decoding skills. Through first language studies, the literature suggests that learning outcomes are most influential for reading in terms of decoding accuracy and fluency. While reading is what is most frequently studied, the impact on phonics in a foreign language context within linguistic outcomes is less explored. This review focuses on listening, speaking, reading, and writing in order to comprehensively understand the ways in which explicit phonics instruction influences these outcomes specifically in a foreign language context. To preface, the research conducted in second language and ESL studies offer an insight into how explicit phonics instruction might support foreign language research.

A meta-analysis regarding phonics instruction for L2 English word reading revealed that phonological awareness training and explicit phonics instruction improved reading outcomes (Murphy Odo, 2021). Similarly, Martínez (2011) found that explicit phonics instruction improved reading comprehension and literacy skills for EFL students in Columbia. When assessing writing, outcomes can be assessed through spelling and word construction through a phonological intervention. For example, in a study assessing the effectiveness of a synthetic phonics program, *Jolly Phonics*, in English learners in Nigeria, exposed that those who were in the experimental group receiving phonics instruction through the program showed higher scores in spelling tests compared to the control group, thus supporting the notion that explicit phonics instruction can be beneficial for writing (Umezinwa and Udogu, 2018). Phonics instruction has also been said to improve writing skills through GPC (Macaro, 2007). Furthermore, listening skills are potentially impacted by phonics instruction by increasing the learners ability to perceive and process

phonemes and discriminate sounds within the target language. Listening skills are also associated with phonemic awareness and the ability to recognize, identify and manipulate any phonological unit in a word. In a meta-analysis investigating the effects of phonemic awareness instruction for L2 listening comprehension compiled 13 articles and found that explicit instruction improved listening skills (Choe et al., 2020). When evaluating speaking outcomes, pronunciation, fluency, and phonological accuracy are characteristics that phonics instruction can influence (Arcos Guangashi, 2020). It has also been argued that speaking skills are harder to acquire than other linguistic measures (Bailey, 2003). Studies have been conducted for L2 learners regarding speaking are catered to improve English accentedness (Satio, 2011; Munro and Derwing, 1995). A study found that explicit phonetic instruction showed improvements in reading out loud and specifically looked at comprehensibility at which the learner produced the sounds. The learners were evaluated by native speakers to determine comprehensibility (Satio, 2011). While accentedness occurs when learning and speaking sounds that are different from the L1, the research conducted does not discuss any other speaking characteristics apart from accent and some fluency. Accentidness and explicit phonics instruction focus on different aspects of spoken language despite both being related to pronunciation. Accentedness generally refers to the extent to which a speaker's pronunciation diverts from what is considered the local variety of the language spoken along with the impact that this has on the larger language community (Derwing and Munro, 2009). Since explicit phonics instruction often relies on a systematic and organized approach to teaching the learners the sounds and patters of the target language (Ehri et al, 2001), research focusing solely on accentedness when determining comprehensibility and not acknowledging measurable linguistic skills should not be the most reliable measure of speaking abilities.

1.5 Current Research in Phonics Instruction

Recently, there has been an increase in the investigation of phonics instruction in language acquisition. Studies focusing on explicit phonics instruction in L2 acquisition has provided some empirical evidence to support linguistic outcomes such as reading. In a systematic review conducted in 2024 titled, "A systematic review into the effectiveness of teaching phonics in English as a Foreign Language", investigated the effects of teaching phonics in EFL contexts. The review compiled 25 studies that synthesized interventions pertaining to literacy skills. Reading comprehension was deemed the most common linguistic outcome that phonics

instruction showed a positive effect on (Albalooshi, 2024). Furthermore, the recent research conducted focuses on the learning of English and little exists pertaining to other foreign languages. Based on the literature available, the goal of this review is to explore all linguistic outcomes such as reading, listening, speaking, and writing to determine the extent to which phonics instruction affects these outcomes because a systematic review helps to minimize the bias of a narrative review by creating a transparent and replicable search strategy.

1.6 Summary and Research Questions

Although previous research investigates the use of phonics instruction can be beneficial in first language studies along with some L2 studies, there needs to be a focus on foreign language studies to determine how effective it can be through all language learning avenues. The use of explicit phonics instruction within foreign language contexts has shown to have a lack of clear empirical evidence surrounding its use in this regard. Due to the assertions by policy makers and scholars around language learning, pedagogy suggests that explicit phonics instruction should be valued. Although some policy suggests that explicit phonics instruction be used in first language contexts but also in some L2 contexts demonstrates that there is a need for a systematic review to extrapolate the totality of evidence regarding explicit phonics instruction. The goal of this review is outlined by the research question below:

RQ1: What research has been conducted that investigates the effects of explicit phonics instruction on listening, speaking, reading, and writing outcomes of learners of foreign languages?

Chapter 2: Methodology

This chapter highlights the methods used within a systematic review and the protocol in which research is conducted. Section 2.1 introduces the eligibility criteria of what this review seeks to identify. Section 2.2 discusses the informational sources that are included to locate and gather the potential records for this review. Section 2.3 describes the search strategy that is involved in locating the exiting studies that are relevant to the scope of this review. Section 2.4 discusses the data management involved in this type of review. Section 2.5 provides information on the selection and screening process of the studies. Section 2.6 discusses the data collection process and how data is extracted from the studies. Section 2.7 provides information on the risk of bias and trustworthiness of the individual studies. The final section, 2.8, discusses data synthesis for this review.

This study adopted a systematic review in order to address the research question as well as an exploration of the relevant literature pertaining to phonics. A systematic review's purpose is to locate and gather as much evidence that relates to the particular research question (Cherry, Boland, and Dickson, 2024). The process allows for explicit and transparent methods that comply with standards. This systematic review will follow the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) 2020 reporting guidelines that create a guided path for the research process (e.g., Hamilton et al., 2024; Huang & Chalmers, 2023; Schulz et al., 2023). I also followed the guidance of the book, "Doing a Systematic Review: A Student's Guide" by Cherry et al. (2024) as it provides a reference guide for academic and graduate level systematic reviews. Before conducting any research, I collaborated with Robert Woore at the University of Oxford's Education Department as he had begun a systematic review that investigated the effects of phonics instruction in a foreign language context in its entirety. To support his research, I ran his initial search strategy to compile the totality of evidence that exists within the field. To ensure the systematic reviews' transparency and reproducibility, this review's protocol was registered with IDESR in May 2025 [link to be added after examination].
The entire protocol is included in Appendix A.

2.1 Eligibility Criteria

Table 2. Eligibility Criteria

Item	**Include/ Exclude Criteria**	**Rationale**

Bibliographic Information	**Include**: Studies with a full reference or sufficient information. **Exclude**: Studies with insufficient bibliographic information.	Without sufficient bibliographic information, retrieval of works cannot occur.
1. Date of Publication	**Include**: Published on any date.	Attempting to collect the totality of available evidence that exists.
2. Participants	**Include:** Studies on typically developing foreign language learners. Include studies even if no explicit reference is made to learning ability if reasonable assumptions can be made that participants are composed mainly of typically developing individuals. **Exclude**: Studies that exclusively target non-typically developing learners or learners with Developmental Language Disorder.	This review seeks to examine existing phonics interventions in a FL, EFL, and SL. The findings for non-typically developing populations may not generalize to a larger population, thus such results will not be extrapolated or included in this review. This also focuses on the outcomes of phonics instructions in all educational settings.
3. Language Context and Publication	**Include**: Studies conducted in FL (foreign language), EFL (English	This review seeks to examine the impact of phonics instruction in FL and EFL classrooms and will

		as a Foreign Language), and SL (second language) context. **Include**: Studies published in any language **Exclude**: Studies conducted on first language acquisition.	not be investigating first language acquisition. SL is retained in the search to minimize the risk of missing papers that do not have a clearly defined language learning context.
4.	Intervention	**Include**: Studies where explicit phonics interventions are implemented in the classroom and/or studies where phonics is taught explicitly as part of a wider intervention (even if not explicitly labeled phonics). **Exclude**: Studies that do not include phonics interventions or explicit phonics teaching.	This review seeks to examine the totality of evidence pertaining to phonics interventions and teaching, even where it is not labeled as phonics.
5.	Outcomes	**Include**: Primary research studies reporting any language acquisition outcome of phonics instruction measuring one or more specific outcomes of phonics instruction, including but not limited to: speaking, reading, and phonological decoding, vocabulary knowledge, listening, writing, and spelling.	A synthesis of empirical findings in this field of literature could not occur without the reporting and evaluation of any data.

	Exclude: Systematic reviews that only report secondary data.	
6. Study Design	**Include**: Experimental and quasi-experimental designs **Exclude**: Non-experimental designs (e.g., case studies, observational studies)	Since this review seeks to ascertain what can be reliably concluded about the effects of explicit phonics instruction on language-learning outcomes, study designs where a causal implication can be draw are included and those designs which are not suited for drawing causal implications are excluded.
7. Publication Status	**Include**: Grey literature. **Exclude**: Do not exclude studies based on publication status.	This paper seeks to offset potential publication bias by including a wider range of research, including grey literature.

2.2 Information Sources

This review seeks to locate and gather the totality of information regarding the effects of phonics instruction and how it impacts speaking outcomes. To obtain the maximum of available information, eight databases were used when searching. The information sources and databases consulted for this systematic review encompass the fields of education, linguistics, psychology, as well as multidisciplinary sources that cover wider fields of the social sciences. The list was created with support from the Bodleian Library research librarians. All databases listed in Table 2 can be accessed electronically via the University of Oxford's Bodleian Library subscription.

Table 3. Information Databases

Discipline	Database
Education	Education Collection (including ERIC), British Education Index
Linguistics	Linguistics Collection (including LLBA)
Multidisciplinary	SCOPUS, Web of Science (Social Sciences Citation Index (SCCI), Google Scholar
Psychology	PsycINFO
Grey Literature	ProQuest Dissertations and Theses

2.3 Search Strategy

Systematic reviews should incorporate a specific and balanced search to target the totality of evidence pertaining to the research question. The search methods must be written with transparency as they must be replicable by the reader or future researchers (Møller and Myles, 2016). An appropriate search strategy should be sensitive and specific to capture the most relevant studies that pertain to the research question (Aromataris and Riitano, 2014). The search strategy for this systematic review was developed with support from Catherine Hamilton, Robert Woore, and Bodleian Library librarians. The search strategy was adapted from Robert Woore's previous protocol for a systematic review (registered on IDESR.org under protocol number IDESR000023). With his permission, his search terms are used for this research. The Boolean search string consists of combining related terms in each category using the 'OR' operator and connecting different terms and topics with 'AND'. The search terms were assigned to "NOFT" meaning "anywhere except full text" or limiting the search to titles, abstracts, and key words, to obtain the most relevant results to the research question. This ensures that the papers identified will describe studies pertaining to the key words within the search stream. Asterisks are used to allow for the search of different forms of keywords (Aromataris and Riitano, 2014). For

example, searching phon* within a database might elicit results including phonics, phonology, phonemes, and so on. Before conducting any searches, in order to ascertain whether there are any existing systematic reviews on this topic, a search was conducted on the ProQuest Education Database using the below search string. There were no pertinent results out of 27 potentially relevant records, suggesting that there were no previous reviews that aligned with the objective of this current study.

Boolean Search string used to locate existing secondary studies:
noft("second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL) AND noft(teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program*) AND noft(phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding) AND abstract("systematic review" OR "meta analysis" OR "research synthesis" OR "rapid evidence assessment")

The final search string used in this review is made up of four elements which are listed in Table 4.

Table 4: Search Strategy

Topic: Language Learning Context	"second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL
AND (Teaching Program or Intervention)	teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program*
AND (Linguistic Element)	phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding

AND (Study Design)	experiment* OR quasi-experiment* OR intervention OR RCT OR randomized control* trial OR randomised control* trial OR Regression Discontinuity Design OR RDD OR compar* OR evaluat*

The entire Boolean Search string is listed below

noft("second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL) AND noft(teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program*) AND noft(phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding) AND noft(experiment* OR quasi-experiment* OR intervention OR RCT OR randomized control* trial OR randomised control* trial OR Regression Discontinuity Design OR RDD OR compar* OR evaluat*)

2.3.1 Citation chaining

Citation chaining was included in this review to find related literature pertaining to the identified records. The process conducted in this review was 'backwards searching' or the looking at the key publication and searching their bibliography to find other relevant studies that the search may have missed (Cherry, Boland, and Dickson, 2024). Citation chaining brought forth one additional study that is included in this review.

2.4 Data Management

Once the searches were conducted, the titles and abstracts were uploaded to ASReview LAB (https://asreview.nl/). This software shuffles through the titles and abstracts from the initial search and the researcher decides on whether to include or exclude each abstract as it is presented. In this training process, the software begins to learn what is relevant based on the inclusion criteria, sifting through the remaining papers to present the next most relevant paper until all potentially relevant records are exhausted (the cut-off being set at 100 exclusions in a row). Deduplicating through ASReview proved to be inefficient so I resorted to using Rayyan (https://www.rayyan.ai/), which is a web-based tool that is designed to assist in the development

of systematic reviews (Johnson and Phillips, 2018). Rayyan eliminated the duplicates that arose from searching multiple databases. After deduplicating, ASReview LAB was used to conduct the screening process. Rayyan was used again for collaboration efforts. Included studies were screened, quality assessed, and the data was extracted according to the procedures of the selection process.

All records are stored on Zotero, an open-source software that manages bibliographic data. Copies of these records will be stored on Microsoft OneDrive, on the University of Oxford secure system. All work is conducted on a MacBook Air running macOS Ventura.

2.5 Selection Process

This section dives into the selection process of records for this systematic review. This includes the initial screening of the titles and abstract, quality assurance, and the full-text screening.

2.5.1 Initial Screening:

Screening of the titles and abstracts returned by the database search were reviewed through ASReview LAB and then shared with a colleague from the University of Oxford. The titles and abstracts were assessed for relevance against the inclusion and exclusion criteria listed in Table 1. Due to the nature of ASReview LAB, the screening process can be expedited as it continues to learn based on what matches the inclusion and exclusion criteria.

2.5.2 Quality Assurance

To assess quality assurance on the appropriate records, a random 10% of the titles and abstracts will be reviewed independently by a colleague. Rayyan offers a blind screening functionality which ensures that neither reviewer can see the other's decisions during the screening process. The decisions for this 10% will be compared for consistency when applying the inclusion criteria. In the case of any disagreements, the reviewers meet, and discrepancies were resolved through a discussion. In this case, consistency between reviewers is important for the review to minimize the risk of bias.

2.5.3 Full-text Screening

Following abstract screening, the full text of potentially relevant articles identified in the abstract screening process were used. The full texts that were obtained were screened again based on inclusion and exclusion criteria in order to prepare for data collection. The full texts were stored on Zotero.

2.6 Data Collection Process

The sources that are marked for inclusion will be read closely and data extracted using a Data Extraction Form that was created in Microsoft Forms. The data collected from the forms will be exported into an Excel sheet which categorizes the information. Data collection items and example information are listed in Table 5.

Table 5: Data Collection

Category	**Data Item(s)**
General Information	Reference citation, Author details, publication type, document source, data base
Participants	L1, country, age, relevant socio-demographics
Study Design	RCT, experiment, descriptive, longitudinal study, quasi-experiment
Data Type	Quantitative, interviews, test results
Intervention	Descriptions of intervention, length, comparison
Outcomes	Outcome type, outcome name, time points, measurements
Limitations	Describe author limitations
Conclusions	Describe author conclusions

2.7 Risk of Bias and Trustworthiness of Individual Studies

Trustworthiness of individual studies will be determined using an appraisal tool that is catered toward quantitative research designs. Studies that use quantitative research designs can be examined and quality assessed through scalable measures. This review uses Gorard (2024) "Sieve" entitled *Judging the relative trustworthiness of research results: How to do it and why it matters*. When judging the quality of research, the data presented in individual studies is based on its robustness or internal validity (Gorard, 2024). The sieve (shown in Table 6) presents a 0-4 scale evaluating the design, scale, data, and measurement quality.

Table 6: Judging the relative trustworthiness of research results (Gorard, 2024)

Design	Scale	Missing Data	Measurement Quality	Rating
Strong design for research question	Large number of cases (per comparison group)	Minimal missing data, no impact on findings	Standardized, independent, accurate	4
Adequate design for research	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	3
Weak design for research question	Small number of cases (per comparison group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	2
Very weak design for research	Very small number of cases (per group)	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	1
No consideration of design	A trivial scale of study	Huge amount of missing data, or not reported	Very weak measures	0

For the purpose of this review, I took Gorard (2020)'s star descriptors and combined them with the recent sieve to better visualize the quality ratings of the included studies.

4 ★★★★- all criteria is met

3- ★★★- at least three criteria is met

2-★★- at least two criteria is met

1-★- when one criteria is met

0- does not meet criteria

2.8 Data Synthesis

Following the data extraction process, the data was categorized and then synthesized by the language-learning outcomes of listening, reading, speaking, and writing. Once the Excel sheet was created from the data extraction form, it was uploaded to R Studio to produce visualizations and summary statistics pertinent to the narrative synthesis.

If it is determined that the review includes sufficient studies with comparable methods and outcomes to permit a statistical synthesis, a meta-analysis will be carried out. In the event that there are no studies suitable for a meta-analysis, then a solely narrative synthesis will be conducted. A narrative synthesis "relies primarily on the use of words and text to summarise and explain the findings" (Popay et al., 2006). It also produces "a summary of the current state of knowledge" to the review question (Popay et al., 2006). The narrative data synthesis in this review will follow Petticrew and Roberts (2006)'s three step process to narrative synthesis that include:

1. Organizing the description of the studies into logical categories
2. Analyzing the findings within each of the categories
3. Synthesizing the findings across all included studies.

This synthesis summarizes and explains the findings of the systematic review, highlights the relationships within the data, as well as assess the available evidence.

Chapter 3: Results

This chapter presents the results and data analysis of this systematic review. Section 3.1 describes the results of the search and study selection process. Section 3.2 describes study characteristics including publication details, geographical context, educational context, and sample size of the included studies. Section 3.3 describes the study design including outcome type, duration, and the description of the focal outcome measures of reading, speaking, listening, writing. This section also discusses additional measures that were discovered during the analysis of included studies such as phonological awareness measures and vocabulary knowledge. The final section, 3.4, discusses the risk of bias of included studies plus the quality appraisal and trustworthiness of each study.

3.1 Study Selection

The PRISMA flow diagram in Figure 1 (Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D., 2021), displays the results of the selection and screening process. The initial number of records obtained from database searches was 7,854. All records were uploaded onto Zotero using a .ris bibliographic file format. They were initially categorized in Zotero by database. Deduplicating through ASReview proved to be inefficient. I therefore uploaded the 7,854 records onto Rayyan to complete the deduplication process manually. After 2,242 duplicates were removed, the number of potentially relevant records was 5,613. A further. ris file containing the deduplicated list was uploaded to ASReview for title and abstract screening. Once the files were uploaded to ASReview, 2 relevant and 2 irrelevant records were identified to commence the active-learning phase of technology-assisted screening, which continues as records are screened by the reviewer and the software learns what to include or exclude. Screening ended after 134 consecutive records had been excluded, which suggested that ASReview could not find any more papers that met the inclusion criteria.

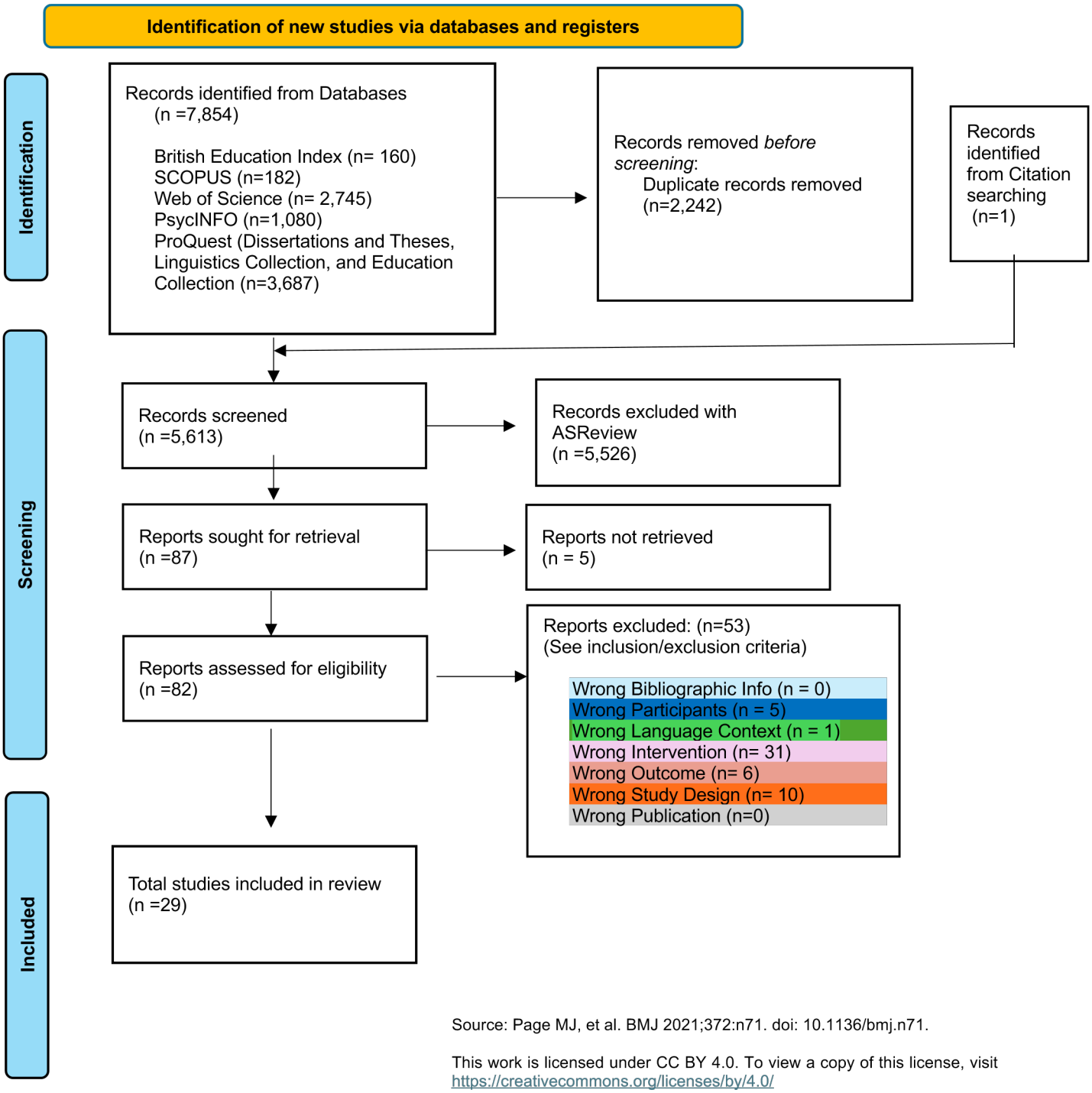


Fig. 1 PRISMA 2020 flow diagram for updated systematic reviews which included searches of databases, registers and other sources (Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D., 2021).

A colleague screened a random 10% of the titles and abstracts using the 'blind mode' in Rayyan, and there was 94% agreement between screening decisions. There was one discrepancy that was discussed and resolved. During the full-text screening phase, of the 87 records that were sought for retrieval, five records were found to be inaccessible through the University Library System and contacting the authors was unsuccessful. These full texts were therefore not obtained. During the full text screening, a further 10% sample was screened by a colleague, resulting in 87.5% agreement on inclusion decisions. The discrepancy was also resolved through discussion. Citation chaining uncovered one study that was included in the final total. The total number of studies included in this systematic review is 29.

3.2 Study Characteristics

Table 7 provides characteristics for the 29 included studies, detailing their publication details, study design, country of location, participant sample size, educational setting, and duration.

RCT= Randomized Control Trial

Table 7: Study Characteristics

Study ID	Reference	Publication	Study Design	Location	Setting	Length of Intervention (in weeks unless stated otherwise)	Sample Size
1	Patel et al. (2022)	Journal Article	Stratified matched-pair RCT- students were randomized individually within each classroom	India	Primary School	5 weeks, 5 days per week, 20 minutes per session	n=136
2	Romero et al. (2021)	Journal Article	RCT	Spain	Primary School	7 months, 3 times per week, 45 minutes per lesson	n=52

3	Abiy Zewdu et al. (2023)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Ethiopia	Primary School	12 weeks, 40 minutes per session	n=70
4	Giambo, and McKinney (2004)	Journal Article	RCT	United States	Kindergarten	19 weeks long, 60 lessons that lasted approx. 20-25 minutes per lesson, 4 times per week.	n= 80
5	Yeung et al. (2013)	Journal Article	RCT	Hong Kong	Kindergarten	12 weeks, 24 sessions, 2 lessons per week, 30 minutes per session	n=76
6	Konerding et al. (2020)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Germany	Primary School	9 weeks	n=26
7	Sun et al. (2015)	Journal Article	Cluster RCT (cluster or groups assigned randomly)	China	Primary School	10 weeks	n=80
8	FonsecaMora et al. (2015)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Spain	Primary School	11 weeks	n=63

9	Chu and Chen (2014)	Journal Article	Cluster RCT (cluster or groups assigned randomly)	Taiwan	Primary School	7 weeks (5 weeks for intervention, 2 week delayed posttest)	n=117
10	Li and Woore (2021)	Journal Article	Cluster RCT (cluster or groups assigned randomly)	China	University	12 weeks	n=138
11	Rahman et al. (2023)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Malaysia	Preschool	4 weeks, 3 times per week, 30minute sessions	n=60
12	Martin (2024)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	United States	University	13 weeks.	n=158
13	Kashiwagi et al. (2005)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Japan	University	1 year	n=126
14	Dixon et al. (2011)	Journal Article	Cluster RCT (cluster or groups assigned randomly)	India	Primary School	6 months	n=506

15	Ghorbani et al. (2016)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Iran	University	8 weeks	n=38
16	Algethami (2016)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Saudi Arabia	University	11 weeks	n=9
17	Sturm (2013)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	United States	University	15 weeks, meeting twice per week, 75 min per session	n=22
18	Hubert and Vigil (2017)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	United States	University	16 weeks, 75 minutes, 2 times per week	n=28
19	Tweedie et al. (2015)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Qatar	University	10 weeks	n=67
20	Johnson and Tweedie (2010)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	Malaysia	Primary School	6 weeks	n=862
				Egypt	Adult Teachers	3 months	n=40

21	Ghoneim and Elghotmy (2015)	Journal Article	Single group pre/ post-test (no control)				
22	Ashmore et al. (2002)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	China	Primary School	10 weeks, 5 10-minute sessions per week	n=202
23	Valbuena (2014)	Journal Article	Single group pre/ post-test (no control)	Colombia	Primary School	6 months, one our classes 3 times per week	n=25
24	Taylor (2008)	PhD Thesis	Non-equivalent groups pre/posttest (ex. control group and experimental group)	United Arab Emirates	University	16 weeks, 2 times per week, 15 minutes per session	n=74
25	Coates et al. (2017)	Journal Article	RCT	Italy	High School	10 weeks, 60 minutes per session	n=3
26	Cui et al. (2021)	Journal Article	RCT	China	University	12 weeks (2 hours/week); control group had regular classes only (no phonological training)	n=160
27	Burnham (2013)	Journal Article	Non-equivalent groups pre/posttest (ex. control group and experimental group)	United States	University	4 weeks, 100 modules, at their own pace	n=24

28	Geer and Keane (2018)	Journal Article	Between-subjects experimental design with matched groups	United States	University	3 weeks, 30-40 minutes per session	n=18
29	Nishanimut, et al. (2013)	Journal Article	Cluster RCT (cluster or groups assigned randomly)	India	Primary School	5 weeks, 1 hour each day	n=60

3.2.1 Publication Details

Figure 2 illustrates the publication trends from this area of research from the oldest paper (2002) to the most recent (2024).

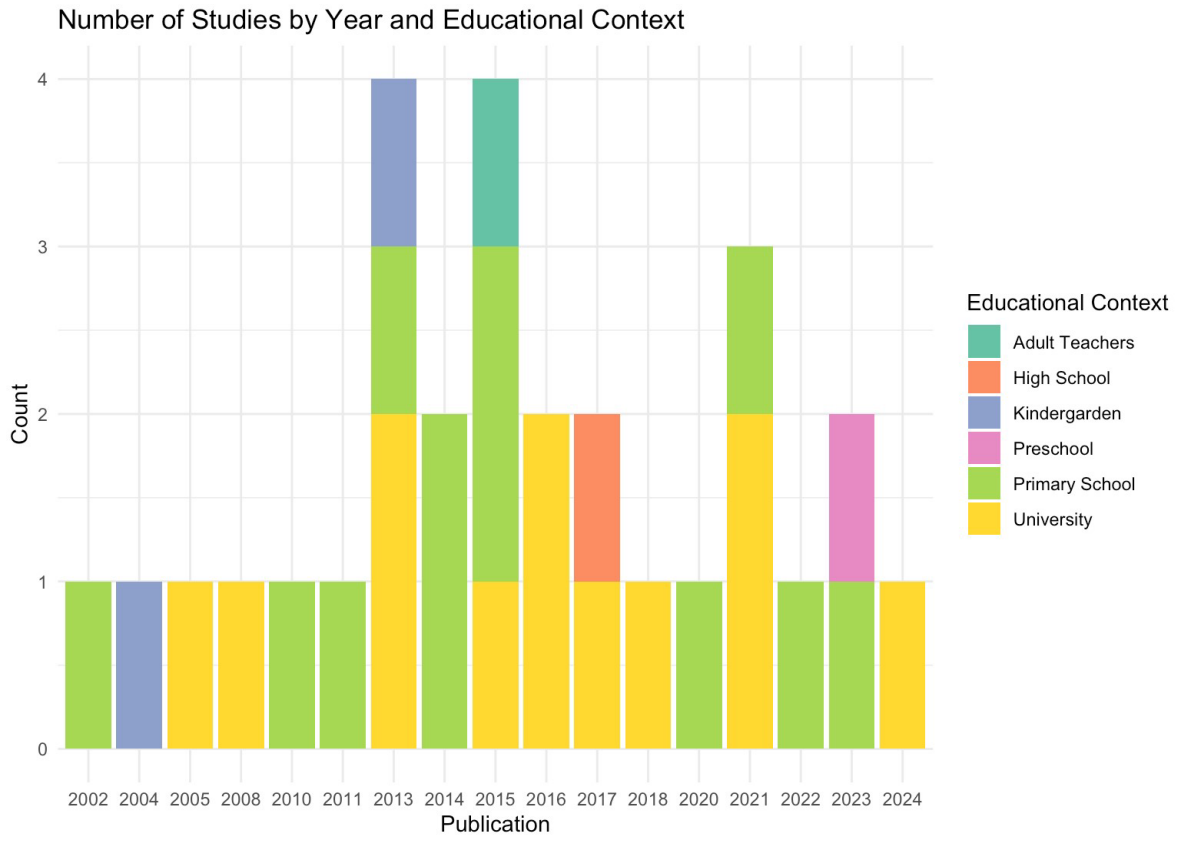


Fig. 2 Number of studies by publication year and educational context.

3.2.2 Geographical Context

The studies were conducted across 17 countries worldwide. All records were published in English with the exception of three studies that were published in English along with other languages such as no.19 (French), no.23 (Spanish), and no.25 that was published in French, Spanish, and Italian.

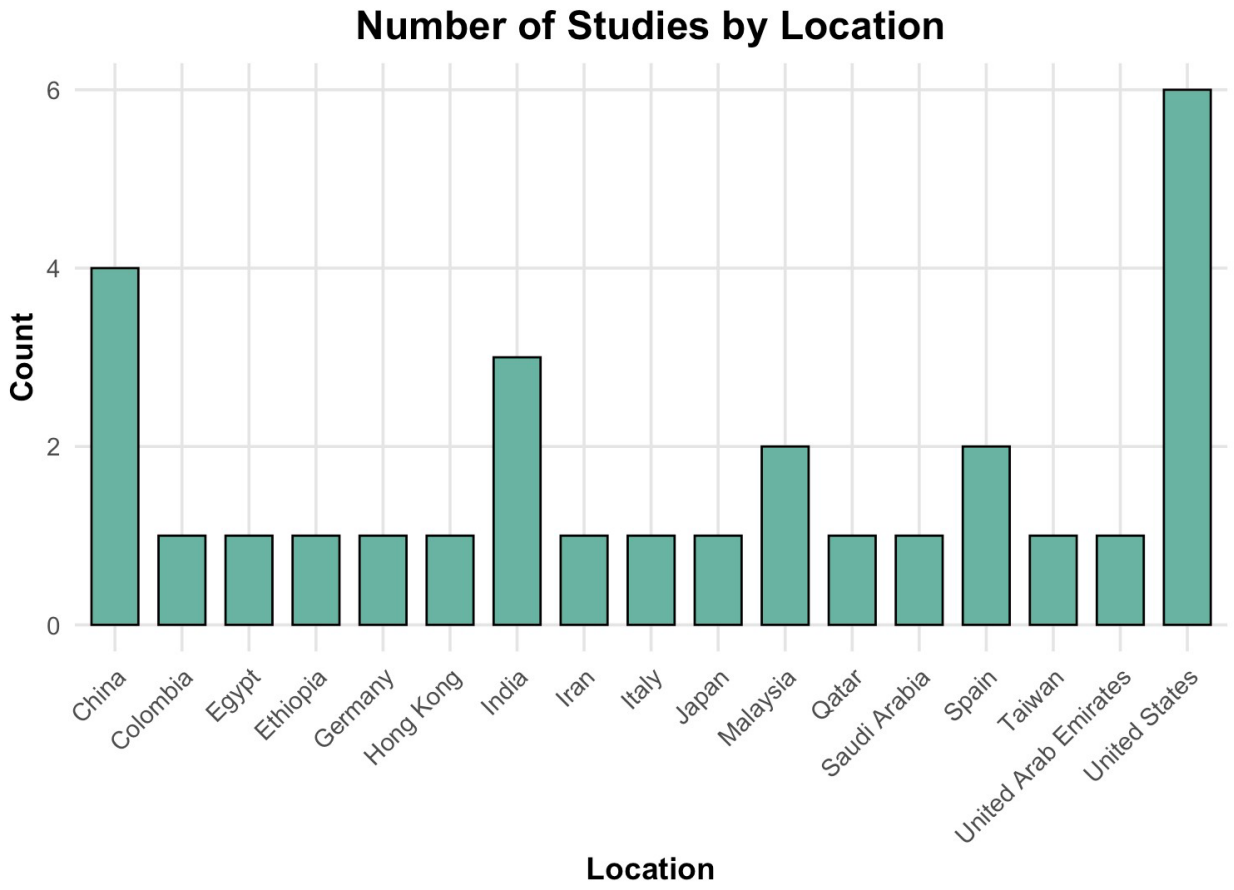


Fig 3. Geographic region.

Figure 3 illustrates the geographic region of the included studies. The country with the most studies was the US with two studies (4 and 18) learning Spanish, two studies (12 and 17) learning French, one study (27) learning Arabic, and one study (28) learning American Sign Language (ASL). ASL is considered a foreign language and is accepted by schools and universities to count towards foreign language curriculum (Peyton, 1998). China was the next

largest study group with four studies followed by India with three, Malaysia with two, Spain with two, and all other counties with one study.

3.2.3 Educational Context

Figure 4 illustrates the instructional contexts of the included studies. The majority of studies were conducted in Primary School (41%), followed by University (41%). The remaining educational settings include Preschool (3%), Kindergarten (7%), High School (3%), and Adult Teachers (3%).

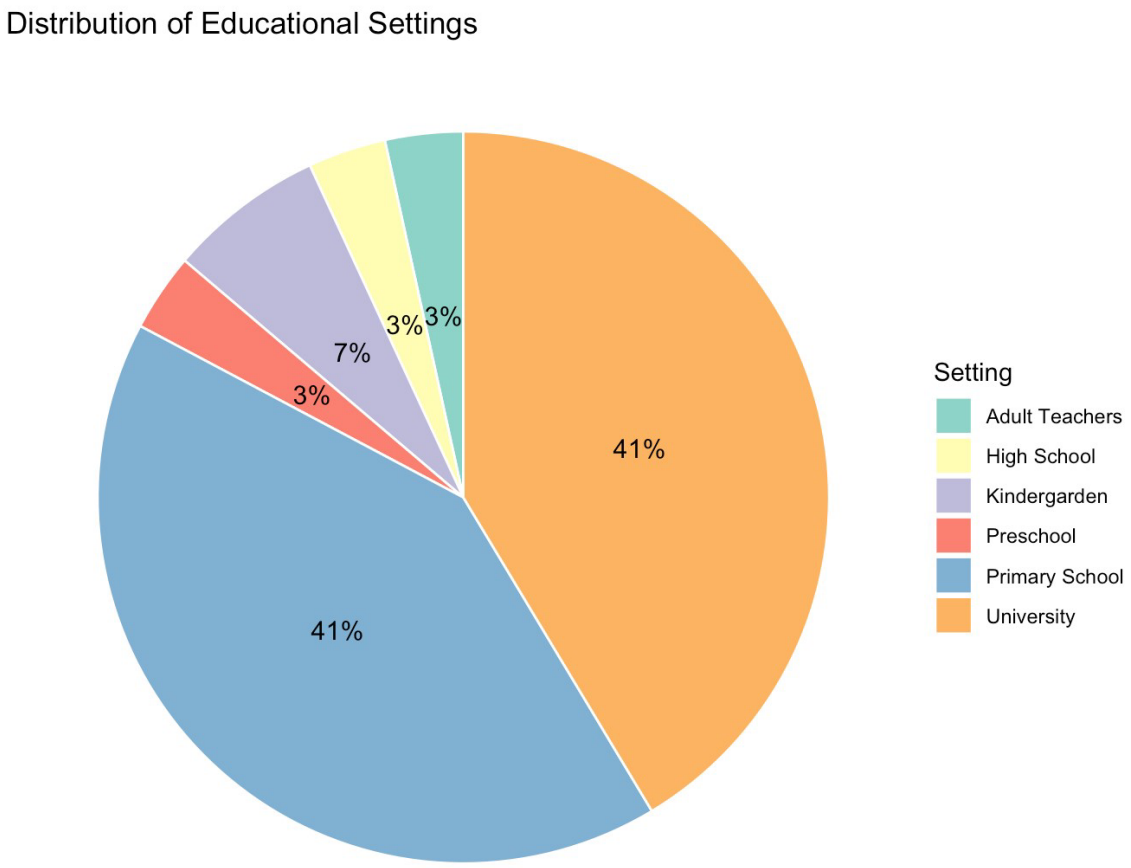


Fig 4. Instructional context of studies

3.2.4 Sample Size

Figure 5 illustrates the sample sizes of the included studies. The number of participants vary from 3 to 506. Two studies had under 10 participants, eight studies had 10-50 participants, ten studies had 50-100 participants, and lastly, nine studies had over 100 participants. Sample size of interventions are essential in making generalizations about the effects of the study (Ahrens and Zaščerinska, 2014).

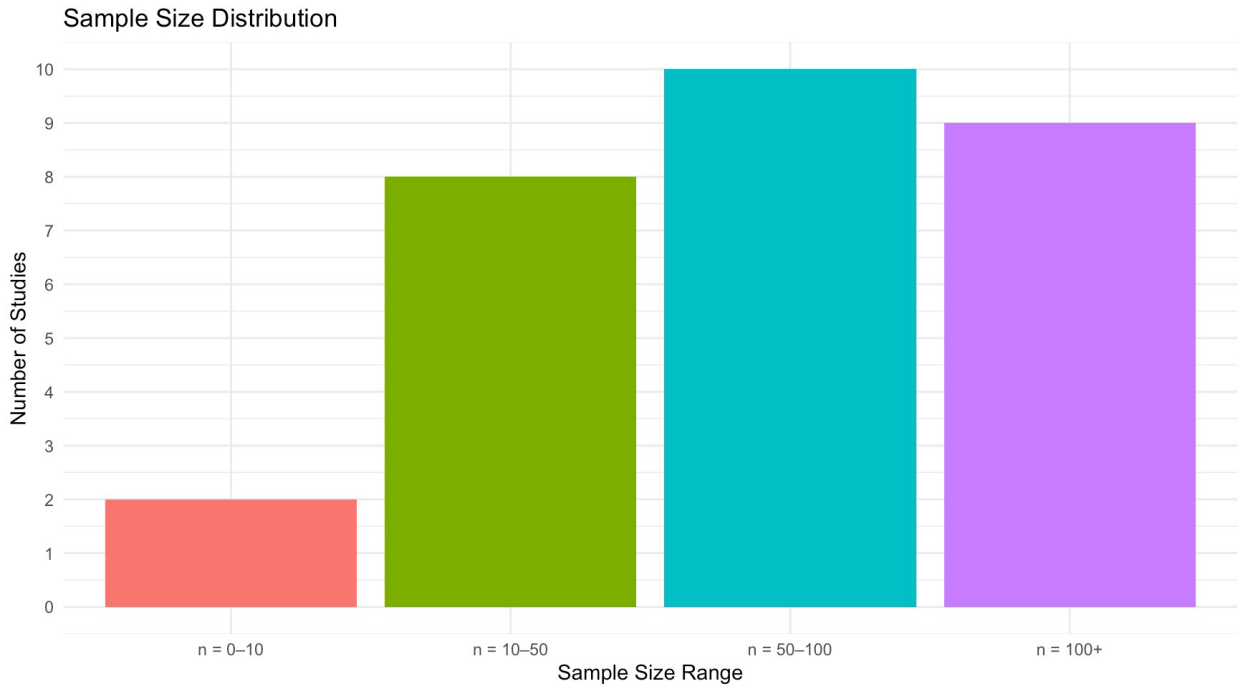


Fig 5. Sample sizes of the included studies.

3.3 Study Design

Table 8 summarizes the various study designs used within the included studies. 51% of studies (n=15) employed a non-equivalent groups pre/post-test design. All studies exclusively used quantitative methods. Study design is essential when determining the reliability of the intervention. Studies with pre/post-test design offer an insight into how effective the intervention was on the experimental group while also including a control to further measure the effects.

Table 8: Study Design

Study Design	**No. of Works**	**Study ID**

Non-equivalent groups pre/post-test (ex. control group and experimental group)	15	3, 6, 8, 11, 12, 13, 15, 16, 17, 18, 19, 20, 22, 24, 27
Cluster RCT (cluster or groups assigned randomly)	5	7, 9, 10, 14, 29
RCT (randomized control trial)	5	2, 4, 5, 25, 26
Single group pre/ post-test (no control)	2	21, 23
Between-subjects experimental design with matched groups	1	28
Stratified matched-pair RCT- students were randomized individually within each classroom	1	1

The study design that included the stratified matched-pair RCT is different from a normal RCT due to the differences in randomization and allocation strategy. Participants were randomized within each classroom after the matching occurred. This ensures that groups are balanced and thus reduces bias associated with any pre-existing differences between the classrooms.

3.3.1 Outcome Type

Figure 6 visualizes the outcomes measured in included studies. There were 14 studies that included reading outcomes; 6 studies included speaking outcomes; one included writing outcomes; and one included listening outcomes. Studies that included measures of literacy primarily focused on reading so they are included in Table 9. Additional outcomes (n= 7) included phonological awareness and vocabulary knowledge and these are summarized in Table 13.

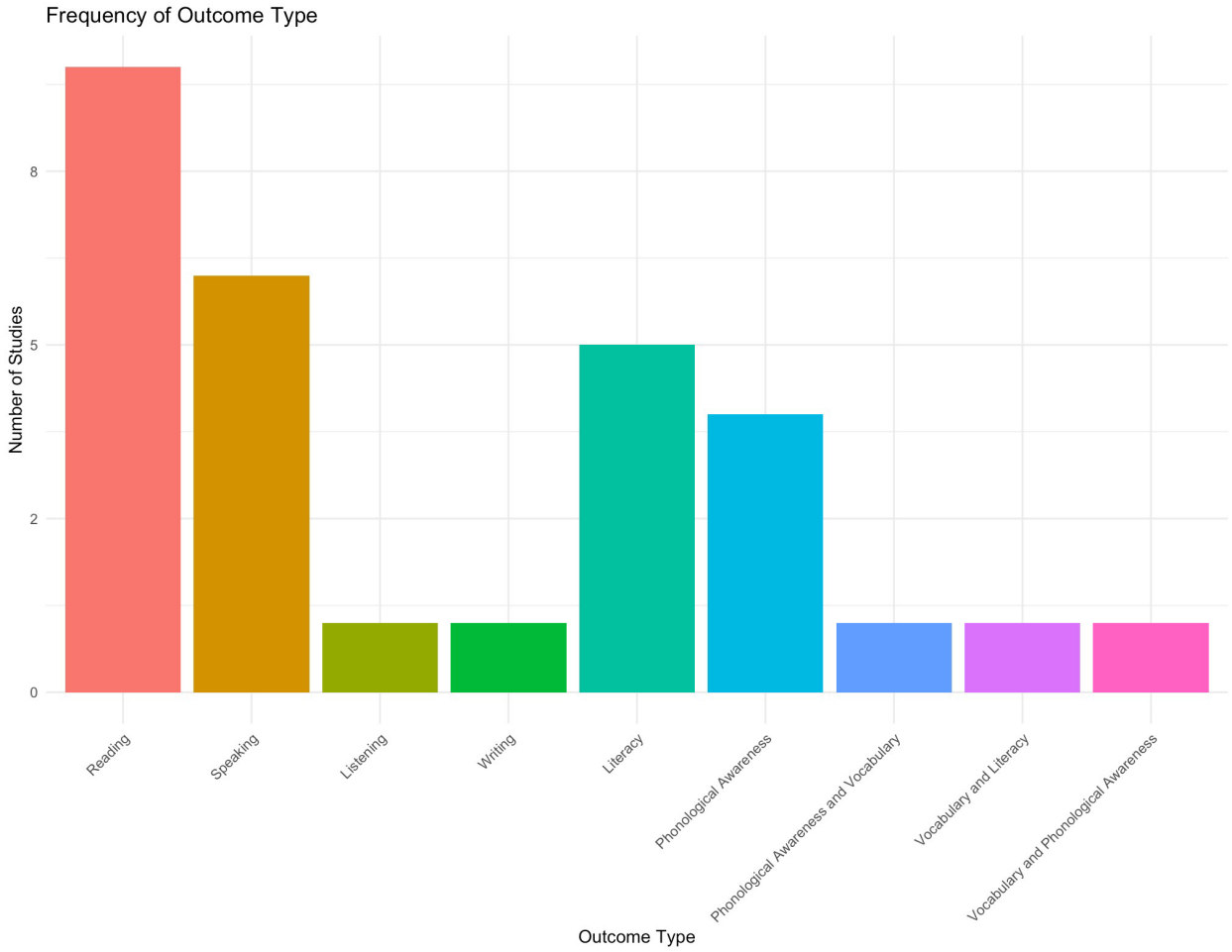


Fig. 6 Frequency of outcome types of the studies.

3.3.2 Study Duration

Figure 7 illustrates the study duration of each study. Study duration ranged from 3 weeks to one year. Studies 1, 2, 3, 4, 5, 9, 11, 17, 18, 22, 23, 24, 25, 26, 27, 28, and 29 reported how many lessons were taught and the duration of the intervention. 96% of the studies (n=28) were over four weeks (one month). Duration of the studies is an important aspect of foreign language research as the amount of exposure is important to acquisition. With the shortest study being three weeks provides some insight into how much the intervention actually affected the language outcome.

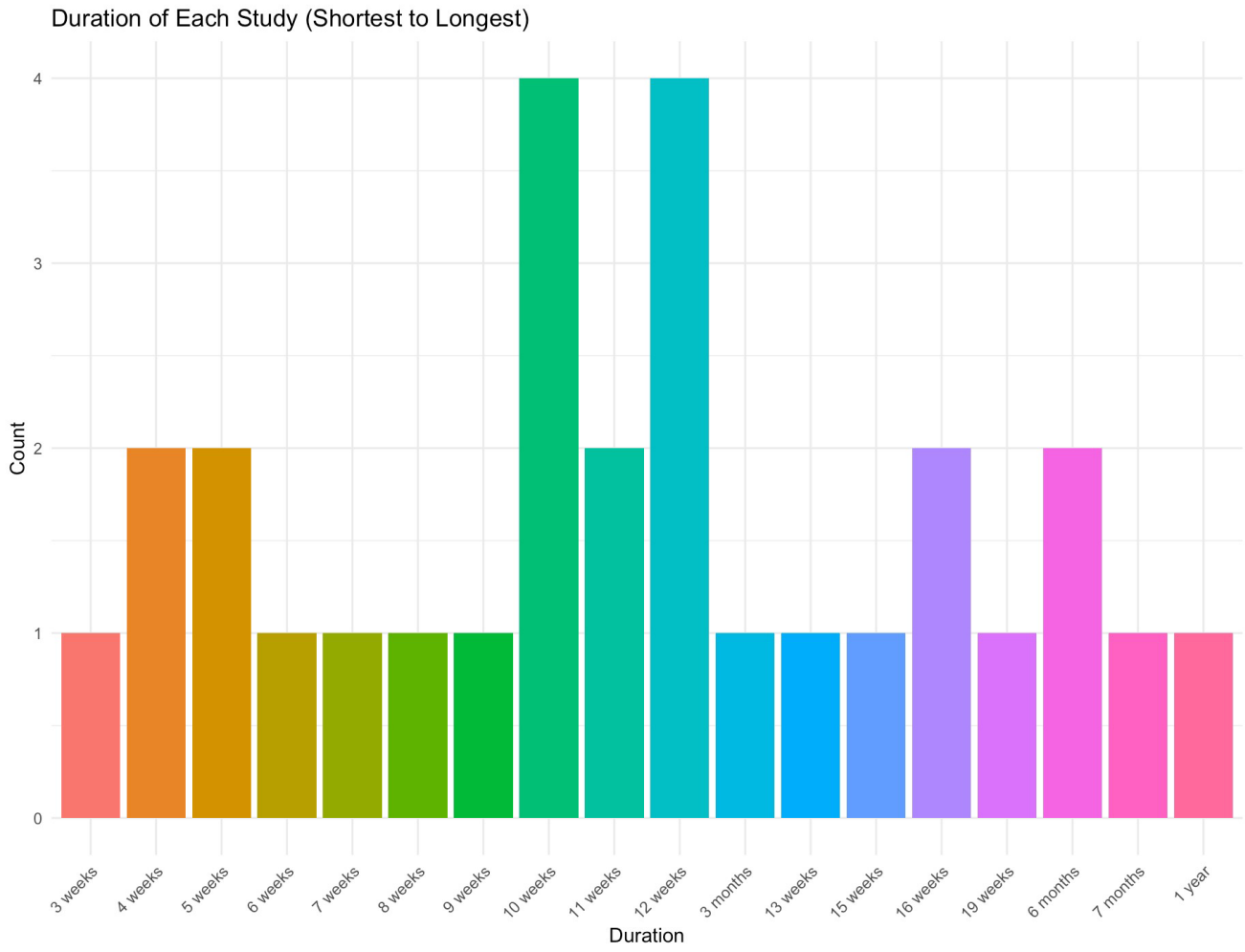


Fig. 7 Study duration of the included studies.

3.3.3 Reading Measures

The largest group of included studies, 48% (n=14), reported reading measures. Table 9 summarizes the reading measures used, the claims made by the author(s) about the effect of phonics on the outcome measure, and the study quality appraisal rating. Studies that reported literacy outcomes have been categorized as 'reading' and are therefore included in this section.

Table 9: Studies reporting Reading Measures

Study ID	Reading Measure	Claim made by author about findings Green= Positive, Orange= Mixed, Purple= Negative	Quality Appraisal and Trustworthiness indicator

			Green= Strong, Orange= Moderate, Purple= weak
1. Patel et al. (2022)	Phonics Program "GraphoLearn", systematic phonics program	Claimed positive effects of the phonics intervention for reading scores and literacy. "At the end of a 5-week intervention period, the GL group made significant improvements, particularly on in-game assessments of lettersound knowledge, rime unit recognition, and word recognition."	Rating: ★★★ Adequate and clear study design. Data collected from the online phonics program allows for use outside of the research context. Relatively small sample size cannot generalize for the phonics program.
2. Romero et al. (2021)	Synthetic Phonics program for (Phonological Awareness (PA), naming (N), speed naming (SN), letter reading (LR), word reading (WR), pseudoword reading (PWR)	Claimed positive effects for reading and literacy measures. "The results reveal that the use of phonics can be a beneficial alternative for teaching English literacy in the first year of Primary Education. The method helped the participating children to acquire a good level of English phonics in order to be able to decode and read any word or pseudo-word in English"	Rating: ★★★ Small sample size, adequate study design, clear and concise measures.
7. Sun et al. (2015)	Phonological Awareness Training (PAT) that taught the sequence from syllable, final sound, rhyme, onset, phoneme identification and matching, phoneme deletion and addition, to phoneme blending and segmentation.	Claimed no significant effects after PAT. "After a ten-week PA training program, there was still no significant difference between the treatment and contrast classes in terms of their recognition of English letters and English picture naming"	Rating: ★★★ Adequate study design, confusing discussion that depicts no effects but then supports PA training.

8. Fonseca-Mora et al. (2015)	Phonics program plus music for Correct letters read in English and Sound identification were better in both experimental groups compared to the control. This points towards the notion that the effects of a phonological based teaching approach were beneficial to both experimental groups vs the control.	Claimed positive effects for phonics interventions. Claimed that even without a musical element, phonics instruction should be implemented for FL learners "acquiring good phonological and decoding skills is of uttermost importance for foreign language learners and these abilities are not necessarily directly transferred from L1 knowledge"	Rating: ★★ Small sample size, unclear data and musical measures/ interventions.
9. Chu and Chen (2014)	Synthetic Phonics group: Phonics-only group received phonics teaching and a Phonics+ group received phonics plus decodable text instruction.	Claimed positive outcomes: "English word reading statistically improved from the pre-post and delayed posttests in both experimental groups"	Rating: ★★ Low sample size, RCT, Clear concise findings No research questions.
11. Rahman et al. (2023)	Systematic Phonics intervention: Integrated Sound Word Method (ISWM) for reading proficiency through games	Claimed positive effects from ISWM: "The games improved early reading comprehension by sharpening the children's decoding skills with pictures to comprehend. By the end of the intervention program, the findings show that preschoolers had mastered Level 4 of early reading comprehension and improved their English language proficiency"	Rating: ★★ Small sample size, low language exposure (6 hours of instruction total).
14. Dixon et al. (2011)	Synthetic Phonics, "Jolly Phonics" assessed	Claimed positive effects from the Jolly Phonics program for	Rating: ★★★ Large sample size,

	by the Burt Reading test and Schonell Spelling test	children in a low socioeconomic region of India: "The findings show that the improvements in the test scores of students experiencing the phonics method were statistically significantly higher than those in the control group when assessing reading and spelling and the ability to sound out letters and words for all the tests given"	Extensive analysis, clear and concise results with implications for children in a low socioeconomic region. No research questions.
19. Tweedie et al. (2015)	Treatment group received explicit phonics instruction through Reading Horizons that provided phonemic awareness and reading	Claimed positive effects from explicit phonics instruction. "Results suggest direct phonemic instruction lead not only gains in vowel recognition skills, but also to moderate gains for ELLs when navigating and comprehending academic English text"	Rating: ★★ Small sample size, missing data from tests No research questions.
20. Johnson and Tweedie (2010)	Explicit PA training coupled with Total Physical Repose (TPR) and were tested through 5 different categories such as: 1. Sounds fluency: 2. Basic reading: 4. Reading of Nonsense words 5. Simple writing test	Claimed positive effects from explicit phonological awareness instruction. "The results of the study showed that PA instruction made larger gains in tested skills than the control group in literacy measures"	Rating: ★ No research questions, missing data, confusing study design, lack of allocation strategy despite large sample size. Relatively unclear measures although most centered on reading measures with one writing test.

21. Ghoneim and Elghotmy (2015)	Explicit Phonics through a Synthetic Phonics approach for Adult Teachers using a Multi-	Claimed positive effects from MSPP. "The result of the study demonstrates that using a multisensory phonics program had a significant impact on	Rating: ★ No control, weak study design, small sample size.
	Sensory Phonics Program (MSPP)	reading accuracy, decoding, and PA"	
22. Ashmore et al. (2002)	Explicit PAT through: -Word reading: Sight Word Efficiency (SWE) -Test of Word Reading Efficiency (TOWRE) Phonological awareness: -Comprehensive Test of Phonological Processing (CTOP): Elision, Blending Words, and Segmenting Words	Claimed mainly positive but also mixed effects for PAT: Specifically with word reading. "Both groups started with five correct words. After 10 weeks, the control group gained 3 words (total 8), while the experimental group gained 5 (total 10). The difference was statistically significant ($p = 0.04$)" Authors claim that positive effects were statistically small but there were small gains with the experimental group	Rating: ★ Lack of comprehensive data results. Authors' discussion was confusing and could not clearly indicate positive findings but still suggested that PA training is effective. Large sample size.
23. Valbuena (2014).	Tucker Signing Synthetic Phonics program on developing phonemic awareness for reading	Claimed positive effects of the Tucker Signing Synthetic Phonics Program: "The study also shows that the Tucker Signing program is a useful tool not only for native speakers of English who are learning to read, but also for Spanish speakers who are learning to read in English as a second language"	Rating: 0 No research questions, small sample size, no statistical data included in the study. There are a few graphs with no description on what data was used and where it came from.

26. Cui et al. (2021)	Explicit Suprasegmental phonological training on English reading for EFL students	Claimed positive effects. "Suprasegmental training improved reading accuracy and speed; especially effective for lower-proficiency students"	Rating: ★★★ Detailed and clear data, Attention to study design, adequate sample size.
29. Nishanimut et al. (2013)	Synthetic Phonics Program: Samveda	Claimed positive effects. "Kannada speaking children learning to read English were	Rating: ★★ Small sample size, no research questions.
	Synthetic Phonics Program (SSPP)	able to benefit from a phonics approach, even though many of the words on which they were tested were unlikely to be in their oral vocabularies"	Relatively short duration (5 weeks).

3.3.4 Speaking measures

Table 10 summarizes the findings of the studies that reported speaking measures. 20% (n=6) of included studies reported speaking measures.

Table 10: Speaking Measures

Study ID	Speaking Measures	Claim made by author about findings Green= Positive, Orange= No Change, Purple= Negative	Quality Appraisal and Trustworthiness indicator Green= Strong, Orange= Moderate, Purple= weak

13. Kashiwagi et al. (2005)	Explicit pronunciation training that included: explicit explanation of the articulatory positions of both English consonants and vowels, practice both in drill-type activities and in more communicative interactions, discrimination exercises, more global listening activities, peer feedback and self-monitoring exercises.	Claimed positive effects on pronunciation training for speaking and long-term acquisition. "This provides some evidence that one year of explicit pronunciation instruction resulted in the experimental group's being better than the control group in not only their metalinguistic knowledge and perception, but also their production both segmental and global".	Rating: ★★★ No report of allocation strategy, moderate sample size, clear data analysis, long-term effects assessed (1 year long).
15. Ghorbani et al. (2016)	Systematic Phonics approach for Pronunciation Training: Vowel perception	Claimed mixed perceptions of phonics for pronunciation training: "The findings indicate that explicit vowel instruction, through raising learners' awareness or consciousness, is more effective than implicit teaching of vowels through the intuitive-imitative approach (repetition and imitation of sounds) in which learners are made to listen to vowel pronunciation drills and imitate them".	Rating: ★★ Authors provided mixed effects on explicit instruction but still expressed the value of instruction for learners. Small sample size.

16. Algethami (2016)	Explicit pronunciation training for long-term effects on the production of English sounds	Claimed mixed effects, one participant improved in their speech production of the target sounds.	Rating: ★ A trivial sample size (n=9), short duration to assess "long-term" effects if the intervention was only 15 weeks.
17. Sturm (2013)	Explicit Phonics and pronunciation course was taught to improve speech	Claimed positive effects. "A semester-long course of explicit phonetics and pronunciation instruction, such as the course taken by the learners in this study, is beneficial to reduction of pronunciation errors and intelligibility".	Rating: ★★ Small sample size, students volunteered to be in the experimental/ control group. Students were judged by native French speakers.
18. Hubert and Vigil (2017)	Explicit phonics with writing to improve Spanish pronunciation and speaking	Claimed mixed effects. While the target sounds were acquired, the results were not statistically significant thus showing that Phonics+ writing	Rating: ★ Small sample size, unclear discussion of results, missing data, weak study design and details.
		may not be as effective as a stand-alone phonics program.	
25. Coates et al. (2017)	Systematic Phonics approach for pronunciation and orthography	Claimed positive effects of the phonics program and that participants improved in both pronunciation and orthography in long vowel and digraph categories.	Rating: ★★ No research questions Very small sample size.

3.3.5 Listening Measures

Table 11 summarizes the findings of one study that reported listening outcomes. Study no. 27 claimed positive findings for explicit phonics instruction for the perception of foreign language sounds.

Table 11: Listening Measures

Study ID	Listening Measure	Claim made by author about findings Green= Positive, Orange= Mixed, Purple= Negative	Quality Appraisal and Trustworthiness indicator Green= Strong, Orange= Moderate, Purple= weak
27. Burnham (2013)	High Variability Phonetic Training (HVPT) for sound perception of Arabic	Claimed positive effects of the HVPT program in improving listening and sound perception and discrimination of Arabic sounds.	Rating: ★★★ Adequate study design, no research questions, trivial sample size.

3.3.6 Writing Measures

Table 12 summarizes the findings of one study that reported writing outcomes. Study no. 12 claimed positive effects for phonics instruction for writing with learners with varying L1 backgrounds but the primary positive effects came from learners who had a Roman-alphabet background. This study included learners with around 12 different L1's which provides an adequate sample to how phonics instruction could affect a variety of learners.

Table 12: Writing Measures

Study ID	Writing Measure	Claim made by author about findings Green= Positive, Orange= Mixed, Purple= Negative	Quality Appraisal and Trustworthiness indicator Green= Strong, Orange= Moderate, Purple= weak

12. Martin, K. I. (2024)	Systematic Phonics program for English learners with varying L1 background including: French, German, Italian, Portuguese, Spanish, Turkish, Vietnamese, Korean, Arabic, Thai, Japanese, and Chinese	Claimed positive effects of phonics instruction specifically for learners with a Roman alphabet L1.	Rating: ★★ No research questions, Clear measures and allocation strategy, adequate sample size.

3.3.7 Additional measures

Table 13 summarizes the findings of the studies that reported measures other than reading, speaking, listening and writing. These include phonological awareness and vocabulary knowledge. Studies no. 3, 4, 5, 6, 10, 24, and 28 discussed phonics instruction for phonological awareness and vocabulary knowledge.

Table 13: Additional Measures

Study ID	Measure	Claim made by author about findings Green= Positive, Orange= Mixed, Purple= Negative	Quality Appraisal and Trustworthiness indicator Green= Strong, Orange= Moderate, Purple= weak

3. Abiy Zewdu Agegnehu et al. (2023)	**Phonological Awareness**- Systematic Phonics intervention for PA along with multisensory activities to engage learners.	Claimed positive effects on phonics intervention for PA. "The results indicated that a twelve-week explicit rime-based phonics training-based intervention, can accelerate the development of phonological awareness".	Rating: ★★★ Moderate sample size, clear data and analysis of the intervention.
5. Yeung et al. (2013)	**Phonological Awareness-** Systematic Phonics program to increase phonological awareness of English	Claimed positive effects on PA test scores than the control group.	Rating: ★★★ Adequate research design, small sample size. Provided clear results to the tests of the intervention.
24. Taylor (2008)	**Phonological Awareness-** Systematic Phonics program Get Reading (2006) to improve phonological awareness of English	Claimed no significant statistical effects on phonics program to improve phonological awareness.	Rating: ★★★ Small sample size, clear measurements and analysis of findings.
28. Geer and Keane (2018).	Phonics instruction to improve fingerspelling of ASL learners	Claimed positive effects of explicit phonics instruction. "Explicit training improved fingerspelling	Rating: ★ Missing data, no control group, trivial sample size, moderate study design.
		comprehension accuracy significantly".	

6. Konerding et al. (2020)	**Phonological awareness and Vocabulary Knowledge** Systematic Phonics program, “Lautarium” that provides explicit phonics instruction on phoneme perception, phonological awareness, and grapheme-phoneme correspondences	Claimed positive effects on phonics program for PA and vocabulary measures. "The most important finding from the current study is that computerized grapho-phonological training evokes enduring beneficial effects on vocabulary acquisition in L2 learners.”	Rating: ★★★ No research questions, small sample size, clear data and design discussed.
10. Li and Woore (2021)	**Vocabulary and Literacy-** Systematic Phonics program, Read Write Inc: Fresh Start, to promote phonological decoding and vocabulary knowledge	Claimed positive effects of phonics intervention. The experimental group improved more than the comparison group in English decoding and vocabulary knowledge.	Rating: ★★★★ Large sample size, strong research design, clear findings and measurements.
4. Giambo and McKinney (2004)	**Phonological awareness and Vocabulary Knowledge**- Systematic Phonics intervention through introducing word sets to the participants	Claimed positive effects. “The outcome was that explicit phonics instruction improved the test scores involving phonological awareness and receptive vocabulary”.	Rating: ★★ No research questions, large sample size, confusing measures- aim describes oral English proficiency but the discussion focuses on PA and vocabulary.

3.4 Risk of Bias

3.4.1 Quality Appraisal and Trustworthiness

Table 14 summarizes the quality appraisal and trustworthiness of the included studies. Using Gorard’s (2024) sieve, each study underwent a quality appraisal and measurement of

trustworthiness. Gorard's (2024) sieve provides an appraisal tool for quantitative research methods. Since this study only explored quantitative research design, this tool provided a comprehensive method of judging the reliability and trustworthiness of the included studies. Discussion on why this appraisal tool was chosen can be found in Chapter 2.

Table 14: Quality Appraisal and Rating

Study ID	Study Design	Scale	Missing Data	Measurement Quality	Rating
1	Adequate design for research	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★
2	Adequate design for research	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★
3	Adequate design for research	Very small number of cases (per group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★
4	Weak design for research question	Small number of cases (per comparison group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★
5	Adequate design for research	Very small number of cases (per group)	Minimal missing data, no impact on findings	Standardized, independent, some errors	★★★
6	Very weak design for research	A trivial scale of study	Moderate missing data, likely impact on findings	Standardized, independent, some errors	★★★
7	Strong design for research question	Small number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★

8	Very weak design for research	Very small number of cases (per group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★
9	Weak design for research question	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★
10	Strong design for research question	Adequate number of cases (per comparison group)	Minimal missing data, no impact on findings	Standardized, independent, accurate	★★★★
11	Adequate design for research	Very small number of cases (per group)	High level of missing data, clear impact on findings	Not standardized/ independent errors	★★
12	Weak design for research question	Adequate number of cases (per comparison group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★
13	Adequate design for research	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★
14	Very weak design for research	Large number of cases (per comparison group)	Moderate missing data, likely impact on findings	Standardized, independent, some errors	★★★
15	Adequate design for research	Very small number of cases (per group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★
16	Adequate design for research	A trivial scale of study	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	★

17	Weak design for research question	A trivial scale of study	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★
18	Adequate design for research	A trivial scale of study	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	★
19	Weak design for research question	Small number of cases (per comparison group)	High level of missing data, clear impact on findings	Not standardized/ independent errors	★★
20	Very weak design for research	Large number of cases (per comparison group)	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	★
21	Very weak design for research	Small number of cases (per comparison group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★
22	Adequate design for research	Large number of cases (per comparison group)	Some missing data, possible impact on findings	Not standardized/ independent errors	★
23	No consideration of design	A trivial scale of study	Huge amount of missing data, or not reported	Very weak measures	0
24	Adequate design for research	Small number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★
25	Weak design for research question	Very small number of cases (per group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★
26	Adequate design for research	Adequate number of cases (per comparison group)	Some missing data, possible impact on findings	Standardized, independent, some errors	★★★

27	Weak design for research question	A trivial scale of study	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	★★★
28	Weak design for research question	A trivial scale of study	High level of missing data, clear impact on findings	Weak measures, high level of error, or many outcomes	★
29	Adequate design for research	Small number of cases (per comparison group)	Moderate missing data, likely impact on findings	Not standardized/ independent errors	★★

3.4.2 Cumulative confidence across studies

Figure 8 depicts the cumulative confidence across studies based on each category. 58% (n=17) of the studies received 0-2 stars thus receiving a 'weak' weight of evidence. 37% (n=11) of studies received 3 stars thus receiving a 'moderate' weight of evidence and 3% (n=1) of the studies received the full 4 stars and receiving a 'strong" weight of evidence.

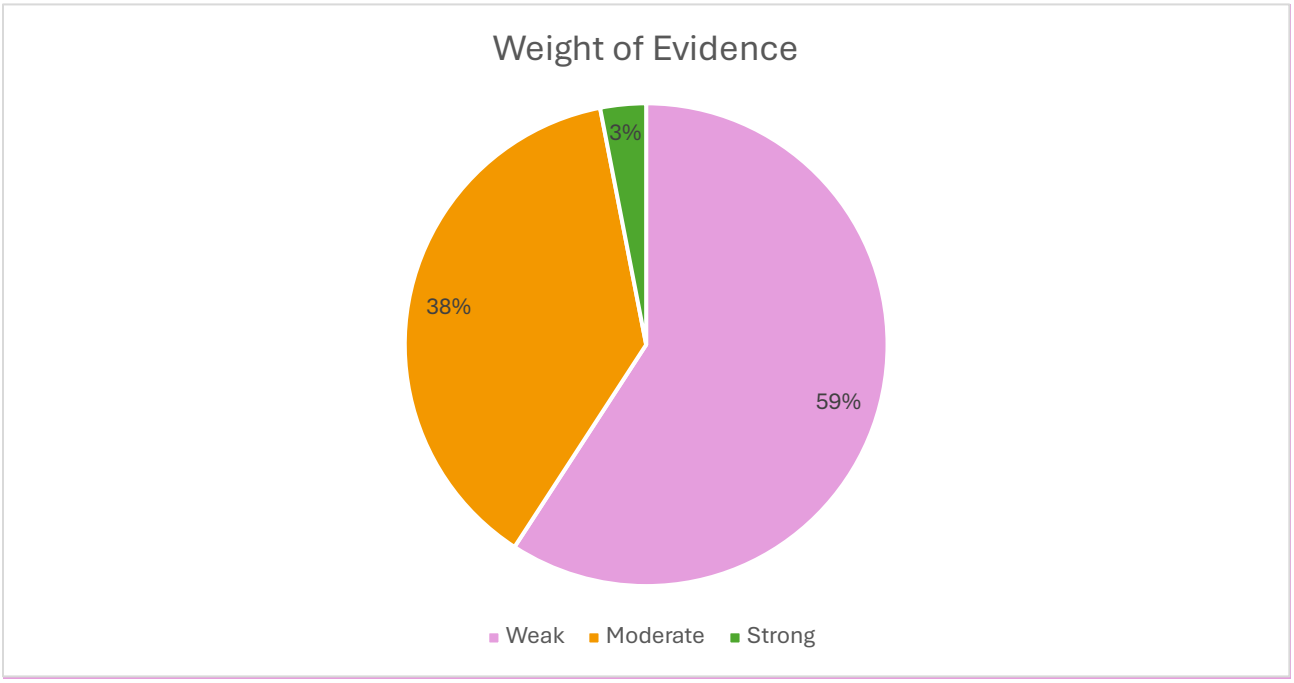


Fig 8. Weight of evidence based on quality appraisal

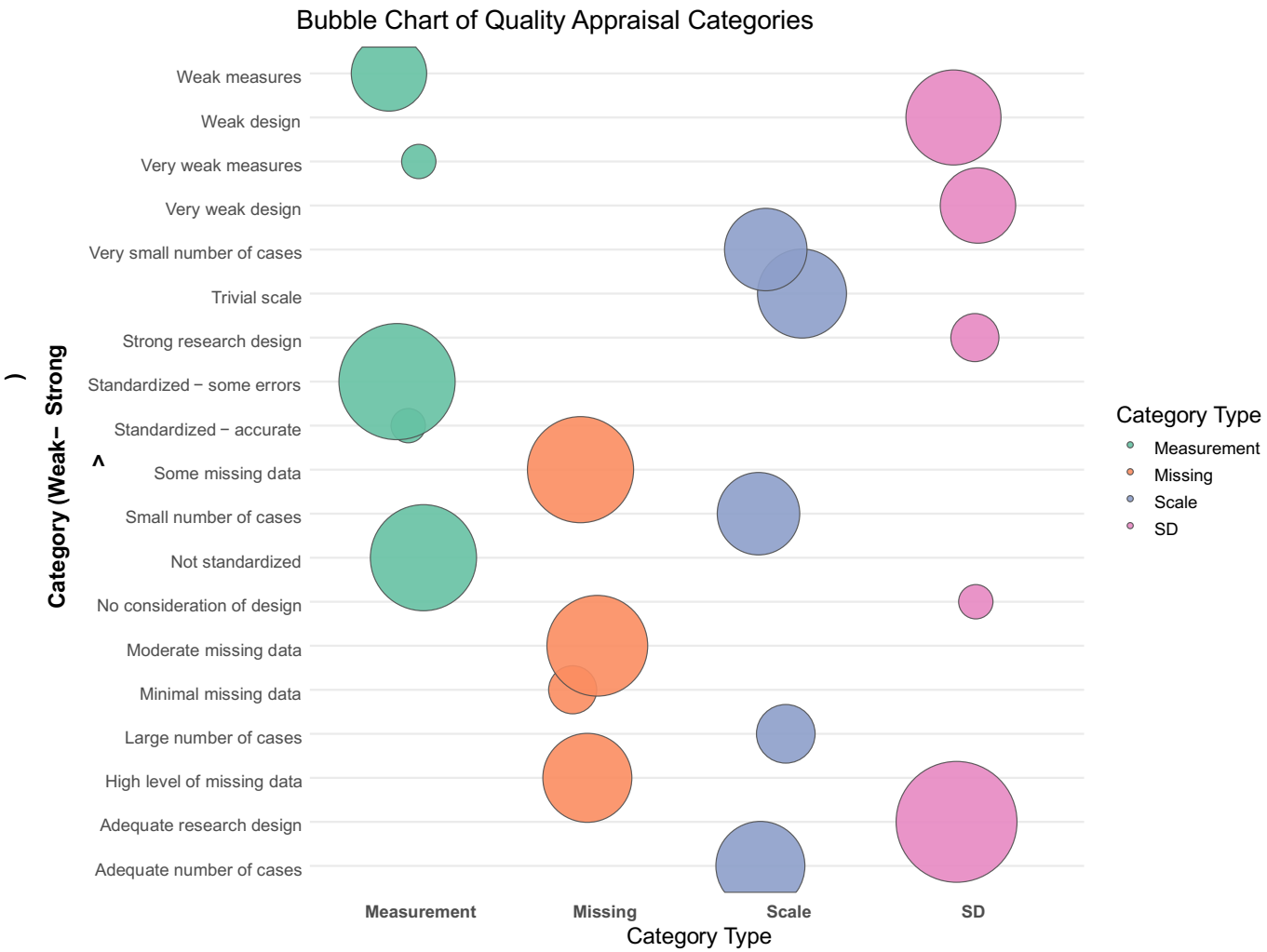


Fig. 9 Bubble "heat" chart of Quality Appraisal categories.

Measurement= Measurement Quality

Missing= Missing Data

Scale= Sample Scale

SD= Study Design

Figure 9 depicts a "heat" map of the categorical variables that were used during the quality appraisal process. The map has been categorized from weak to strong to further visualize the quality of the included studies. Based on this chart, most of the studies demonstrated lower

ratings thus receiving a ‘weak’ score. The areas with large bubbles indicate a higher frequency of the ratings. Further detailed comments of each study can be found in tables 9-13.

Chapter 4: Discussion

This chapter provides a deeper analysis of the results of this review. Section 4.1 provides analysis on the geographical and educational contexts of the included studies. Section 4.2 discusses the study designs that the included studies employed. Section 4.2.1 discusses each of the focal language learning outcomes of reading, speaking, listening, writing and additional measures of phonological awareness and vocabulary knowledge. Section 4.3 provides information on the risk of bias along with the quality appraisal and measurement of trustworthiness. The final section, 4.4, discusses the limitations of this systematic review. The goal of this review is to ascertain the totality of evidence pertaining to explicit phonics instruction in a foreign language context. The results of the review have provided insight into this area of study.

4.1 Geographical and Educational contexts

The trend in research reveals that implementing explicit phonics instruction in a foreign language context began in the early 2000's and peaked with the most research being published in 20132015. The most recent study occurred in 2024 suggesting that research is still relevant and ongoing. Of the 29 included studies, 41% of the studies were conducted in Asia, 21% in the United States, 14% in Europe, 14% in the Middle East, 7% in Africa, and 3% in South America. Across the studies, the majority of educational contexts include 41% of studies are around primary school and 41% of studies are in university contexts. Based on this information, explicit phonics instruction is taught to children and adults with little focus in secondary contexts. Interventions in other educational settings could be explored in future research.

4.2 Study Design

The focus of this review exclusively identified quantitative methods to ascertain what can be reliably concluded about the effects of explicit phonics instruction on language-learning outcomes. In this case qualitative methods and study designs have been avoided to ensure standardization, replication, and to reduce bias when drawing implications (Malterud, 2001). Of the 29 included studies, 51% (n=15) employed a non-equivalent groups pre/post-test design, 17% (n=5) were cluster RCTs, 17% were RCTs, 6% (n=2) comprised a single group pre/posttest (no control) design, 3% (n=1) employed a between-subjects experimental design with matched groups, and 3% (n=1) employed a stratified matched-pair RCT. Further information is summarized in Table 8.

Studies that use a pre/posttest design are essential in measuring before and after effects of the intervention (Alam, 2019), thus providing reliable assessment of the impact of phonics instruction in a foreign language context. Study design is critical in determining reliability because it provides a clear indicator of how an intervention affected the sample. Without a pretest, it is impossible to accurately assess baseline abilities or to determine the magnitude of change resulting from an intervention. Similarly, the absence of a control group makes it difficult to attribute any observed or statistical improvements. High-quality study designs such as cluster RCTs or stratified matched-pair RCTs can help address these limitations by ensuring comparability between groups and strengthening the validity of the findings. Quantitative designs also allow for statistical analysis of outcomes thus enabling researchers to make evidence- based conclusions regarding the effectiveness of explicit phonics instruction.

4.2.1 Language-Learning Outcome Measures

4.2.1.1 Reading Measures

The language learning outcome measures include reading, speaking, listening, and writing. Table 9 summarizes reading measures along with the quality and trustworthiness rating. The largest group of studies, 48% (n=14), reported reading measures. Of those studies, 85% (n=12) claimed positive effects of explicit phonics instruction on reading outcomes. Two studies, no.7 and no. 22, reported mixed findings of phonics instruction that either showed no change or very little effect on participants' reading outcomes. Despite many studies claiming positive findings, the quality appraisal demonstrated that the trustworthiness of the studies is either 'moderate' or 'weak'. Five studies had no research questions and were inadequately reported, making it hard to determine the focus of the studies. 64% (n=9) of the studies discussing reading measures had a 'weak' rating which brings the reliability of their findings into question. According to Gorard (2024), weaker rated studies tend to be more positive about the effects of the intervention of the study. For studies where this is the case when authors make generalizations or conclusions, they should be interpreted cautiously. It is also important to note that within the context of foreign language studies, the amount of exposure to the target language is limited and the longest study duration in this category was 7 months. Due to the shorter duration of the interventions, the amount of exposure to the language demonstrates that the effectiveness of the intervention

should be interpreted cautiously because the amount of exposure is crucial during foreign language learning (Lightbown and Spada, 2013).

Studies no. 1, 11, 14, and 23 used pre-existing phonics programs including the Jolly Phonics and Tucker Signing Phonics programs that have been explored throughout first language studies (Goouch and Lambirth, 2007). Based on these results, these phonics programs can be used in foreign language contexts, but further research should be conducted to ensure the reliability of the findings. The studies primarily used explicit phonics instruction for phonological decoding to support reading and prepare learners to read unfamiliar words and enhance reading speed. The summary of findings for this category includes increased in word reading speed, identifying pseudowords, phonological decoding, literacy skills, and fluency. Overall, phonics instruction for reading seems to be the most studied outcome in a foreign language context and has some empirical evidence for its effectiveness. Based on this information, phonics instruction should be explored more to identify its fullest potential in this context.

4.2.1.2 Speaking measures

Table 10 summarizes speaking outcome measures along with the quality and trustworthiness rating. Speaking encompasses pronunciation and speech production (Bailey, 2003). During this investigation, speaking outcomes were the most surprising because it was the second largest group at 20% (n=6) of the studies reported. Of these studies, three reported mixed or no effects, while only two reported positive effects of explicit phonics instruction on speaking outcomes. Four studies in this category scored ‘weak’ and only one scored “moderate” on the quality appraisal. Three studies, no. 13, 17, and 25 studies claimed positive effects the methods of phonics instruction. One study, no. 13, had the longest study duration of all included studies at 1 year long. Study no. 13 had the highest quality rating of this category suggesting that longer exposure to an intervention could lead to greater learning outcomes. However, due to the lack of quality and reliability of most studies in this category, explicit phonics instruction for speaking outcomes may not be as effective or would need careful consideration and time to determine further effects in a foreign language context. This finding was the most surprising as the investigation of speaking outcomes inspired the creation of this review because during the initial investigation, speaking outcomes were the hardest to find literature around and speaking is often a harder language skill to acquire (Bailey, 2003). The summary of these findings include an

improvement of speech production, pronunciation, phonological awareness, reduction of speech errors.

4.2.1.3 Listening measures

Table 11 summarizes listening outcome measures along with the quality and trustworthiness rating. In this category, one study reported listening outcomes. Study no. 27 claimed positive findings for explicit phonics instruction for the perception of Arabic sounds. This study received a 'moderate' score on the quality appraisal. Despite claiming positive effects, the study had no clearly stated research questions and a trivial sample size (n=24), which may limit the generalizability of the findings. Sound Perception test scores revealed that the experimental group improved on their percentage of correct scores during the posttest. Specifically, the average test score for the experimental group was 80.6% versus the control group's average at 68.4%, indicating that there was a statistical improvement in their sound perception. Despite these findings, the limited sample size and lack of research questions demonstrate that caution should be exercised when interpreting these results. While this information warrants caution, there is some evidence for explicit phonics instruction for listening outcomes through enhancing the learners' ability to accurately perceive and discriminate unfamiliar sounds in a foreign language.

4.2.1.4 Writing measures

Table 12 summarizes writing outcome measures along with quality and trustworthiness rating. In this category, one study reported writing outcomes. Study no. 12 claimed positive findings for explicit phonics instruction for the improvement of spelling for English learners with varying L1 backgrounds. This study employs a systematic phonics approach and positive claims demonstrate some evidence of phonics instruction supporting writing skills with varying L1 backgrounds. The sample size (n=158) provides an adequate sample with 12 different L1's including Systematic Phonics program for English learners with varying L1 background including: French, German, Italian, Portuguese, Spanish, Turkish, Vietnamese, Korean, Arabic, Thai, Japanese, and Chinese, provides some evidence on the effects of phonics instruction in a foreign language context. Although findings claim that participants with Roman-alphabetic L1 backgrounds showed greater improvements, the experimental group including all other L1 backgrounds outperformed the control group. However, this study receiving a 'weak' score on the quality

appraisal due to the lack of research questions and reported findings. It is evident that more research needs to occur in order to make reliable conclusions on the way explicit phonics instruction effects writing.

4.2.1.5 Additional measures

Table 13 summarizes other measures and outcomes identified from the included studies. Approximately 24% (n=7) of the studies reported additional measures beyond the focal outcomes of reading, speaking, listening, and writing outcomes, specifically focusing on phonological awareness and vocabulary knowledge. Among these, study, no. 10, received the only 'strong' score of the quality appraisal highlighting its methodological rigor and reliability. Study no. 10 also claimed positive effects of explicit phonics instruction on decoding skills and vocabulary knowledge. Based on this study, there is some evidence that shows clear, positive effects and provides reliable information on this topic. This study provides some of the most reliable evidence in this review, highlight the potential benefits of systematic phonics instruction in a foreign language context.

Moreover, four studies received a 'moderate' score and two studies received a 'weak' score. Among these studies, the majority reported positive findings thus demonstrating that explicit phonics instruction was associated with an increase in phonological awareness and vocabulary knowledge. One study, no. 24, claimed no statistical effects after their intervention. Based on these findings, there is some evidence that explicit phonics instruction can enhance phonological awareness, decoding skills, and vocabulary knowledge. However, this area of research remains underexplored and further high-quality studies are needed in order to establish stronger, and more generalizable evidence regarding the effects of phonics instruction in foreign language contexts. The summary of findings include increased in phonological awareness, increased in vocabulary and receptive vocabulary knowledge, and decoding skills.

4.3 Risk of Bias, Quality Appraisal, and Trustworthiness

The Risk of Bias and Quality appraisal and trustworthiness is summarized in Table 14. Of the 29 studies, 79% (n=23) claimed positive effects on the use of explicit phonics instruction. The overall analysis of the effectiveness is limited by lack of rigorous and reliable design, data

collection, and reporting. Only one of the studies received a ‘strong’ rating using explicit phonics instruction for phonological decoding and vocabulary measures. Despite research from first language studies and the growing amount of research in second language studies, the exploration of phonics instruction in foreign language contexts lacks reliable and quality research. Most of the studies, 58% (n=17) received a ‘weak’ score on the quality appraisal. Based on quality ratings of the studies, the need for trustworthy and reliable research is clear. It is evident that in foreign language contexts, the amount of language exposure is extremely important because there is a limited amount of input that learners receive (Lightbown and Spada, 2013). Based on the studies included, the shortest intervention lasted 3 weeks. With only one study lasting one year, this demonstrates that there needs to be further research that focuses on the long-term effects of explicit phonics instruction.

To address the research question, the studies that this review identified provide some exploration into the topic of explicit phonics instruction in a foreign language context. The main effects that can be concluded about the use of explicit phonics instruction is that there is a lack of reliable research and reporting to make any generalizations about the use of phonics. Despite these instances, there is a small number of studies that provide some evidence to suggest that phonics instruction has an effect on foreign language learning. However, there is a need for further investigation into these effects in order to provide reliable information to other researchers and teachers. Due to the claims and evidence from first language, along with insight from the growing number of second language research, foreign language studies require more reliable research. The limited body of reliable research on phonics in foreign language contexts presents challenges in understanding how instructional approaches translate across different languages and learning environment. While first and second language research provides a relatively strong foundation suggesting that phonics instruction can support decoding, reading fluency, and language development, foreign language contexts provide additional complexities such as varying orthographies, phonological systems, as well as levels of exposure. Furthermore, the implications for teaching are substantial as educators depend on clear, evidence based guidance to implement instructional strategies that support their learners language skills. From a teaching perspective, evidence informed practice cannot occur if the research that exists within the field is unclear, lacks reliability, and is consequently untrustworthy. Ensuring access to reliable research is crucial because teachers and educational practitioners who engage with

research require rigorous and trustworthy evidence to make informed decisions that directly benefit their learners' language development. Access to quality evidence enables educators to implement strategies that are grounded in reliable and comprehensible research allow for the improvement of teaching methods and practices.

4.4 Limitations

There are some inevitable limitations in this review. Although the search strategy was deliberately designed to be comprehensive while aiming to identify every available study examining the use of explicit phonics instruction in a foreign language context, the inclusion criteria were restricted to quantitative research design. As a result of this methodological limitation, studies employing qualitative or mixed-method approaches where excluded in this review. Additionally, the searches targeted studies published in English which may have possibly overlooked research conducted in other languages. Searches also primarily targeted language learning outcomes such as reading, speaking, listening and writing, and did not target any additional outcomes that may be affected by explicit phonics instruction. For example, six studies reported other measures including phonological awareness and vocabulary knowledge. Additional searches, or broadening the research scope further could possibly reveal other Although there were studies that involved other measures that are included in this review, there could possibly other measures that are affected by phonics instruction.

During the screening process, incorporating ASReview posed several challenges that impacted collaboration efforts. Using the software required downloading the program and having access to Python (https://www.python.org/), a computer programming language, in order to effectively interact with the tool. This technical requirement may create barriers for some researchers which limits the broader applicability and replicability of the screening process. Additionally, this review may be subject to bias due to the omission of five unretrievable records during the full-text screening stage, which may contain relevant information and data that might influence the overall conclusions. However, based on the quality ratings of the included studies, the current lack of reliable evidence pertaining to this area of research is unlikely to substantially alter the findings of this review. While the included studies provide only limited insight, they provide the beginning process in understanding the use of explicit phonics instruction in foreign

language contexts. Addressing these procedural challenges in future reviews could strengthen confidence in findings and support the development of a more robust evidence base.

Chapter 5: Conclusion

This systematic review investigated the totality of information regarding the effectiveness of explicit phonics instruction in foreign language studies. Based on the literature, it is evident that explicit phonics instruction has been supported by first language studies and a growing number of second language research, highlighting its foundational role in language development. This area of research as been supported by policymakers across the United states and the United Kingdom, with initiative emphasizing that structured phonics preprograms are essential components of early literacy and decoding. These assertions provide a rationale for investigating phonics instruction within a foreign language context as educators seek evidence based strategies to support learners throughout their language acquisition journey.

Regarding the extent of empirical studies that explored this area, the review included 29 studies that were published between 2002 and 2024. The geographic distribution of the included studies was diverse, with the highest concentration in the United States (six studies), followed by China (four studies), India (three studies), Malaysia (two studies), Spain (two studies), and one study from each of the several other countries. Most studies focused on primary and university educational contexts, suggesting that there is a particular emphasis on the foundational language skills along with advance foreign language learning. Across these studies, authors reported overall positive effects of explicit phonics instruction on reading, speaking, listening, and writing outcomes. Reading measures comprised the largest group, suggesting that much of the research on explicit phonics instruction is designed to support reading development. Positive reading outcomes included increased fluency, faster reading speed, and an improvement of decoding pseudowords. Speaking outcomes comprised the second largest group of the focal language learning outcomes. These findings indicated enhances speech production, improvement on pronunciation, and reduced speech errors. Listening outcomes demonstrated improvements in

sound perception of unfamiliar words and a greater ability to discriminate phonological contrasts in the target language. Writing outcomes showed gains for learners whose L1 shared similar alphabetic features within the target language. Studies reporting additional measures suggested that phonics instruction increased phonological awareness, decoding skills, vocabulary knowledge, and receptive vocabulary knowledge.

The research question driving this review aimed to ascertain the extent and reliability of evidence regarding explicit phonics instruction in foreign language contexts. While some evidence suggests that phonics instruction has beneficial effects, the overall quality of research raises concerns. Most studies (n=17) received a 'weak' quality rating, indicating limitations in the study design, reliability, and validity. Only one study received a 'strong' rating, demonstrating that rigorous research is possible, but currently limited. These findings have important implications for educators and researchers. In order for teachers to confidently adopt evidence-based practices, research must provide robust and trustworthy findings. Future investigations should focus on long-term effects, as foreign language learning requires sustained exposure over time as well as adopt stronger methodological designs to clarify how phonics instruction affects learning outcomes across reading, speaking, listening, and writing. Expanding research in this field will support both the development of pedagogical strategies as well as the broader advancement of foreign language education.

References

Alam, T. G. M. R. (2019). Comparative analysis between pre-test/post-test model and post-testonly model in achieving the learning outcomes. *Pakistan Journal of Ophthalmology, 35*(1).

Alba, J. M. (2021). *Synthetic phonics strategies to improve reading and writing for English as a second language students*(Master's thesis, Greensboro College).

Ahrens, A., & Zaščerinska, J. (2014). Factors that influence sample size in educational research. *2014 ATEE Spring University proceedings Changing Education in a Changing Society*, 19-32.

Aromataris, E., & Riitano, D. (2014). Constructing a search strategy and searching for evidence. *Am J Nurs, 114*(5), 49-56.

Bailey, K. M. (2003). Speaking. *Practical English language teaching*, 47-66.

Beck, I. L., & Juel, C. (1995). The role of decoding in learning to read. *American educator, 19*(2), 8.

Birch, B. M., & Fulop, S. (2020). *English L2 reading: Getting to the bottom*. Routledge.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic analysis, 20*(1-2), 3-49.

Choe, S., Lee, K., & So, Y. (2020). The effects of phonemic awareness instructions on L2 listening comprehension: A meta-analysis. *Journal of Asia TEFL, 17*(4), 1294-1309.

Cherry, M. G., Boland, A., & Dickson, R. (Eds.). (2024). *Doing a systematic review : a student's guide* (3rd edition.). SAGE Publications.

Connelly, V., Johnston, R., & Thompson, G. B. (2001). The effect of phonics instruction on the reading comprehension of beginning readers. *Reading and Writing, 14*(5), 423-457.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490. doi:10.1017/S026144480800551X

Durgunoğlu, A. Y., & Öney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing, 11*, 281-299.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of educational research, 71*(3), 393-447.

Ehri, L. C. (2020). The science of learning to read words: A case for systematic phonics instruction. *Reading research quarterly, 55*, S45-S60.

Gillon, G. T. (2004). *Phonological awareness : from research to practice*. Guilford Press.

Goouch, K., & Lambirth, A. (2007). *Understanding phonics and the teaching of reading : critical perspectives* (1st ed.). McGraw Hill/Open University Press.

Gorard, S. (2024). Judging the relative trustworthiness of research results: How to do it and why it matters. *Review of Education, 12*(1), e3448.

Hamilton, C., Schulz, J., Chalmers, H., & Murphy, V. (2024). Investigating the substantive linguistic effects of using songs for teaching second or foreign languages to preschool,

primary and secondary school learners: a systematic review of intervention research. *System, 124*(103350). https://doi.org/10.1016/j.system.2024.103350

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading. *Language, speech, and hearing services in schools, 36*(4), 285293.

Huang, X., & Chalmers, H. (2023). Implementation and effects of pedagogical translanguaging in EFL classrooms: A systematic review. *Languages, 8*(3), 194.

Huo, S., & Wang, S. (2017). The effectiveness of phonological-based instruction in English as a foreign language students at primary school level: A research synthesis. In *Frontiers in Education* (Vol. 2, p. 15). Frontiers Media SA.

Johnson, N., & Phillips, M. (2018). Rayyan for systematic reviews. *Journal of Electronic Resources Librarianship, 30*(1), 46-48.

Lightbown, P., & Spada, N. M. (2013). *How languages are learned* (Fourth edition.). Oxford University Press.

Macaro, E. (2007). Do near-beginner learners of French have any writing strategies? *The Language Learning Journal, 35*(1), 23–35. https://doi.org/10.1080/09571730701315600

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The lancet, 358*(9280), 483-488.

Martínez, A. M. M. (2011). Explicit and Differentiated Phonics Instruction as a Tool to Improve Literacy Skills for Children Learning English as a Foreign Language. *GIST Education and Learning Research Journal, 5*, 25-49.

Mesmer, H. A. E., & Griffith, P. L. (2005). Everybody's selling it—But just what is explicit, systematic phonics instruction?. *The Reading Teacher, 59*(4), 366-376.

Moats, L. C., & Tolman, C. (2004). *Language essentials for teachers of reading and spelling*. Longmont, CO: Sopris West Educational Services.

Monro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 45*(1), 73-97.

Multon, K., & Coleman, J. (2018). Inter-rater reliability. In *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vol. 4, pp. 863-865). SAGE Publications, Inc., https://doi.org/10.4135/9781506326139.n344

Murphy Odo, D. (2021). A meta-analysis of the effect of phonological awareness and/or phonics instruction on word and pseudo word reading of English as an L2. *SAGE Open, 11*(4), 21582440211059168.

Møller, A. M., & Myles, P. S. (2016). What makes a good systematic review and meta-analysis?. *BJA: British Journal of Anaesthesia, 117*(4), 428-430.

Norris, J. M. (2006). The Issue: The Why (And How) of Assessing Student Learning Outcomes in College Foreign Language Programs. *The Modern Language Journal, 90*(4), 576–583. http://www.jstor.org/stable/4127045

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj, 372*.

Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., ... & Wagner, R. K. (2020). How the science of reading informs 21st-century education. *Reading research quarterly, 55*, S267-S282.

Peyton, J. K. (1998). ASL as a Foreign Language. *The ERIC Review, 6*(1).

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Reinking, D., Hruby, G. G., & Risko, V. J. (2023). Legislating phonics: Settled science or political polemics?. *Teachers College Record, 125*(1), 104-131.

Rose, J. (2006). *Independent review of the teaching of early reading* (Nottingham, DfES Publications).

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language awareness, 20*(1), 45-59.

Scull, J., & Lyons, D. (2024). Teaching phonics in context—stories of teachers' practice and students' outcomes. *The Australian Journal of Language and Literacy, 47*(2), 181-201.

Schulz, J., Hamilton, C., Wonnacott, E., & Murphy, V. (2023). The impact of multi-word units in early foreign language learning and teaching contexts: A systematic review. *Review of Education, 11*(2), e3413.

Sitthitikul, P. (2014). Theoretical review of phonics instruction for struggling/beginning readers of English. *PASAA, 48*(1), 113-126.

Stahl, S. A., Duffy-Hester, A. M., & Stahl, K. A. D. (1998). Everything you wanted to know about phonics (but were afraid to ask). *Reading research quarterly, 33*(3), 338-355.

Swank, L. K., & Catts, H. W. (1994). Phonological awareness and written word decoding. *Language, speech, and hearing services in schools, 25*(1), 9-14.

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education, 34*(2), 208-238.

Treiman, R. (2018). What research tells us about reading instruction. *Psychological Science in the Public Interest, 19*(1), 1-4.

Umezinwa, R. N., & Udogu, C. O. (2018). Effect of jolly Phonics on primary school pupils achievement on reading and spelling skills in Anambra state. *European Journal of Research and Reflection in Educational Sciences Vol, 6*(6).

Urbanová, L. (2017). Effectiveness of teaching synthetic phonics to EFL students.

Woore, R. (2021). Teaching phonics in a second language. *Debates in Second Language Education* (pp. 222-246). Routledge. https://doi.org/10.4324/9781003008361-15

Wyse, D., & Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal, 34*(6), 691-710.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental science, 9*(5), 429-436.

Zsargo, L. (2021). Learning to read in the era of systematic synthetic phonics as prime: young children's views on their experiences (Doctoral dissertation, University of Huddersfield).

Appendix A: IDESR Protocol

Title

Investigating the Effects on Language-learning Outcomes of Phonics Instruction in Foreign Language Teaching: A Systematic Review

Review Question

What research has been conducted that investigates the effects of explicit phonics instruction on listening, speaking, reading, and writing outcomes of learners of foreign languages?

Rationale

Phonics is integral to the interaction and acquisition of any language. As highlighted by the United Kingdom Government, phonics is a part of the language 'pillar' that supports linguistic and language development (Ofsted, 2021), highlighting its role in education.

By indicating that phonics instruction is an essential component of language learning, why do some learners feel unprepared to participate in the target language community? Despite assertions in curriculum and policy, explicit phonics instruction may not be given as much attention within a language classroom as pedagogy might suggest (Levis, 2005). This systematic review seeks to identify the totality of evidence regarding phonics instruction from intervention studies where phonics instruction is investigated in FL and EFL contexts. Therefore, this systematic review will seek research that investigates the effects of phonics teaching on linguistic outcomes. The synthesis of studies will include categorizing them by the linguistic outcomes of listening, reading, speaking, and writing. The goal is to highlight the existing research that involves explicit phonics instruction and its effects on substantive linguistic outcomes in order to assess what literature already exists and what can be reliably concluded about the effects of phonics instruction on language-learning outcomes. This review also seeks to highlight any gaps in research in relation to phonics instruction to suggest research for the future of phonics instruction within language acquisition. Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. TESOL quarterly, 39(3), 369377. Ofsted. 2021. Curriculum research review for languages (OCRR). https://www.gov.uk/government/publications/ curriculum-research-review-serieslanguages).

Inclusion Criteria

This systematic review will include research that investigates the effects of phonics teaching on linguistic outcomes.

Bibliographic information

Include 1: Studies with a full reference or sufficient information.

Exclude 1: Studies with insufficient bibliographic information.

Rationale: Without sufficient bibliographic information, retrieval of works cannot occur.

Date of publication

Include 2: Published on any date.

Rationale: Attempting to collect the totality of available evidence that exists.

Participants

Include 3: Studies on typically developing foreign language learners. Include studies even if no explicit reference is made to learning ability if reasonable assumption can be made that participants are comprised mainly of typically developing individuals.

Include 3: Studies conducted in all educational settings including all age groups from preschool to university and adults.

Exclude 3: Studies that exclusively target non-typically developing learners or learners with Developmental Language Disorder.

Rationale: This review seeks to examine existing phonics interventions in a FL, EFL, and SL. The findings for non-typically developing populations may not generalize to a larger population, thus such results will not be extrapolated or included in this review. This also focuses on the outcomes of phonics instructions in all educational settings.

Language Context and Publication

Include 4: Studies conducted in FL (foreign language), EFL (English as a Foreign Language), and SL (second language) context.

Exclude 4: Studies conducted on first language acquisition.

Rationale: This review seeks to examine the impact of phonics instruction in FL and EFL classrooms and will not be investigating first language acquisition. SL is retained in the search to minimize the risk of missing papers that do not have a clearly defined language learning context.

Intervention

Include 5: Studies where explicit phonics interventions are implemented in the classroom and/or studies where phonics is taught explicitly as part of a wider intervention (even if not explicitly labeled phonics)

Exclude 5: Studies that do not include phonics interventions or explicit phonics teaching.
Rationale: This review seeks to examine the totality of evidence pertaining to phonics interventions and teaching, even where it is not labeled as phonics.

Outcomes

Include 6: Primary research studies reporting any language acquisition outcome of phonics instruction measuring one or more specific outcomes of phonics instruction, including but not limited to: speaking, reading, and phonological decoding, vocabulary knowledge, listening, writing, and spelling.

Exclude 6: Systematic reviews that only report secondary data.

Rationale: A synthesis of empirical findings in this field of literature could not occur without the reporting and evaluation of any data.

Study Design

Include 7: Experimental and quasi- experimental designs

Exclude 7: Non-experimental designs (e.g., case studies, observational studies)

Rationale: Since this review seeks to ascertain what can be reliably concluded about the effects of explicit phonics instruction on language-learning outcomes, study designs where a causal implication can be draw are included and those designs which are not suited for drawing causal implications are excluded.

Publication status

Include 8: Grey literature (i.e. Master's and Doctoral Theses and dissertations; book chapters; articles from practitioner magazines that we can access)

Exclude 8: Do not exclude studies based on publication status.

Rationale: This paper seeks to offset potential publication bias by including a wider range of research, including grey literature. We include practitioner magazines because we aim to look at classroom practices and interventions. Theses and dissertations have been reviewed by examiners.

Information Sources

The information sources and databases collected for this systematic review encompass the fields of education, linguistics, psychology, as well as multidisciplinary sources that cover wider fields of the social sciences. The list was created in support of the Bodleian Library Librarians. All databases can be accessed electronically via the University of Oxford's Bodleian Library.

Web of Science (Social Sciences Citation Index (SCCI)

Education Collection (including ERIC)

SCOPUS

Linguistics Collection (including LLBA)

British Education Index

PsycInfo

ProQuest Dissertations

Theses Google Scholar

Search Strategy

For this review, the search strategy was developed with support from Catherine Hamilton, Robert Woore, and the Bodleian Library librarians. The search strategy was adapted from Robert Woore's previous protocol for a systematic review (registered on IDESR.org under protocol number IDESR000023). With his permission, similar search terms are used for this research.

To ascertain whether there are any existing systematic reviews on this topic, a search was conducted on the ProQuest Education Database using the following search string. There were no pertinent results out of 27 hits.

Boolean Search Stream:

noft("second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL) AND noft(teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program*) AND noft(phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding) AND abstract("systematic review" OR "meta analysis" OR "research synthesis" OR "rapid evidence assessment")

The final search string is made up of four elements: language learning context (second or foreign language); teaching program or intervention (teaching or training); linguistic element (phonics, phonetics, phonological decoding); study design (experiment, randomised control).

Language Learning Context:

"second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL

AND (Teaching Program or Intervention)

noft(teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program* AND (Linguistic Element)

noft (phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding) AND (Study Design)

noft(experiment* OR quasi-experiment* OR intervention OR RCT OR randomized control* trial OR randomised control* trial OR Regression Discontinuity Design OR RDD OR compar* OR evaluat*)

Boolean Search String:

noft("second language*" OR SL OR ESL OR "foreign language*" OR FL OR "modern language*" OR "modern foreign language*" OR MFL OR "additional language*" OR L2 OR EFL) AND noft(teach* OR learn* OR instruct* OR pedagog* OR acqui* OR train* OR study* OR educat* OR program*) AND noft(phon* OR phonological decoding OR phonological recoding OR phonological coding OR decoding) AND noft(experiment* OR quasi-experiment* OR intervention OR RCT OR randomized control* trial OR

randomised control* trial OR Regression Discontinuity Design OR RDD OR compar* OR evaluat*)

Data Management

Once the searches are conducted, the titles and abstracts will be uploaded to ASReview LAB(https://asreview.nl/).

All records will be stored on Zotero, an open-source software that manages bibliographic data. Copies of these records will be stored on Microsoft OneDrive, on the University of Oxford secure system. All work is conducted on a MacBook Air running macOS Ventura. **Selection Process**

Initial Screening:

Screening of the titles and abstracts returned by the database search will be reviewed through ASReview and then be shared with a colleague. Titles and abstracts will be assessed for relevance against the inclusion/exclusion criteria listed above. The ASReview software shuffles through the titles and abstracts from the initial search and the researcher makes the decisions on whether to include or exclude each abstract as it is presented. ASReview will also eliminate duplicates that may arise from the databases. In this training process, the software begins to learn what is relevant based on the inclusion criteria, sifting through the remaining papers to present the next most relevant paper until all potentially relevant records are exhausted (the cut-off being set at 100 exclusions in a row). Included studies will be screened, quality assessed, and the data will be extracted according to the procedures of the selection process.

Quality assurance:

To assess quality assurance on the appropriate records, a random 10% of the titles and abstracts will be reviewed independently with a colleague. ASReview offers a blind screening functionality which ensures that neither reviewer can see the other's decisions during the screening process. The decisions for this 10% will be compared for consistency when applying the inclusion criteria. If there are any disagreements, the reviewers will meet, and discrepancies will be resolved through a discussion. In this case, consistency between reviewers is important for the review. If a decision is not made, a third reviewer will be brought in to discuss any discrepancies.

Full-text screening:

Following abstract screening, the full text of potentially relevant articles identified in the abstract screening process will be sought. The full texts that are obtained will be screened again based on inclusion criteria to prepare for data collection. The full texts will be stored on Zotero.

Data Collection Process

The sources that are marked for inclusion will be read closely and data extracted using a Data Extraction Form that was created in Microsoft Forms and piloted on two relevant records that have previously been identified. The second reviewer will independently review 10% of the data extraction process and the first and second authors will discuss any discrepancies, resolving these through discussion. Then, the first author will extract the rest of the data. The data collected from the forms will be exported onto an Excel sheet which categorizes the information.

Data Items

Data items will include information about: General Information (reference citation, author details, publication type, document source, data base), participants (L1, country, age, relevant socio-demographics), study design (observational study, RCT, experiment, descriptive, longitudinal study), data type (qualitative/quantitative, e.g. interviews, test results), intervention (descriptions, length, comparison group if present), outcomes (outcome type, outcome name, time points, measurements), limitations (descriptions of authors limitations), conclusions (descriptions of authors conclusions).

Risk of bias/trustworthiness of individual studies

Trustworthiness of individual studies will be determined using an appraisal tool that is catered toward qualitative research designs.

Quantitative Quality Assessment:

Gorard, S. (2024). Judging the relative trustworthiness of research results: How to do it and why it matters. Review of Education, 12(1). https://doi.org/10.1002/rev3.3448

Studies that use quantitative research designs can be examined and quality assessed through scalable measures. This review uses Gorard's (2024) "Sieve" entitled Judging the relative trustworthiness of research results: How to do it and why it matters. When judging the quality of research, the data presented in individual studies is based on its robustness or internal validity (Gorard, 2024). The sieve presents a 0-4 scale evaluating the design, scale, data, and measurement quality. For the purpose of this review, I will be taking Gorard (2020)'s star descriptors and combining them with the recent sieve.

4- ★★★★- all criteria are met

3- ★★★- at least three criteria are met

2-★★- at least two criteria are met

1-★- when one criteria is met

0- does not meet criteria

Data Synthesis

Following the data extraction process, the included studies will be classified according to the data extraction items that are categorized from the selection process. The data will be categorized and then synthesized by the language-learning outcomes of listening, reading, speaking, and writing. Once the Excel sheet is created from the data extraction form, the Excel sheet will be uploaded to R Studio to produce visualizations and summary statistics pertinent to the narrative synthesis. If it is determined that the review includes sufficient studies with comparable methods and outcomes to permit a statistical synthesis, a meta-analysis will be carried out. In the event that there are no studies suitable for a meta-analysis, then a sole narrative synthesis will be conducted. A narrative synthesis "relies primarily on the use of words and text to summarise and explain the findings" (Popay et al., 2006). It also produces "a summary of the current state of knowledge" to the review question (Popay et al., 2006). This synthesis will summarise and explain the findings of the systematic review, highlight the relationships within the data, as well as assess the available evidence.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Meta-biases

To minimise language bias, the selection process includes all languages. However, only studies published in English will be examined in detail. This factor will be highlighted in the limitations of the systematic review.

Confidence in cumulative evidence

AMSTAR Checklist: Assessing the Methodological Quality of Systematic Reviews

The AMSTSR Checklist will be used to assess confidence in cumulative evidence. https://amstar.ca/Amstar_Checklist.php

Appendix B: Bibliographic Information of the included studies

1. Patel, P., Torppa, M., Aro, M., Richardson, U., & Lyytinen, H. (2022). Assessing the effectiveness of a game-based phonics intervention for first and second grade English language learners in India: A randomized controlled trial. Journal of Computer Assisted Learning, 38(1), 76-89.

2. Rendón Romero, S. I., Navarro Pablo, M., & Garcia-Jimenez, E. (2021). Using phonics to develop the emergent English literacy skills of Spanish learners.

3. Abiy Zewdu Agegnehu, Mebratu Mulatu Bachore & Zeleke Arficho Ayele (2023) Effects of an explicit rime-based phonics method on the phonological awareness outcomes of Ethiopian EFL children, Cogent Education, 10:1, 2203582, DOI: 10.1080/2331186X.2023.2203582

4. Giambo, D. A., & McKinney, J. D. (2004). The effects of a phonological awareness intervention on the oral English proficiency of Spanish-speaking kindergarten children. Tesol Quarterly, 38(1), 95-117.

5. Yeung, S. S., Siegel, L. S., & Chan, C. K. (2013). Effects of a phonological awareness program on English reading and spelling among Hong Kong Chinese ESL children. Reading and writing, 26, 681-704.

6. Konerding, M., Bergström, K., Lachmann, T., & Klatte, M. (2020). Effects of computerized grapho-phonological training on literacy acquisition and vocabulary knowledge in children with an immigrant background learning German as L2. Journal of Cultural Cognitive Science, 1-17.

7. Sun, B., Zhu, B., Chen, J., & Zhou, H. (2015). The effects of phonological awareness training on the reading performance by child EFL learners in China. Chinese Journal of Applied Linguistics, 38(3), 339-354.

8. Fonseca-Mora, M. C., Jara-Jiménez, P., & Gómez-Domínguez, M. (2015). Musical plus phonological input for young foreign language readers. Frontiers in psychology, 6, 286.

9. Chu, M. C., & Chen, S. H. (2014). Comparison of the effects of two phonics training programs on L2 word reading. Psychological Reports, 114(1), 272-291.

10. Li, S., & Woore, R. (2021). The effects of phonics instruction on L2 phonological decoding and vocabulary learning: An experimental study of Chinese EFL learners. System, 103, 102677.

11. Rahman, M. N. A., Tham, K. E., & Liu, C. (2023). Implementation of Early Reading Games to Teach English Using Integrated Sound Word Method for Preschoolers. Pertanika Journal of Social Sciences & Humanities, 31(4).

12. Martin, K. I. (2024). How a phonics-based intervention, L1 orthography, and item characteristics impact adult ESL spelling knowledge. Education Sciences, 14(4), 421.

13. Kashiwagi, A., Snyder, M., & Craig, J. (2005). The role of explicit instruction in the development of L2 phonology. Journal of Asia TEFL, 2(2).

14. Dixon, P., Schagen, I., & Seedhouse, P. (2011). The impact of an intervention on children's reading and spelling ability in low-income schools in India. School Effectiveness and School Improvement, 22(4), 461-482.

15. Ghorbani, M. R., Neissari, M., & Kargozari, H. R. (2016). The effect of explicit pronunciation instruction on undergraduate EFL learners' vowel perception. Language and Literacy, 18(1), 57-70.

16. Algethami, G. (2016). Long term effect of phonetic instruction on the production of/p/by EFL Arab learners: An exploratory study. Arab World English Journal (AWEJ), 7(4).

17. Sturm, J. L. (2013). Explicit phonetics instruction in L2 French: A global analysis of improvement. System, 41(3), 654-662.

18. Hubert, M. D., & Vigil, D. (2017). Using writing to teach pronunciation: An experimental fourth-year university Spanish phonetics/phonology course. Applied Language Learning, 27(1-2), 18–40.

19. Tweedie, M. G., Johnson, R. C., Kay, D. W., & Shimoda, J. (2015). Direct Phonemic Awareness Instruction as a Means of Improving Academic Text Comprehension for Adult Language Learners. Journal of Educational Thought/Revue de la Pensée Educative, 81-102.

20. Johnson, R. C., & Tweedie, M. G. (2010). Could phonemic awareness instruction be (part of) the answer for young EFL learners? A report on the early literacy project in Malaysia. Tesol Quarterly, 44(4), 822-829.

21. Ghoneim, N. M. M., & Elghotmy, H. E. A. (2015). The Effect of a Suggested Multisensory Phonics Program on Developing Kindergarten Pre-Service Teachers' EFL Reading Accuracy and Phonemic Awareness. English Language Teaching, 8(12), 124143.

22. Ashmore, R. A., Farrier, M. J., Paulson, L. H., & Chu, X. (2002). The Effects of Phonemic Awareness Drills on Phonological Awareness and Word Reading Performance in a Later Learned Alphabetic Script.

23. Valbuena, A. C. (2014). Tucker Signing as a Phonics Instruction Tool to Develop Phonemic Awareness in Children. Gist: Education and Learning Research Journal, (8), 66-82.

24. Taylor, M. (2008). Orthographic and phonological awareness among L1 Arabic ESL learners: A quasi-experimental study. University of Phoenix.

25. Coates, R. A. G., Gorham, J., & Nicholas, R. (2017). The Efficacy of Phonics-Based Instruction of English as a Second Language in an Italian High School: A Randomised Controlled Trial. GIST Education and Learning Research Journal, 15, 29-67.

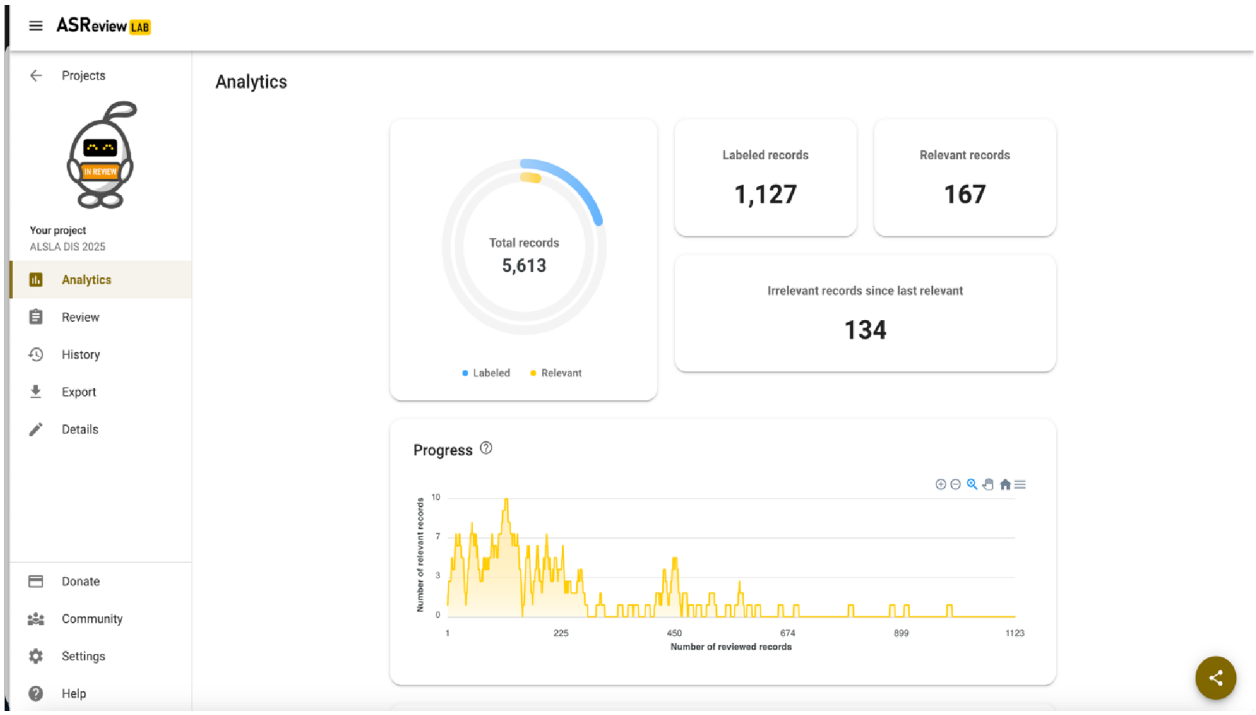
26. Cui, G., Wang, Y., & Zhong, X. (2021). The effects of suprasegmental phonological training on English reading comprehension: Evidence from Chinese EFL learners. Journal of Psycholinguistic Research, 50(2), 317-333.

27. Burnham, K. R. (2013). Phonetic training in the foreign language curriculum. Applied Language Learning, 23(63), 63.

28. Geer, L. C., & Keane, J. (2018). Improving ASL fingerspelling comprehension in L2 learners with explicit phonetic instruction. Language Teaching Research, 22(4), 439-457a

29. Nishanimut, S. P., Johnston, R. S., Joshi, R. M., Thomas, P. J., & Padakannaya, P. (2013). Effect of synthetic phonics instruction on literacy skills in an ESL setting. Learning and Individual Differences, 27, 47-53.

Appendix C: ASReview Lab Analytics



During the screening process, ASReview LAB provided analytics of the progress of screening.

Appendix D: Blank Sample Data Extraction Form

Date	Include data of extraction

Reference citation	List the full APA reference
Publication type	Full report, abstract, letter
Document source	Name of database or website
Research Questions	Copy directly from paper
Study Design	Drop down menu including: -Non-equivalent groups pre/post-test (ex. control group and experimental group) -Non-equivalent groups crossover (ex. non random groups experiencing the same treatment) -Single group pre/ post-test (no control) -RCT (randomized control trial) -Cluster RCT (cluster or groups assigned randomly) -Non-equivalent groups (post-test only) (ex. groups only measured by a post test) -Single group, post-test only - Other (describe)
Data Type	Qualitative or Quantitative
Study Duration	Include start date, end date, duration if possible
Location of Study	Country
Language of Publication	
Participants	Social and educational context
Population Description	Include any information regarding participants' learning disabilities, socioeconomic background, etc.
Languages Spoken	Indicate L1/L2
Age	Age range
Gender	Include breakdowns
Intervention: Languages of instruction	Include hours/ weeks of language study
Intervention	Quote and summarize
Length of intervention	Duration in weeks (if applicable)
Comparison	If any
Number of participants	(n= total participants, n= per experimental conditions, n= comparison)
Class grouping	Describe how they were grouped
How were groups generated	i.e. random allocation, at individual level, no report of allocation strategy, etc.
Baseline imbalances	Were there any significant differences at the start of the study?

Attrition	Did any participants leave the study?
Outcome type	What did the intervention do? Quote from the paper
Outcome name	Reading, speaking, listening, writing, other
Phonics intervention included	Systematic phonics, synthetic phonics, etc..
Unit(s) of measure	How is the outcome operationalized?
Time points measures	How many data collections?
Descriptive outcomes	Describe what was claimed during the study
Effect Sizes	Or other statistics

Appendix E: Completed Data Extraction Form

Date	June, 10th 2025
Reference citation	Patel, P., Torppa, M., Aro, M., Richardson, U., & Lyytinen, H. (2022). Assessing the effectiveness of a game-based phonics intervention for first and second grade English language learners in India: A randomized controlled trial. Journal of Computer Assisted Learning, 38(1), 76-89.
Publication type	Journal Article
Document source	Wiley
Research Questions	1. Do children who play GL significantly outperform children who do not on game-based measures of English reading skills? 2. Do children who play GL significantly outperform children who do not on oral and paper-based measures of English reading skills? 3. How does progress within GL relate to students' pre-test, post- test, and gain scores? Are there differences in effectiveness based on this relation?
Study Design	Stratified matched-pair RCT- students were randomized individually within each classroom
Data Type	Quantitative
Study Duration	Pretest- 30 minutes, Test- 5 days a week, 20 minute sessions during regular school hours. Total time was 5 weeks
Location of Study	India
Language of Publication	English
Participants	Primary school, Grade 1 and Grade 2
Population Description	Public school in Delhi, India
Languages Spoken	L1 majority Hindi
Age	6.2 years old (Range of 5- 7 years)
Gender	Experimental 35 male, 34 female, Control 34 male, 33 female
Intervention: Languages of instruction	EMI + Hindi as an additional language taught for one period each week (however, it is the home language)

Intervention	A matched pairs randomized design was used in which randomization was done within classrooms. Students within each classroom were matched on age and gender, and then randomly allocated to either the GL group (n = 69) or the control group (n = 67) which played a math game... For the intervention sessions, 25 smartphones were set up in a spare

	classroom in the school and students were brought in class-by- class, 5 days a week, for 20 minute sessions during their regular school hours. The students in the intervention group played GL English Rime. GL=GraphoLearn (systematic phonics)
Length of intervention	5 weeks, 5 days per week, 20 minutes per session
Comparison	Control group were students within the same grade but were using a different program to ensure that both groups were exposed to computer-based programs.
Number of participants	n= 136, (Experimental) n= 69, (Control) n=67
Class grouping	136 students across three Grade 1 classrooms and three Grade 2 classrooms. Different grade levels and varying ages, bringing the average to 6.2 years old.
How were groups generated	Matched pair randomized design- random allocation was done within classrooms.
Baseline imbalances	N/A
Attrition	Original total= 143 but 7 left due to technical issues in the GL program and their data did not save so the data from 136 participants was collected.
Outcome type	Assessment of English Literacy skills: letter-sound knowledge, rime unit recognition, and word recognition Rhyme identification, phoneme replacement, semantic fluency, verbal fluency, word reading, pseudoword reading, spelling
Outcome name	Literacy/ Reading
Phonics intervention included	Systematic phonics

Unit(s) of measure	In-game measures and raw scores from the program were used along with oral and paper assessments from the researchers. Letter sounds: -Students were presented individual letter sounds auditorily which they had to match with the correct written form out of the multiple options presented to them Words were presented auditorily which students were required to match with the correct written form out of the multiple options presented to them -Total number of correct responses
	Rime units: -Rime units (i.e., -ip, -at) were presented auditorily which students were required to match with the correct written form out of the multiple options presented to them -Total number of correct responses Word recognition: -Words were presented auditorily which students were required to match with the correct written form out of the multiple options presented to them -Total number of correct responses See table 3 for oral and paper-based task descriptions
Time points measures	Two time points. Pretest and Post test. Data was saved by the GL program

Descriptive outcomes	In this study, we aimed to examine whether GL English Rime, a glob- ally recognized computer-assisted reading intervention, could significantly improve the foundational English literacy skills of Grade 1 and 2 ELLs who were attending an English-medium public school in Delhi, India. At the end of a 5-week intervention period, the GL group made significant improvements, particularly on in-game assessments of letter-sound knowledge, rime unit recognition, and word recognition. Along with significantly higher post-test scores, children who played GL showed faster development across all three in-game mea- sures as compared to children who did not play GL. These results are meaningful in showing that GL was able to quickly and effectively teach letter-sound correspondences, a critical subskill for English word reading, to young ELLs in India
Effect Sizes	In-game stats: Letter Sounds: 0.5 Rime Units: 0.25 Word Recognition: 0.06 Out-of-game oral and paper-based tasks: Rhyme Identification: 0.03 Phoneme replacement: 0.02 Semantic fluency: 0.01
	Verbal fluency: 0.002 Word reading: 0.003 Pseudoword reading: 0.000 Spelling: 0.003 Bottom 50%: Letter Sounds: 0.41 Rime Units: 0.17 Word Recognition: 0.02 Top 50% Letter Sounds: 0.62 Rime Units: 0.39 Word Recognition: 0.10