



A Best-Evidence Synthesis and Meta-analysis on Effective Reading Programs in Spanish

José L. Arco-Tirado 
Francisco D. Fernández-Martín 
Mirian Hervás-Torres 
Gracia Jiménez-Fernández 
Nuria Calet 
Sylvia Defior 
University of Granada

Amanda J. Neitzel 
Robert E. Slavin
Johns Hopkins University

This systematic review and meta-analysis examined the effectiveness of Spanish reading programs in grades K–6. The research designs included were experimental and quasi-experimental. Effect sizes were analyzed using a multivariate meta-regression model with robust variance estimation. To assess the degree of heterogeneity in the effect sizes, a 95% prediction interval was calculated. A total of 11 studies and 51 effect sizes met the inclusion criteria. The full meta-regression model controlling for grade level and outcome type showed a large positive effect across all studies (effect size = 0.49, $p < 0.05$), with a large, positive effect on reading outcomes, and significant impacts on phonological awareness, phonics, fluency, and reading comprehension in K–2. Results suggest that effective instructional programs for K–6 Spanish reading exist. However, there is a need for more rigorous research on reading instructional programs for Spanish-speaking children.

KEYWORDS: educational policy, effectiveness, effect size, instructional practices, literacy, meta-analysis, program evaluation, school/teacher effectiveness, secondary data analysis, Spanish reading instruction, systematic review

Literacy skills are important to students' success in school and later in life. Literate adults report higher health and economic outcomes, civic engagement, and community well-being than nonliterate adults (Baye et al., 2018). However,

literacy skills are unevenly distributed across countries (European Commission, 2018; Organization for Economic Cooperation and Development [OECD], 2019; United Nations Educational, Scientific and Cultural Organization, 2019).

In the case of Spain, for example, the levels of school failure (European Commission, 2018; OECD, 2017), students with low reading performance (Mullis et al., 2012, 2017), youth unemployment (European Commission, 2019), and human capital underdevelopment (World Economic Forum, 2017) are higher than expected. To reverse this situation, the European Commission (European Agency for Special Needs and Inclusive Education, 2018; European Commission, 2017) and the Spanish government (Ministry of Foreign Affairs, European Union, and Cooperation, 2018) have been working on the educational objectives of the Agenda 2020 of the European Commission and the Agenda 2030 (objective 4) of the United Nations. However, the lack of significant progress in those rankings on key indicators like reading comprehension suggests that educational reforms are not bringing about the expected improvements. For some children, success depends on the Spanish region in which they live (Ministerio Educación, Cultura y Deporte, 2016) or the quality of their teachers' training (OECD, 2013). Education as a tool to correct inequalities either inherited or acquired has not worked (OECD, 2017; The Education Endowment Foundation, 2018).

If we expand our focus to include other Spanish-speaking countries, no Spanish-speaking country ranks among the top 10 most literate according to the Progress in International Reading Literacy Study report, and all Spanish-speaking countries rank below the average of OECD countries (Mullis et al., 2017). According to a prevalence study by García et al. (2013), 20% of children in Spain show difficulties in reading comprehension. In sum, this group of countries has a serious reading comprehension problem.

The Rationale for This Review

This review intends to identify effective Spanish reading instruction programs that improve student literacy in Spanish-speaking countries in grades K–6. To do so, we build upon previous reviews in Spanish focusing on this topic and the body of literature focusing on English language literacy development as outlined later.

Evolution and Recent Advances in Reading Instruction Research

A lack of a comprehensive governmental review of reading instruction in Spanish, similar to those developed in English-speaking countries (e.g., National Reading Panel [NRP], 2000), has led researchers in Spanish reading literacy to use the United States' NRP as a reference for research on reading instruction in Spanish (Crespo et al., 2018; Ripoll & Aguado, 2014). The NRP (2000) research revealed a strong scientific consensus around five key components of effective reading instruction, including phonemic awareness, phonics, fluency, vocabulary, and comprehension. The NRP continues to be widely referenced, and these key “pillars” of effective reading instruction have been further strengthened in the ensuing years, with even more studies supporting their inclusion in classrooms. However, our understanding of the development of skilled reading has become more sophisticated as we accept that a broad range of knowledge and skills are needed to become expert readers (Pearson et al., 2020). As one example, the Active View of Reading incorporates additional factors beyond the five pillars,

such as reader motivation and engagement, the importance of background knowledge, and the need for active self-regulation such as through strategy use or other types of executive functioning skills (Duke & Cartwright, 2021). However, while these additional factors may be included as essential elements of instruction, our search has not found studies of Spanish reading including them as predictors.

Integration and Application of English and Spanish Literacy Research

This foundation of English reading research has already been applied in studies of Spanish literacy instruction. For example, previous reviews of Spanish language literacy instruction (e.g., Baker et al., 2022; Balbi et al., 2018; Chávez-Delgado et al., 2022; Ripoll & Aguado, 2014) have also assumed that cracking the alphabetic code is central to learning to read in alphabetic writing systems such as Spanish or English. This transfer is supported by three decades of research in English and Spanish literacy (August et al., 2002; National Academies of Sciences, Engineering, and Medicine, 2017; Vaughn et al., 2006) that has found strong correlations between phonological skills in Spanish and English. Indeed, several authors have demonstrated the effectiveness of Spanish interventions based on the development of basic English reading skills as supported by the NRP (e.g., Pallante & Kim, 2013). Furthermore, cross-linguistic research involving both languages has shown that literacy skills identified as significant predictors of later reading success are similar for English and Spanish, including phonological processing (Bravo-Valdivieso, 1995; Caravolas et al., 2012, 2013; Carrillo, 1994; Defior & Tudela, 1994; Jiménez & García, 1995; González & Valle, 2000), decoding skills (Bravo-Valdivieso, 1995; Caravolas et al., 2019; Lindsey et al., 2003), and oral activities (Bravo-Valdivieso, 1995). Specifically, basic phonemic awareness ability is important in the beginning stages of literacy acquisition, but by first grade, phoneme manipulation is a better predictor (Carrillo, 1994), with some forms of phonemic awareness developing after the onset of reading instruction.

The utility of this systematic review, meta-analysis, and best-evidence synthesis is clear for school-based educational leaders, researchers, and policymakers in Spanish-speaking countries since it intends to provide a set of evidence-based programs ready to be used not just to improve students' reading performance but also to prevent reading problems.

Prior Reviews

Our scoping search identified three prior reviews focused on reading outcomes in Spanish-speaking settings (i.e., Baker et al., 2022; Chávez-Delgado et al., 2022; Ripoll & Aguado, 2014). Baker et al.'s (2022) study analyzed the relation between the essential components of reading and reading comprehension in monolingual Spanish-speaking children in 26 cross-sectional studies and 7 longitudinal studies. Chávez-Delgado et al.'s (2022) review included 24 studies. Finally, Ripoll and Aguado's (2014) meta-analysis of 39 studies of programs for K–12 students in Spanish-speaking countries reported a combined effect size of 0.71.

The Contribution of This Review

To improve national reading outcomes, policymakers and educational practitioners need information about high-quality, evidence-based instructional programs that can increase the number of Spanish-speaking students who read

proficiently. This systematic review contributes to this research topic by building on the prior reviews in three fundamental ways.

The first consists of adopting a conceptual framework for the review with six targeted components: the five pillars of effective instruction included in the NRP (2000) framework (i.e., phonemic awareness, phonics, fluency, vocabulary, and reading comprehension) plus a sixth component—concepts about print (Chall, 1996a, 1996b). Concepts of print means that children understand that print carries meaning, that books contain letters and words, and books “work” in a particular way.

The second contribution has to do with the quality of the review. To ensure high quality, we adopted the rigorous inclusion standards recommended by international organizations for high-quality systematic reviews. Other reviews in Spanish (like Baker et al., 2022; Chávez-Delgado et al., 2022; Ripoll & Aguado, 2014) meet only some of these standards, but our review meets them all. A thorough description of these criteria can be found in the eligibility criteria section.

The third contribution has to do with following statistical considerations that apply to technical differences resulting from the combination of the different reading effects and reading conditions across the selected studies into a single meta-analysis to estimate an overall effect on reading outcomes. The method section describing the effect sizes calculation and statistical procedures provides a thorough description and justification of those considerations.

Perhaps the most important contribution is that, to our knowledge, this will be the first comprehensive, updated, international-standards-based systematic research review, meta-analysis, and best-evidence synthesis on the effectiveness of programs focused on reading instruction in Spanish across Spanish-speaking countries covering K–6.

Research Questions

The following research questions guided this review:

Research Question 1: What is the average treatment effect across included studies of Spanish literacy instruction programs?

Research Question 2: How much variability in effect sizes can be accounted for by studies’ characteristics of grade level and outcome type?

Method

The present review uses a best-evidence synthesis approach (Slavin, 1986), which combines traditional meta-analytic techniques of systematic review and effect size calculations (Lipsey & Wilson, 2001) with narrative descriptions of individual programs and studies. To meet international standards for high-quality systematic reviews, a protocol for this review was written in advance, specifying the main objectives, key design features, and planned analyses as recommended by The Campbell Collaboration (2019) and Piggot and Polanin (2020). This explicit methodology not only ensures transparency and replicability but also helps to increase adherence to the research plan and avoid bias in the research and reporting processes. A detailed protocol of the review was registered in the OSF (Arco-Tirado et al., 2023) with the main methodological

considerations discussed later. Data and code used in the analysis are also available (Arco-Tirado et al., 2024).

Eligibility Criteria

Inclusion Criteria

Studies had to meet the following rigorous criteria to minimize bias and provide educators and researchers with reliable information on a program's effectiveness: (a) focused on Spanish as students' first language and use Spanish as the language of instruction; (b) used Spanish as the target language in Spanish-speaking countries; (c) focused on the effects of classroom/school-based Spanish reading instructional programs on quantitative measures of reading outcomes (i.e., concepts about print, phonological awareness, phonics, vocabulary, fluency and/or reading comprehension or a combination of thereof); (d) the treatment and control groups compared across the different reading conditions were equivalent in the pretest condition in all aspects except for receiving the instruction; (e) for each treatment condition, one group of children was taught one or several reading components (i.e., concepts about print, phonological awareness, phonics, vocabulary, fluency, and/or reading comprehension) while the control group received another type of instruction (i.e., regular curriculum, whole language approach, whole word approach, miscellaneous, or basal programs), involving equal time; (f) following training, the two groups were compared in their ability to read; (g) evaluated instructional practices and/or products implemented in K–6 education (i.e., 5–12 years of age); (h) applied a true- or quasi-experiment to test the instruction, with random assignment or matching with appropriate adjustment for any pretest differences of ± 0.25 standard deviation; (i) use of distal measures (not proximal measures or researcher-made due to the distortion introduced on effect sizes) (Cheung & Slavin, 2016); (j) the level of assignment was schools, teachers, or students, taking clustering into account; (k) minimum duration was 12 weeks or an approximately equivalent number of sessions, as the minimum period required for programs to show their full effect (Cheung & Slavin, 2012), as well as the cutoff value for studies with larger effect sizes (Cheung & Slavin, 2016; Galuschka et al., 2014); (l) evaluated programs that could be replicated, that is, the article gave readers enough detailed information that the research could be replicated; (m) articles had to be published or written in Spanish or English; (n) all publication types or publication status were included; (o) no cultural restrictions; (p) no geographical limits; (q) no publication time restriction; and (r) group size of ≥ 30 (Bloom, 2003).

Exclusion Criteria

Studies were excluded for the following reasons: (a) Spanish as students' second language while taught in any other language (for example, Blackford et al., 2012, since Spanish scores could be conditioned on English native level of proficiency); (b) bilingual education studies because of potential confounding effects (for example, Flores & Duran, 2016b, since it is not possible to isolate potential confounding effects arising from Spanish in this case); (c) evaluation studies aimed at Spanish-speaking English language learner students (for example, Solari & Gerber, 2008) since potential English instruction could distort the effect of

Spanish reading instruction; (d) evaluation studies including special education populations; (for example, Favila & Seda, 2010, since students sampled showed reading difficulties); and (e) studies in which the instruction was conducted by the researchers for replicability and sustainability reasons as recommended by Case et al. (2010) to control for the potential extraneous effects linked to the researchers' characteristics, and/or because of potential unrealistic levels of support that could not be maintained for a semester or more (Cheung & Slavin, 2016), respectively (for example, Bizama et al., 2013, since the intervention was delivered by the researchers themselves).

Search for Eligible Studies

Relevant studies were identified through two main search strategies, which ended on April 25, 2022. The primary search included a wide range of electronic platforms and databases: Web of Science, Proquest, Scopus, OvidSP, EBSCOhost, Taylor & Francis, Springer Link, Science Direct, REDINED, REDUC, ÍnDICES-CSIC, Redalyc, and Dialnet.

The complementary search included hand searching of included studies and reference lists of relevant reviews, relevant websites, institutions, and (evidence) networks (e.g., American Institutes for Research, Empirical Education's Investing in Innovation/Education Innovation and Research, United States Department of Education's Institute of Education Sciences, and studies associated with the projects listed on the NSF Community for Advancing Discovery Research in Education website, What Works Clearinghouse [WWC], Evidence for ESSA, The Best Evidence Encyclopedia [BEE], EPPI Centre, and Educational Evidence Portal), literature snowballing, contacting experts, personal contacts, and Google Scholar to identify potential unpublished studies. Additionally, the tables of contents of the key journals for the last 22 years and the journals from the manual revision of references from key BEE reports were examined (Table 1).

The search strategy was modified according to the specifications of each platform, database, and website. The search terms were selected using the Education Resources Information Center Thesaurus and reflected the inclusion criteria defined in the previous section. For websites or databases with basic search functions, the review team adjusted the search terms due to the limited functionality of search functions. The preferred search strategies were based on keyword searches and/or topic/theme searches. For databases/websites, which do not allow the combination of keywords, separate keyword searches were conducted for the terms. For example, the terms and strings used for the Web of Science search were: TS = (reading AND Spanish) AND (intervention* OR program* OR train* OR practice* OR treatment*) AND ("literacy achievement" OR "literacy knowledge" OR "literacy skills" OR "print awareness" OR "phonological awareness" OR "phonemic awareness" OR "phonics" OR "vocabulary" OR "fluency" OR "decoding" OR "comprehension" OR "prosod*" OR "school" OR "district*" OR "Kindergarten" OR "pre-k" OR "K-6" OR "elementary" OR "primary" OR "control group" OR "comparison group" OR "quasi-experiment" OR "true experiment" OR "randomized control design" OR "matched") NOT ("qualitative study" OR "case study" OR "action research" OR "single subject design" OR "descriptive study" OR "correlational study" OR "university" OR "high school" OR "vocational education" OR "higher education").

TABLE 1.
Journals Searched

Key Journals	Journals From the Manual Review of References From Key BEE Reports
American Educational Research Journal	Bilingual Research Journal
Bordón	Current Directions in Psychological Science
British Educational Research Journal	Dissertation Abstracts International
Comunidad Educativa	Education and Treatment of Children
Contemporary Educational Psychology	Educational Evaluation and Policy Analysis
Educational Research Review	Educational Leadership
Educational Researcher	Educational Psychology Review
Elementary School Journal	Educational Psychologist
Estudios de Psicología	Evaluation Review
Estudios Pedagógicos	Exceptional Children
European Journal of Psychology of Education	Harvard Educational Review
Infancia y Aprendizaje	International Journal of Educational Research
Journal of Educational Psychology	Journal of Advanced Academics
Journal of Educational Research	Journal of Education for Students Placed at Risk
Journal of Experimental Education	Journal of Educational and Behavioral Statistics
Journal of Learning Disabilities	Journal of Educational Computing Research
Journal of Research on Educational Effectiveness	Journal of Experimental Child Psychology
Learning and Instruction	Journal of Literacy Research
Lenguaje y Textos	Journal of Research in Reading
Psicothema	Journal of School Effectiveness and School Improvement
Review of Educational Research	Journal of School Psychology
Reading Research Quarterly	Learning Disabilities Research and Practice
Review of Research in Education	Learning Disability Quarterly
Revista de Educación	Machine-Mediated Learning
Revista de Investigación Educativa	Psychological Bulletin
Revista de Psicodidáctica	Reading and Writing: An Interdisciplinary Journal
Revista de Psicología General y Aplicada	Reading Improvement
The Spanish Journal of Psychology	School Psychology Review
	Social Science Research
	Teaching and Teacher Education
	The British Journal of Educational Psychology
	The Future of Children

Selection of Studies for Review and Coding

Training was conducted for each stage of the review and coding processes, with review team members practicing their screening, reviewing, and coding until they reached 90% agreement. Weekly meetings of the review team provided opportunities for reviewers to present decisions they made, questions they had, and challenges they faced. These decisions and issues were documented through a living codebook for all reviewers to access. The screening processes were completed using Covidence, yielding the results shown on a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021) flow chart (Figure 1).

The two main search strategies yielded a total of 9243 results. The search on databases yielded 7389 results, targeted journals 1053 results, study referrals 209 results, and hand searching 592 results. A total of 8732 results were excluded on pre-pass, resulting in 511 results as eligible studies.

The first screening consisted of eliminating studies that were obviously not eligible for inclusion based on the title and abstract (e.g., studies that are not evaluations of a reading instructional practice and/or product). Each study was assessed by a single reviewer. This stage was conducted by the authors in second, third, fourth, and fifth position, and a random sample of 10% of studies removed at this stage were rescreened by an additional reviewer to ensure coding consistency, which yielded a 99.61% of inter-rater reliability. We retrieved the full-text version of all 511 remaining studies except for three that were inaccessible and screened each using our inclusion criteria for final eligibility determination.

The second screening consisted of authors from first to sixth positions organized in pairs reading each document in full to determine if it met all inclusion and exclusion criteria. Disagreements were resolved by the first and fifth author. Average percent agreement at this stage was 90.2%. This screening process resulted in the identification of 11 studies fully meeting the inclusion criteria.

Codes were verified by three senior research team members—that is, the first, sixth, and seventh authors. Each study was coded for the following descriptors: publication type (journal, article, dissertation/thesis, conference presentation, book chapter, and other); year of publication; design type (experimental, quasi-experimental); randomized (yes, no); clustered assignment (yes, no); *N*, mean, and standard deviation of intervention and control groups at pretest; *N*, mean, and standard deviation of control group at posttest; grade (K–6), treatment intensity; teachers' training duration; urbanicity; nationality; post-test measurement instrument; and outcome variable (concept of print, phonological awareness, phonics, vocabulary, fluency, reading comprehension, and a combination of thereof). Each study was coded independently by two authors, who resolved their differences. Coding categories for reading comprehension were analyzed and discussed by the first six authors. Any initial disagreement between the trained coders was again examined by the first and seventh authors and resolved. This coding process resulted in the identification of 51 effect sizes. Because coding categories for reading comprehension were extracted through discussion and collaboration, reliability and agreement statistics are not available.

Search and Exclusion Process (*k* = Number of Studies, *n* = number of outcomes)

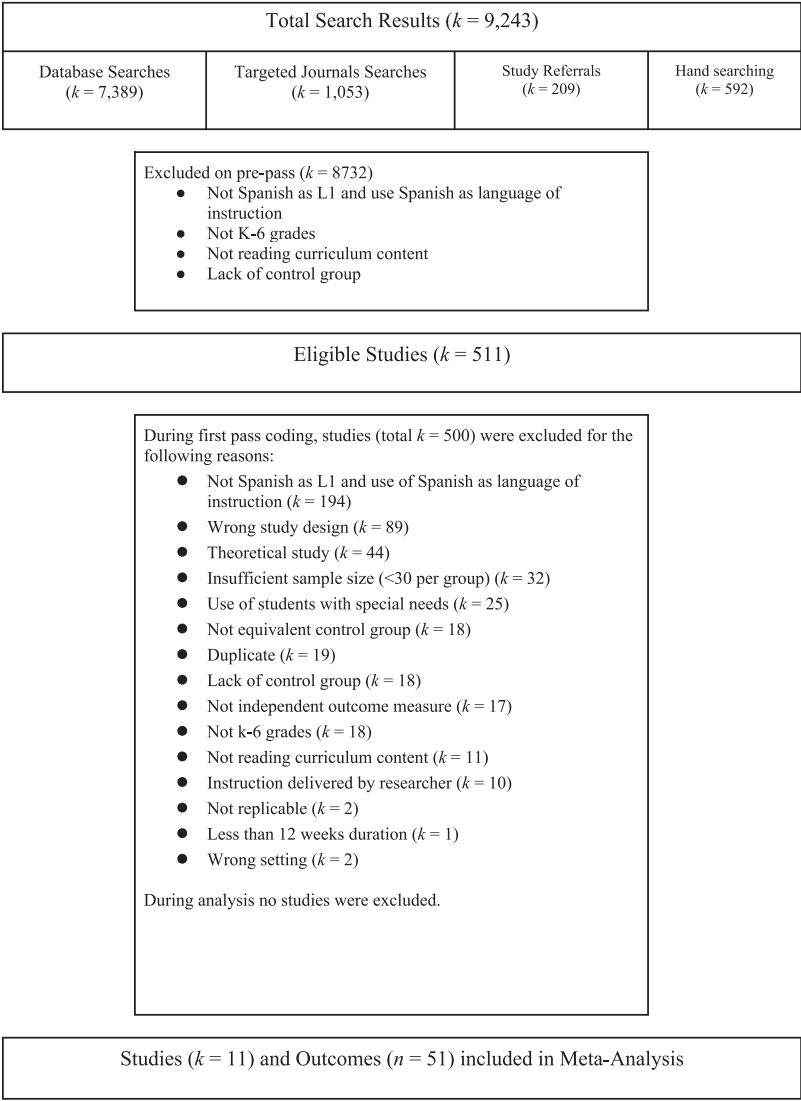


FIGURE 1. Search and exclusion process (*k* = number of studies, *n* = number of outcomes).

Effect Size Calculations and Statistical Procedures

Effect sizes were calculated in terms of Hedges’ *g*. Standardized mean difference effect sizes were calculated using procedures for Hedges’ *g* as the difference

between adjusted post-test scores for treatment and control students, divided by the pooled standard deviation of unadjusted post-test scores for treatment and control, with a correction applied for small sample sizes (Hedges, 1981). Alternative procedures were used to estimate effect sizes when unadjusted post-tests or unadjusted standard deviations were not reported, as described by Lipsey and Wilson (2001). Overall mean effect sizes were calculated across studies and programs, weighted by inverse variance, and adjusted for clustering as described by Hedges (2007).

Mean effect sizes across studies were calculated after assigning each study a weight based on inverse variance (Lipsey & Wilson, 2001), with adjustments for clustered designs suggested by Hedges (2007). The decision about conflating all six components into an overarching concept of reading outcomes (i.e., concepts about print, phonological awareness, phonics, vocabulary, fluency, and/or reading comprehension) to estimate the average reading outcome seemed consistent with the conceptual framework adopted, the studies selected, the compared reading instruction conditions, the statistical analysis conducted, and the research objectives set. In this vein, others' meta-analyses in English adopting a conceptual and statistical rationale comparable to ours include an overarching concept measurement called "foundational skills" (i.e., Wanzek et al, 2016, 2018); "reading outcomes" (i.e., Gersten et al., 2020; Roberts et al., 2020); or "norm-referenced reading outcomes" (i.e., Hall et al., 2023).

In combining across studies and in moderator analysis, we used random-effects models, as recommended by Borenstein et al. (2009).

Meta-regression

We used a multivariate meta-regression model with robust variance estimation to conduct the meta-analysis (Hedges et al., 2010). This approach has several advantages. First, our data included multiple effect sizes per study, and robust variance estimation accounts for this dependence without requiring knowledge of the covariance structure (Hedges et al., 2010). Second, this approach allows for moderators to be added to the meta-regression model and calculates the statistical significance of each moderator in explaining variation in the effect sizes (Hedges et al., 2010). Tipton (2015) expanded this approach by adding a small-sample correction that prevents inflated Type I errors when the number of studies included in the meta-analysis is small or when the covariates are imbalanced. We estimated three meta-regression models. First, we estimated a null model to produce the average effect size without adjusting for any covariates. Second, we estimated a meta-regression model with the identified moderators of interest. This model took the general form:

$$T_{ij} = \beta_0 + \beta_k X_{ij} + \beta_m X_j + \eta_j + \phi_{ij} + \varepsilon_{ij}$$

where T_{ij} is the effect size estimate i in study j , β_0 is the grand mean effect size for all studies, β_k is a vector of regression coefficients for the covariates at the effect size level, X_{ij} is a vector of covariates at the effect size level, β_m is a vector of regression coefficients at the study level, X_j is a vector of covariates at the

study level, η_i is the study-specific random effect, ϕ_{ij} is the effect size specific random effect, and ε_{ij} is the effect size specific sampling error.

All moderators and covariates were grand-mean centered to facilitate the interpretation of the intercept. All reported mean effect sizes come from this meta-regression model, which adjusts for potential moderators and covariates. The packages *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2020) were used to estimate all random-effects models with robust variance estimation in the R statistical software (version 4.2.3) (R Core Team, 2020).

To assess the degree of heterogeneity in the effect sizes, a 95% confidence interval and a 95% prediction interval were calculated for each of the full meta-regressions (Borenstein et al., 2017). A prediction interval is an estimate of an interval in which a future observation will fall, with a certain probability, given what has already been observed. Prediction intervals must account for the uncertainty in estimating the population mean, plus the random variation of the individual values. So, a prediction interval is always wider than a confidence interval. Also, the prediction interval will not converge to a single value as the sample size increases. The key point is that the prediction interval tells you about the distribution of individual values, as opposed to the uncertainty in estimating the population mean, and will not converge to a single value as the sample size increases.

The 95% prediction interval was calculated by:

$$u - 1.96\sqrt{\tau^2 + \omega^2}, u + 1.96\sqrt{\tau^2 + \omega^2}$$

Where u is the average effect size, τ^2 is the between-study variance in the effect sizes, and ω^2 is the within-study variance in the effect sizes. While robust variance estimation does not require a normality assumption, estimates of τ^2 and ω^2 are accurately estimated when the normality assumption is met; if the normality assumption is not met, these estimates are approximations.

Moderator Analyses

Moderator analyses were conducted to determine if excess variability could be accounted for by identifiable differences between studies and outcomes. Study characteristics examined in these analyses included: grade level (K–2 vs. 3–6) and outcome type (concepts about print, phonological awareness, phonics, vocabulary, fluency, and reading comprehension). Moderator analyses tested the combined effects of study characteristics. To determine if different grade levels may be a source of variation, we divided the study outcomes into those relating to grades K to 2 and those relating to grades 3 to 6. To determine if different outcome types may be a source of variation, we coded each effect size for the outcome domain concepts about print, phonological awareness, phonics, vocabulary, fluency, and reading comprehension. While we originally intended to include a broader range of moderators, most of the planned moderators were ultimately removed because either there was missing data (e.g., studies rarely reported the urbanicity of the sample) or a lack of variation in other potential moderators to the degree that the item could not be included (e.g., all but one study was randomized, so including research design as a moderator wasn't feasible). Therefore, moderator analyses included only outcome type and grade level.

Publication Bias

Prominent sources of bias include publication bias and selection bias. These are a type of systematic error that occurs when the likelihood of publishing a study or finding is contingent on it producing a desirable outcome (i.e., significant results in the predicted direction) (Myers et al., 2021). We were particularly careful to search for unpublished as well as published studies because of the known effects of publication bias in research reviews (Cheung & Slavin, 2016; Chow & Ekholm, 2018; Polanin et al., 2016). Ultimately, there were no unpublished studies included in the analyses. In our case, we followed two approaches to assess the degree to which publication or selection bias was present in the included sample—that is, a funnel plot for visual inspection and selection modeling for statistical inspection (Vevea & Hedges, 1995), although both assessment approaches must be carefully interpreted.

Results

A total of 11 studies and 51 effect sizes involving 8,839 K–6 students were found that examined Spanish reading programs’ effectiveness. Table 2 presents the characteristics of the individual studies selected by study level and outcome level (i.e., design type, grade, nationality, and dependent variable type). The study outcomes are summarized in Figure 2.

The main characteristics and findings of individual studies are summarized in Table 3.

Reading Instruction

Grades K–2

Seven studies were based on the initial phases of reading acquisition (5-year-old K [5K], grade 1, and grade 2), and addressed phonological awareness, knowledge of letters and sounds, and reading of words and pseudowords (Gutiérrez-Fresneda, 2017 [4K and 5K], Gutiérrez-Fresneda et al., 2017 [5K], Gutiérrez-Fresneda, 2018 [grade 1], Gutiérrez-Fresneda et al., 2021 [grade 2], Muñoz et al., 2018 [5K], Pallante & Kim, 2013 [K–1], and Rolla et al., 2006 [K]). All studies were implemented with children who had typical reading development, using RCT designs, except for Rolla et al. (2006) with a QED. Pallante and Kim (2013) implemented a Spanish adaptation of the original American English intervention *Collaborative Language and Literacy Instruction Project (CLLIP)*. This program emphasizes the five components identified in the NRP (2000), with a preventive approach; the educators received professional development. The rest of the implemented programs were designed by the researchers. Three of them used shared reading as a procedure (Gutiérrez-Fresneda, 2017, 2018; Gutiérrez-Fresneda et al., 2017) and dialogic reading practices. The program by Muñoz et al. (2018) aimed to improve the phonological awareness of low-income Chilean preschool children and included a professional development course designed for teachers. Rolla et al. (2006) showed that a combination of three early literacy interventions (tutoring, classroom activities, and working with families) had an impact on the emerging literacy skills of low-income Costa Rican children. The program implemented by Gutiérrez-Fresneda et al. (2021) included 40 sessions of

TABLE 2.
Description of Included Studies

Category	Level	Overall
Study level		
Research design	RCT	10 (90.91%)
	QED	1 (9.09%)
Grade	3–6	4 (36.36%)
	K–2	7 (63.64%)
Nationality	Spain	7 (63.64%)
	Chile	2 (18.18%)
	Argentina	1 (9.09%)
	Costa Rica	1 (9.09%)
Total studies		11 (100%)
Outcome level		
Outcome type	Concepts of print	3 (5.88%)
	Phonological awareness	13 (25.49%)
	Phonics	13 (25.49%)
	Vocabulary	3 (5.88%)
	Fluency	2 (3.92%)
	Reading comprehension	17 (33.34%)
	Total effect sizes	51 (100%)

45 minutes’ duration, each including phonological awareness, decoding, reading comprehension, and prosodic activities. Specifically, the first part of the sessions focused on segmental phonology (lexical segmentation tasks, syllabic awareness, and phonemic awareness).

Research results in this age group (ES=0.51) illustrate that these interventions can successfully increase foundational reading processes, including phonological awareness, knowledge of letters and sounds, and reading of words and pseudo-words. The magnitude of these effects corresponds to a medium-high effect size following Cohen’s (1988) benchmarks or above the 90th percentile in reading following Kraft’s (2020).

Grades 3–6

Four studies were based on the later stages of reading consolidation (grades 3–6) and addressed reading comprehension (Carriedo & Alonso-Tapia, 1996 [grade 6], Flores & Duran, 2016a [grades 3–6], Fonseca et al., 2019 [grade 4], and Núñez et al., 2022 [grades 3–4]). All of them were applied to children with typical reading development, with RCT designs. The programs were designed by the

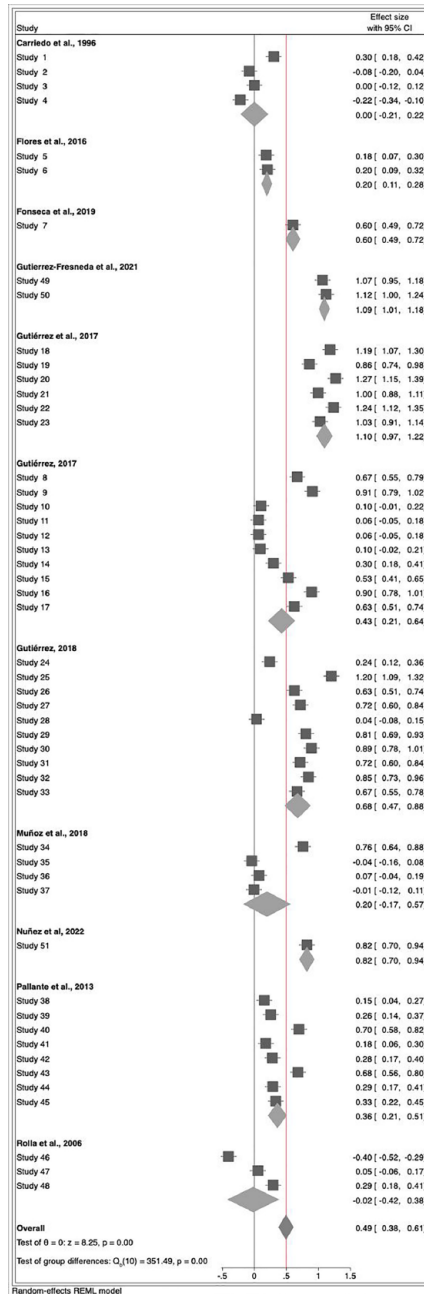


FIGURE 2. Forest plot illustrating outcomes of included studies.

TABLE 3.

Main characteristics and outcomes of the individual studies

Study	Country	Study design	Sample size	Grade	Duration	Experimental group instruction	Control group instruction type*	Effect size by outcome				
								Concepts of print	Phonological awareness	Phonics	Vocabulary	Fluency
Carriedo & Alonso-Tapia (1996)	Spain	RCT	T, 138 C, 73	6	12 weeks	Researchers-made program: Main idea identification, previous knowledge activation, text structure knowledge	Traditional instruction				+0.301	-0.081 0.0 -0.218
Flores & Duran (2016a)	Spain	RCT	T, 441 C, 136	3-6	24 half-hour sessions for 12 weeks	Leemos en pareja [We read as a Couple] program: Self-regulated learning and reading comprehension strategies before, during, and after reading	Regular curriculum					+0.185 +0.205
Fonseca et al. (2019)	Argentina	RCT	T, 80 C, 47	4	16 80-minute sessions	LEE comprensivamente [READ comprehensively] program: Vocabulary and comprehension strategies (e.g., inferences, self-regulation, and textual structure)	Basal program					+0.603
Gutiérrez-Fresneda et al. (2017)	Spain	RCT	T, 220 C, 212	5K	60 45-minutes sessions for 7 months	Researchers-made program: Oral language, phonological awareness, rapid automatized naming, and alphabetic knowledge	Regular curriculum		+1.185 +0.861	+1.271 +0.997		+1.237 +1.027
Gutiérrez-Fresneda (2017)	Spain	RCT	T, 206 C, 196	4-5K	65 50-minute sessions for 7 months	Researchers-made program: Phonological skills, rapid automatized naming, and meta-comprehension strategies	Basal program	+0.064	+0.67 +0.906	+0.105 +0.095 +0.297 +0.532		+0.064 +0.896 +0.626
Gutiérrez-Fresneda (2018)	Spain	RCT	T, 206 C, 202	1	60 45-minute sessions, 5 sessions per week	Researchers-made program: Phonological awareness, rapid automatized naming and alphabetic knowledge and reading comprehension strategies	Basal program	+0.035	+0.239 +1.204	+0.627 +0.808 +0.894 +0.718		+0.721 +0.846 +0.666
Gutiérrez-Fresneda et al. (2021)	Spain	RCT	T, 220 C, 218	2	40 45-minute sessions	Researchers-made program: Phonological awareness, decoding, reading comprehension, and prosodic skills	Basal program		+1.066 +1.122			

(continued)

Table 3. (continued)

Study	Country	Study design	Sample size	Grade	Duration	Experimental group instruction	Control group instruction type*	Effect size by outcome					
								Concepts of print	Phonological awareness	Phonics	Vocabulary	Fluency	Reading comprehension
Muñoz et al. (2018)	Chile	RCT	T, 81 C, 81	5K	From July 2014 to December 2014	Researchers-made program: Phonological awareness	Regular curriculum		+0.76 −0.041 +0.073 −0.006				
Núñez et al. (2022)	Spain	RCT	T, 403 C, 355	3–6	12 weekly 50-minute sessions	Arco Iris [The Rainbow] program: Self-regulated learning (e.g., work planning, time management, goal setting) and reading comprehension strategies (e.g., main ideas summary)	Regular curriculum						+0.821
Pallante and Kim (2013)	Chile	RCT	T, 80 C, 83 T, 155 C, 150 T, 114 C, 35	K–1	From March to September	Adaptation of the original American English intervention Collaborative Language and Literacy Instruction Project (CLLIP): Phonological awareness, alphabets and phonics, fluency, vocabulary, reading comprehension and writing	Regular curriculum		+0.255	+0.178 +0.682	+0.155 +0.283	+0.698 +0.288	+0.334
Rolla et al. (2006)	Costa Rica	QED	T, 41 C, 55	K	From July to December	Researchers-made program: Phonological awareness and letter-sound relationship	Regular curriculum	−0.404		+0.054			+0.294

Note. *The NRP (2000) used the following categories to classify the type of program taught to the control group: (a) regular curriculum; (b) whole language approach (e.g., Big Books); (c) whole word approach; (d) miscellaneous (e.g., child reads with tutor, study skills); (e) basal programs, which consist of a teacher's manual and a complete set of books and materials that guide the teaching of beginning reading and whose defining characteristic is that they do not provide explicit, systematic phonics instruction (pp. 2–121).

researchers, except for that of Fonseca et al. (2019), who applied the *LEE comprensivamente* (*READ comprehensively*) program by Gottheil et al. (2011), adding vocabulary training to the three other comprehension strategies (inferences, self-regulation, and knowledge of the textual structure) that the program trains. Flores and Duran (2016a) implemented the program *Leemos en pareja* [*We read as a Couple*], whose main objective is the development of reading comprehension through peer tutoring, according to the previously assigned role (tutor/tutor/recipient) depending on their level of reading comprehension. The program of Carriedo and Alonso-Tapia (1996) was mainly dedicated to working on the main idea of texts with children in grade 6. The study by Núñez et al. (2022) used the *Rainbow Program* whereby Spanish students employed self-regulated learning macro-strategies (e.g., work planning, time management, goals setting) and reading comprehension strategies (e.g., main ideas summary).

Research results in this age group focused on the ultimate reading goal, reading comprehension. The average effect size obtained in these studies ($ES=0.34$) is medium following Cohen's (1988) benchmark or between the 80th and 90th percentile in reading following Kraft's (2020). These results demonstrate that comprehension strategies such as inferences, detection of the main idea, and knowledge of text structure can lead to higher reading outcomes.

Concepts of Print Studies

Three of the studies included studies that examined outcomes of concepts of print, specifically, Gutiérrez-Fresneda (2017, 2018) and Rolla et al. (2006). In Gutiérrez-Fresneda (2017), 206 students were in the intervention group and 196 in the control group. The reading program was made up of 65 sessions of 50 minutes. Its objective was to check whether shared reading practices translated into higher decoding skills and a better understanding of reading. Gutiérrez-Fresneda (2018) provided a similar reading learning program (60 sessions of 45 minutes) to 206 students in the intervention group and 202 in the control group (reading teaching according to the textbook). Rolla et al. (2006) used three types of interventions: family (e.g., structured activities around oral and written language at home), tutors (e.g., reading stories), and classroom (e.g., a combination of reading and reciting well-known Costa Rican children's poetry and activities from different phonological awareness curricula) to check their impact on early literacy skills. The program included 18 sessions of 45 minutes. The intervention group included 41 children, and 55 students were assigned to the control group.

Impacts on concepts of print outcomes were negative and nonsignificant in these three studies. Future research is needed to provide evidence about this instructional practice on reading outcomes.

Phonological Awareness Studies

Six studies included phonological awareness activities using either a commercially available program (Pallante & Kim, 2013) or a researcher-designed program (Gutiérrez-Fresneda, 2017, 2018; Gutiérrez-Fresneda et al., 2017, 2021; Muñoz et al., 2018). Pallante and Kim (2013) implemented CLLIP, aimed at phonological awareness, alphabetics and phonics, fluency, vocabulary, reading comprehension, and writing. In this study, seven kindergarten and five first-grade

classrooms were in the CLLIP condition ($n=349$), five kindergarten and five first-grade classrooms were in the control condition ($n=268$). CLLIP teachers received professional development over five scheduled workshops distributed throughout the year about phonological processing, vocabulary, reading fluency, reading comprehension, and writing along with strategic orientation to collaboration and training in assessment and walkthrough demonstrations. Muñoz et al. (2018) implemented an intervention consisting of a professional development approach based on phonological awareness instruction. Participants were selected from two schools: one school was assigned to the control group ($n=81$) and the equivalent school was an intervention group ($n=81$) with a duration of 15 minutes daily and two sessions per week. Phonological awareness processing in Gutiérrez-Fresneda (2017, 2018), and Gutiérrez-Fresneda et al. (2017, 2021) was addressed during tasks of lexical, syllabic, and phonemic awareness using story content.

Phonological awareness studies show one of the highest effect sizes (0.57), which corresponds to a medium-high effect size following Cohen's (1988) benchmarks or above the 90th percentile in reading following Kraft's (2020). Similar results were found by the NRP (2000) with a weighted effect size average of 0.53.

Phonics Studies

A total of five studies targeted phonics and/or reading words. One of them used a commercially available program (Pallante & Kim, 2013) and four used a researcher-designed program (Gutiérrez-Fresneda, 2017, 2018; Gutiérrez-Fresneda et al., 2017; Rolla et al., 2006). Pallante and Kim (2013) used the CLLIP model targeting phonological awareness; alphabetics; and phonics, fluency, vocabulary, reading comprehension, and writing. The alphabetic knowledge in Gutiérrez-Fresneda (2017, 2018) and Gutiérrez-Fresneda et al. (2017) was practiced by making representations with the sounds and the words known at a multi-sensorial level and teaching letter names using phonetic-based mixed methods. In Rolla et al. (2006), the activities included were a combination of reading words and activities from phonological awareness published materials.

Phonics studies showed an effect size of 0.45, which corresponds to a medium-high effect size following Cohen's (1988) benchmarks or between the 80th and 90th percentile in reading following Kraft's (2020). Similar results were found by the NRP (2000) with a weighted effect size average of 0.44. These results reflect the impact of working on the letter-sound relationship on growth in reading.

Vocabulary studies

Two studies addressed vocabulary using either a commercially available program (Pallante & Kim, 2013) or a researcher-designed program (Gutiérrez-Fresneda et al., 2017). Pallante and Kim (2013) used the CLLIP model, which targeted vocabulary, among other skills. Gutiérrez-Fresneda et al. (2017) provided 60 sessions of 45 minutes of reading instruction focused on the development of those words that are currently considered the main precursors of learning to read. The semantic development aimed at enhancing the lexical scope was exercised through recognition tasks of elements in pictures, photographs and drawings; elaboration of lists of objects by semantic fields; identification of intrusive words in sentences; and searches of synonyms and antonyms. The experimental group

consisted of 220 students, whereas the control group had 212 students. Carriedo and Alonso-Tapia (1996) assessed the impact of training main idea comprehension on vocabulary, among other skills such as reading comprehension. Here, a researcher-designed curriculum was used. This study involved 138 students in the intervention group and 75 in the control group. In this study, teachers received a 30-hour course on how to teach reading comprehension strategies (mainly those related to text structure) in the classroom and 15 practice sessions.

Vocabulary studies showed an effect size of 0.52, which represents a medium-high effect size following Cohen's (1988) benchmarks or above the 90th percentile in reading following Kraft's (2020). These results stem from only two studies so must be interpreted with caution.

Fluency Studies

Only one study targeted oral word-reading fluency using a commercially available program (Pallante & Kim, 2013). The authors used the CLLIP model with seven kindergarten and five first-grade classrooms in the CLLIP condition and five kindergarten and five first-grade classrooms in the control condition. Due to Spanish language transparency, children reach a ceiling in word reading accuracy at the end of first grade, although improvements in speed take longer. CLLIP teachers received professional development over five scheduled workshops spread throughout the year.

Results showed an effect size of 0.66, which represents a high effect size following Cohen's (1988) benchmarks or above the 90th percentile in reading following Kraft's (2020). A bit smaller result was found by the NRP (2000), with a weighted effect size average of 0.41. Results also point to some reading fluency difficulties in Spanish students, which means in both cases, that more studies are needed on this core element to be able to draw reliable conclusions or implications for practice.

Reading Comprehension Studies

Eight studies targeted reading comprehension using either a commercially available program (Flores & Duran, 2016a; Fonseca et al., 2019; Pallante & Kim, 2013) or a researcher-designed program (Carriedo & Alonso-Tapia, 1996; Gutiérrez-Fresneda, 2017, 2018; Gutiérrez-Fresneda et al., 2017; Núñez et al., 2022). Flores and Duran (2016a) provided 24 half-hour sessions over 12 weeks of a peer-tutoring program called *Leemos en pareja [We Read as a Couple]* to 3–6 grade students. A total of 441 students formed the intervention group and 136 the comparison group. In Fonseca et al. (2019), 127 4th-grade students were randomly assigned to the *LEE comprensivamente [READ Comprehensively]* reading program ($n=80$) or to a comparison condition ($n=47$). Students received 16 sessions of 80 minutes each. Pallante and Kim (2013) used the CLLIP model, which focused on phonological awareness, alphabetics and phonics, fluency, and vocabulary, along with reading comprehension and writing.

Carriedo and Alonso-Tapia (1996) developed a teacher training program focused on how to teach reading comprehension strategies in the classroom. This teacher training program lasted 50 hours. After that, teachers applied the program

to grade 6 students. In all, 138 students formed the intervention group and 73 the comparison group.

In Gutiérrez-Fresneda et al. (2017), reading comprehension skills were trained through dialogic reading, and comprehension skills were trained through the implementation of the reading strategies using previous knowledge and by promoting the skills that enhance control and regulation during the comprehension process. These strategies were sequenced in three specific moments: before, during, and after reading. Additionally, Gutiérrez-Fresneda (2018) assessed the impact of a researcher-designed reading program on reading comprehension. The most recent study, Núñez et al. (2022), used the *Rainbow Program* in which Spanish students were trained on self-regulated learning macro-strategies (e.g., work planning, time management, goals setting) and reading comprehension strategies (e.g., self-questioning, main ideas summary in one's own words) during 12 weekly 50-minute sessions.

Reading comprehension studies had an effect size of 0.52, which represents a medium-high effect size following Cohen's (1988) benchmarks or above the 90th percentile in reading following Kraft's (2020).

Studies' Quality

In relation to the studies' quality, we utilized the WWC determinants of study quality rating for RCTs and QEDs (What Works Clearinghouse [WWC], 2020); 51 comparisons from 11 studies sampled qualified as *Meets WWC Group Design Standards* and met pretest reading baseline equivalence with statistical adjustments. Three comparisons from one study (Rolla et al., 2006) also qualified as *Meets WWC Group Design Standards with Reservations* based on not randomizing students to condition (researchers decided who were in the control or experimental group) and met pretest reading baseline equivalence with statistical adjustments. All studies included the analytical sample size after attrition and baseline equivalence on pretest reading outcomes within $g = \pm 0.25$ standard deviation.

Meta-regression

For research question 1 (i.e., what is the average treatment effect across included studies of Spanish literacy instructional programs?), a meta-regression analysis was conducted (see Table 4). For all studies, this model controlled for grade level and outcome type. Across the studies included, we obtained a positive, medium, and statistically significant effect (effect size = +0.49, $p < 0.05$) with a 95% confidence interval of -0.15 to -0.88, suggesting that the reading instructional practices and/or products were generally effective.

One way to quantify heterogeneity is with I^2 , which estimates how much of the variance is due to heterogeneity rather than sampling error. This value of the included study is quite large, with 71% of the variance due to heterogeneity. This can be further broken down into between- and within-cluster heterogeneity. Approximately 42% of the total variance is estimated to be due to between-cluster heterogeneity, 29% due to within-cluster heterogeneity, and the remaining 29% due to sampling variance. Another way to quantify heterogeneity is with the

TABLE 4.
Meta-Regression Results

Coefficient	Reference group	β	SE	t	df	p	CI
Null model							
Intercept		0.46	0.12	3.83	9.74	0.003	0.19–0.72
Meta-regression							
Intercept		0.49	0.13	3.75	6.53	0.008	0.17–0.80
Grade level							
K–2	3–6	0.18	0.24	0.74	6.55	0.484	–0.40–0.75
Outcome type							
Concepts about print	Phonics	–0.53	0.16	–3.29	2.17	0.073	–1.17–0.11
Phonological awareness		0.12	0.18	0.66	3.36	0.554	–0.42–0.66
Vocabulary		0.07	0.29	0.24	1.29	0.840	–2.11–2.25
Fluency		0.20	0.13	1.59	2.50	0.227	–0.25–0.66
Reading comprehension		0.07	0.10	0.66	3.53	0.551	–0.23–0.36

Note. β =Standardized coefficient, SE=Standard Error, t = t test, df=degrees of freedom, p = p -value, CI=Confidence interval.

prediction interval. There was substantial heterogeneity across this sample, with a 95% prediction interval of –0.22 to 1.19. The 95% prediction interval gives the range in which the point estimate of 95% of future studies will fall, assuming that true effect sizes are normally distributed. Given the large range of the prediction interval, which implies that “true” effects for these types of studies may fall anywhere within that band and that the prediction interval crosses zero, this implies that some approaches are not effective and may even be associated with lower outcomes for students. Additionally, the high heterogeneity found suggested that an analysis to detect potential moderators was appropriate.

Publication Bias

We analyzed the presence of publication or selection bias using both a funnel plot and selection modeling. The funnel plot (Figure 3) may show some slight asymmetry, with fewer studies present on the lower left side of the funnel, which would be expected if there were indeed a bias toward publishing studies with larger impacts. According to Egger’s test, the significant results ($z=2.38, p<.05$) confirm evidence of an asymmetry. To further explore this, we estimated a selection model with cut points at $p=0.01, 0.05$, and 0.20 . While the adjusted model is not a significantly better fit than the unadjusted model (indicating no selection bias), the adjusted mean effect size is 0.36 , smaller than the mean effect size estimated previously. This suggests there may be a degree of publication bias in these data, but given the small sample size, this finding is inconclusive.

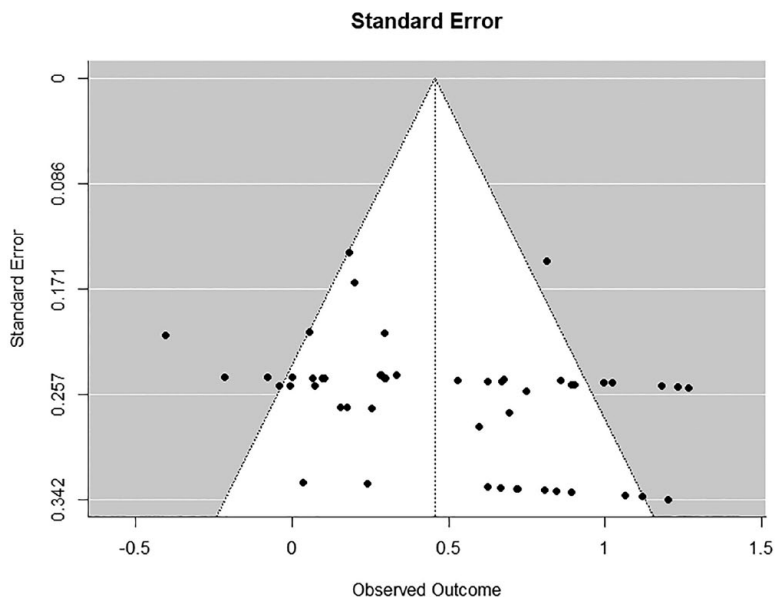


FIGURE 3. *Funnel plot to assess publication bias.*

Moderator analyses

For research question 2 (i.e., How much variability in effect sizes can be accounted for by studies' characteristics of grade level and outcome type?), we examined differences in effect sizes across outcome types and grade level as Table 5 shows. The mean effect size for each outcome was compared with each of the other outcome types, with none of those comparisons yielding significant differences, so we found no evidence that outcome type was a significant moderator of effects. The comparison by grade level after dividing the studies' outcomes into those relating to grades K to 2 and those relating to grades 3 to 6 also yielded no significant differences, so we found no evidence that grade level was a significant moderator of effects.

Discussion

A meta-analysis was conducted to evaluate the evidence of the effectiveness of specific Spanish reading instructional programs. Additionally, a meta-regression was conducted to test the statistical significance of potential moderators to understand better the potential variations in the impacts of these interventions.

The number of studies and effect sizes meeting high-quality methodological standards allows for a picture of the status of effective programs for Spanish reading instruction. Indeed, the rigorous inclusion and exclusion criteria adopted make findings both statistically reliable and relevant to practice and policy. The instructional programs identified in studies meeting rigorous statistical standards for strong and moderate levels of evidence indicate that educators have practical

TABLE 5.
Mean effect sizes by outcome type and grade level

Moderator	ES	SE	<i>t</i>	df	<i>p</i>	CI
Outcome type						
Concepts of print	−0.08	0.14	−0.54	3.10	0.628	−0.52–0.37
Phonological awareness	0.57	0.13	4.39	6.09	0.004	0.25–0.89
Phonics	0.45	0.18	2.54	5.70	0.046	0.01–0.89
Vocabulary	0.52	0.28	1.85	1.89	0.213	−0.76–1.81
Fluency	0.66	0.15	4.50	4.45	0.008	0.27–1.04
Reading comprehension	0.52	0.14	3.75	5.97	0.010	0.18–0.86
Grade level						
3–6	0.34	0.18	1.84	3.39	0.153	−0.21–0.88
K–2	0.51	0.15	3.40	5.80	0.015	0.14–0.89

Note. ES=Effect size, SE=standard error, *t*=*t* test, df=degrees of freedom, *p*=*p*-value, CI=confidence interval.

solutions available for the problems of reading failure in primary education schools. In this vein, the overall effect of reading instructional programs in Spanish was statistically significant at *g*=0.49. This effect was smaller than the *g*=0.71 effect size reported by Ripoll and Aguado (2014). Differences in effect size across both studies may be due to variations in the scope and inclusion criteria. Ripoll and Aguado (2014) included, for example, pre-experimental designs, groups with less than 30 students per group, groups with sample sizes smaller than 30 students per group, or studies using research-made instruments to measure outcomes, which may have resulted in a mean effect size larger than we presented, as other studies like Cheung and Slavin (2016) warn.

These findings suggest that the instructional programs examined in this review are useful for improving the reading achievement of K–6 students. This also means that teachers can leverage a relatively wide array of existing K–2 reading acquisition and 3–6 consolidation instructional programs to provide core instruction that increases students’ access to general reading standards, which may, in turn, enhance their success on other critical school outcomes.

In relation to variability in effect size being affected by grade level and outcome type, in contrast to Ripoll and Aguado’s (2014) results, we did not detect a statistically significant moderator effect. This could be attributed to the limited sample of studies utilized, as the *df* in the meta-regression analyses reflect, and/or the fact that they included less stringent inclusion criteria. Thus, given the underpowered nature of this analysis, as well as Ripoll and Aguado’s (2014), these results should be interpreted with caution. Nevertheless, it is worthy to point out the large effects found on phonological awareness, phonics, fluency and reading comprehension (all of them significant), and vocabulary (marginally significant), as well as the statistically significant levels of evidence on K–2.

Generalization of Results

The reliable evidence on effective reading programs in Spanish for Spanish speakers reported in this systematic review and meta-analysis suggests that guidance about high-quality, evidence-based practices for reading instruction in Spanish could be made available to teachers in multi-component instruction (Gutiérrez-Fresneda, 2018), phonological awareness (Gutiérrez-Fresneda et al., 2021; Pallante & Kim, 2013), and reading comprehension (Gutiérrez-Fresneda, 2018; Pallante & Kim, 2013; Rolla et al., 2006). For the rest of the core reading skills across K–6, we have found marginal evidence, probably attributable to the low number of studies qualifying for rigorous synthesis. This reveals the promising results for the rest of the core reading skills, as well as the need for more evaluation and synthesis research studies using higher-quality research and evaluation designs like those listed in our inclusion criteria. If more teachers and schools adopt an evidence-based educational framework focusing on proven programs and prevail for long enough, as Slavin (2020) recommended, our schools will become more reliable places to deliver the educational promise for all children regarding literacy and, therefore, curricular learning and development.

The transferability of this review's findings in Spanish to other languages with more consistent orthographies and less syllabic complexity and vice versa has been largely debated (Galuschka et al., 2014), and it is not the goal of this study. However, based on the robustness of the inclusion and exclusion criteria adopted by this review and recommended by international standards (WWC, 2020), along with the results of three decades of cross-linguistic research (August et al., 2009; Nakamoto et al., 2008; National Academies of Sciences, Engineering, and Medicine, 2017) showing that language minority students' literacy development parallels with monolingual literacy development, our results may represent a valid and reliable contribution to the arsenal of Spanish reading instructional programs applicable to societies with significant Spanish-speaking populations who are taught to read in Spanish. The potential applicability of these results could also be extended to English language learners (ELL) if their reading difficulties in English were due to their limited proficiency in the English language and not to a learning disability (Klingner et al., 2006).

Neither the NRP (2000) nor this systematic review included students with disabilities because prevention and treatment evaluation standards adopted in both reviews have not been universally accepted or used in reading education research.

Our review also intends to contribute to the ongoing discussion of the importance, need, and impact of developing evidence-based educational legislation capable of discontinuing the cascade of ineffective educational reforms in the last three decades in Spain (Arco-Tirado et al., 2021) and worldwide, to remedy the lack of progress on functional literacy skills (Orellana, 2018). However, as Carroll et al. (2007) point out, advancing in the continuum of evidence identification, dissemination, and adoption, involves training and follow-up measures to assure implementation fidelity, which refers to the degree to which a practice or program is delivered as intended, so that researchers and practitioners gain a better understanding of how and why an intervention works, and the extent to which outcomes can be improved.

The potential economic importance of these results lies in the fact that there is a positive association between education and long-term economic growth (Hanushek & Woessmann, 2008). In this vein, an important motivation for considering the value of basic skills in literacy at school is that it improves employment prospects, productivity, and higher wages as a result (Carneiro & Heckman, 2004). Thus, if the skill level of a country's workforce is correlated with its growth in gross domestic product per person, then the way policymakers go about improving literacy is crucially important (Vignoles, 2016). Therefore, the availability of credible research results from rigorous meta-analyses on effective instructional programs plays a key role in the design, implementation, and evaluation of educational policies and curricula. Furthermore, in terms of academic and scholarly publishing, considering that there are 493 million people having Spanish as their first language (Instituto Cervantes, 2021), it is not difficult to imagine the potential economic benefits of having evidence-based effective Spanish reading instructional programs readily published. For example, if we look at the economic impact of the English language teaching industry in the United Kingdom, teaching English to international students adds £1.1 billion of value to the economy, supporting around 26,500 jobs and generating £194 million in net tax revenues for the government (Chaloner et al., 2015). Furthermore, evidence-based policies in the United Kingdom have helped spur the growth of the creative industries sector, which now contributes £111 billion to the United Kingdom economy (British Council, 2020).

Limitations

In terms of limitations, we found only 11 studies that met our rigorous criteria. Although our sample is relatively large in terms of outcomes analyzed, it is small in terms of statistical analyses that combine empirical studies (Turner et al., 2013). A small sample reduces statistical power for performing moderator analyses, and consequently, the capacity to obtain more precise estimates of the effect size via moderators (Hedges & Pigott, 2004).

Our analysis was also limited in the number of moderators that could be examined. For example, prior research has shown that sample size may be related to effect sizes (Cheung & Slavin, 2016). We believed that by implementing an exclusion criterion to remove the extremely small studies as well as weighting studies by inverse variance, we have adequately addressed this possible concern. However, we also ran a sensitivity analysis where we included sample size as a categorical moderator (larger studies are those with at least 250 students, small studies are those with fewer than 250 students, as described by Cheung and Slavin, 2016)) and found that sample size was not a significant moderator of effect size, and when exploring means of each category, found that larger studies had larger effect sizes than smaller studies. This is an unexpected result, but sample size may be confounded with other factors, such as grade level or type of outcome. This is just one example that highlights that the small number of studies that met our inclusion criteria and the lack of variation across those studies in many factors was an important limitation regarding the degree to which heterogeneity of the studies could be explored statistically.

Directions for Future Research

Future research must focus on study quality. Our full-text article screening excluded a total of 500 studies; 76.2% of those reasons were methodological. This data speaks about the need to seriously reflect on the low quality of the methodological standards applied to educational research and evaluation of instructional programs in this strategic field.

If we compare this review on reading instructional programs in Spanish to the review reported by the NRP in the year 2000, we find that the NRP review was produced by 14 outstanding scholars (out of a list of 300 nominees offered by educational organizations) examining 52 studies on the teaching of phonemic awareness, 38 studies focusing on phonics instruction, 51 studies of oral-reading fluency, 45 on the teaching of vocabulary, and 205 studies of reading comprehension instruction for two years, with more than 400 teachers participating in the public hearings. Our review was conducted about 20 years later and includes just a total of 11 studies. In other words, researchers must conduct more rigorously designed experimental studies examining reading instructional programs for K–6 students to expand the literature available for further meta-analyses. Future investigations should explore the characteristics of those instructional programs (e.g., duration, training intensity, outcome type, grade level) to advance knowledge of these variables' impact on effectiveness. Furthermore, due to the substantial heterogeneity in the effect size estimates found; future research should continue to investigate additional potential moderators that affect the efficacy of reading instructional programs as well.

Although many sources of information can be used as the starting point for reading improvement, the quality of the evidence found here gives special legitimacy to our recommendation of translating these results into brief evidence summary documents, educational policies, and designs for curriculum materials. The need to provide advice and support for teachers about how to use the findings of this review represents another research and implementation gap we need to bridge, as part of a larger endeavor of shaping reading education.

Conclusions

The present study found a large, positive overall effect and significant impacts on phonological awareness, phonics, fluency, and reading comprehension in Spanish-speaking K–2, which suggests that if implemented with fidelity, the number of Spanish readers mastering core reading elements in Spanish can be increased using effective instructional programs.

This meta-analysis provides encouraging findings, suggesting that some instructional programs for Spanish reading exist that can effectively help readers in K–6 education. The need for more rigorous interventions and evaluation research in reading instructional programs should be a priority for all Spanish-speaking countries.

Finally, it is urgent to promote the reading and educational success of children by placing evidence-based practices at the center of educational practice and educational policymaking.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research leading to these results received funding from the Spanish Ministry of Science, Innovation and Universities through type A Program of mobility stays in higher education and research centers for senior researchers “Salvador de Madariaga” under Grant Agreement No[PRX19/00246].

ORCID iDs

José L. Arco-Tirado  <https://orcid.org/0000-0002-8390-9724>
Francisco D. Fernández-Martín  <https://orcid.org/0000-0003-1272-1131>
Mirian Hervás-Torres  <https://orcid.org/0000-0003-0875-6630>
Gracia Jiménez-Fernández  <https://orcid.org/0000-0001-9881-7200>
Nuria Calet  <https://orcid.org/0000-0002-1700-0387>
Sylvia Defior  <https://orcid.org/0000-0002-2786-3485>
Amanda J. Neitzel  <https://orcid.org/0000-0002-4676-9320>

References

Asterisks (*) indicate studies included in this review.

- Arco-Tirado, J. L., Fernández-Martín, F. D., Hervás-Torres, M., Jiménez-Fernández, G., Calet, N., Defior, S. A., Neitzel, A. J., & Slavin, R. (2023, March 21). *Spanish reading review protocol*. OSF. https://osf.io/egsxd/?view_only=95819df08d924356a4bfddc143445367
- Arco-Tirado, J. L., Fernández-Martín, F. D., Hervás-Torres, M., Jiménez-Fernández, G., Calet, N., Defior, S., Neitzel, A. J., & Slavin, R. E. (2024). *Data and code associated with the publication: A best-evidence synthesis and meta-analysis on effective reading programs in Spanish*. Johns Hopkins Research Data Repository, V1. [https://urldefense.com/v3/_https://doi.org/10.7281/T1/4J6ODZ_!!D9dNQwwGXtA!V aQwUsILRrLZb4nYd1Q1nLTZhD3vnfpZ-3XX74m4vJlXr_gAExne6J-hI7Gv-3d8UzumhmOVAIQFrU-4W\\$](https://urldefense.com/v3/_https://doi.org/10.7281/T1/4J6ODZ_!!D9dNQwwGXtA!V aQwUsILRrLZb4nYd1Q1nLTZhD3vnfpZ-3XX74m4vJlXr_gAExne6J-hI7Gv-3d8UzumhmOVAIQFrU-4W$)
- Arco-Tirado, J. L., Fernández-Martín, F. D., & Jagannathan, R. (2021). No jobs, no hope. The future of youth employment in Spain. In R. Jagannathan & M. Camasso (Eds.), *The growing challenge of youth unemployment in Europe and America: A cross-cultural perspective* (pp. 51–78). Bristol University Press.
- August, D., Calderón, M., & Carlo, M. (2002). *Transfer of skills from Spanish to English: A study of young learners*. Center for Applied Linguistics. <https://www.cal.org/acquiringliteracy/pdfs/skills-transfer.pdf>
- August, D., Shanaban, T., & Escamilla, K. (2009). English language learners: Developing literacy in second-language learners-report of the national literacy panel on language-minority children and youth. *Journal of Literacy Research*, 41, 432–452. <https://doi.org/10.1080/10862960903340165>
- Baker, D. L., Crespo, P., Monzalve, M., García, I., & Gutiérrez, M. (2022). Relation between the essential components of reading and reading comprehension in mono-lingual Spanish-speaking children: A meta-analysis. *Educational Psychology Review*, 34, 2661–2696. <https://doi.org/10.1007/s10648-022-09694-1>
- Balbi, A., Von Hagen, A., Cuadro, A., & Ruiz, C. (2018). Revisión sistemática sobre intervenciones en alfabetización temprana: Implicancias para intervenir en español.

- Revista Latinoamericana de Psicología*, 50(1), 31–48. <https://doi.org/10.14349/rlp.2018.v50.n1.4>
- Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2018). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly*, 54(2), 133–166. <https://doi.org/10.1002/rrq.229>
- Bizama, M., Arancibia, M. B., & Sáez, K. (2013). Intervención psicopedagógica temprana en conciencia fonológica como proceso metalingüístico a la base de la lectura en niños de 5 a 6 años socialmente vulnerables [Early psycho-pedagogical intervention in phonological awareness as a metalinguistic process based on Reading, in socially challenged children between 5-6 years old]. *Estudios Pedagógicos*, 39(2), 25–39. <https://doi.org/10.4067/S0718-07052013000200002>
- Blackford, K., Olmstead, G., & Stegman, Ch. (2012). Spanish as a second language for elementary students: A study of participation on literacy benchmark scores. *Journal of Modern Education Review*, 2(2), 77–89. <https://ssrn.com/abstract=2138193>
- Bloom, H. S. (2003). *Sample design for an evaluation of the reading first program* (MDRC Working Papers on Research Methodology). https://www.mdrc.org/sites/default/files/full_498.pdf
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5–18. <https://doi.org/10.1002/jrsm.1230>
- Bravo-Valdivieso, L. (1995). A four-year follow-up study of low socioeconomic status, Latin American children with reading difficulties. *International Journal of Disability, Development, and Education*, 42(3), 189–202. <https://doi.org/10.1080/0156655950420302>
- British Council. (2020). *British Council and design center to lead design economy mapping in the Philippines*. Author. <https://www.britishcouncil.ph/about/press/design-economy-mapping-philippines>
- The Campbell Collaboration. (2019). *Campbell systematic reviews: Policies and guidelines* (Version 1.4). Campbell Policies and Guidelines Series No. 1. <https://doi.org/10.4073/cpg.2016>
- Caravolas, M., Lervåg, A., Defior, S., Seidlová-Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, 24(8), 1398–1407. <https://doi.org/10.1177/0956797612473122>
- Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, 23(5), 386–402. <https://doi.org/10.1080/10888438.2019.1580284>
- Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, 23(6), 678–686. <https://doi.org/10.1177/0956797611434536>
- Carneiro, P., & Heckman, J. J. (2004). *Human capital policy* (Institute of Labor Economics Discussion Papers, No 821). Institute of Labor Economics. <https://EconPapers.repec.org/RePEc:iza:izadps:dp821>

- *Carriedo, N., & Alonso-Tapia, J. (1996). Main idea comprehension: Training teachers and effects on students. *Journal of Research in Reading*, 19(2), 411–431. <https://doi.org/10.1007/BF03172930>
- Carrillo, M. (1994). Development of phonological awareness and reading acquisition: A study in Spanish language. *Reading and Writing: An Interdisciplinary Journal*, 6(3), 279–298. <https://doi.org/10.1007/BF01027086>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40. <https://doi.org/10.1186/1748-5908-2-40>
- Case, L. P., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., Montanaro, E., & Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities*, 43(5), 402–417. <https://doi.org/10.1177/0022219409355475>
- Chall, J. (1996a). *Learning to read: The great debate* (revised, with a new foreword). McGraw-Hill.
- Chall, J. (1996b). *Stages of reading development* (2nd ed.). Harcourt-Brace.
- Chaloner, J., Evans, A., & Pragnell, M. (2015). *Supporting the British economy through teaching English as a foreign language. An assessment of the contribution of English language teaching to the United Kingdom economy*. Capital Economics Limited.
- Chávez-Delgado, M. E., González-Vergara, S., & Sepúlveda-López, F. (2022). Revisión sistemática de literatura sobre programas de intervención en habilidades de lectura inicial [Systematic review on literature on intervention programs on initial reading skills]. *Páginas de Educación*, 15(2), 98–127. <https://doi.org/10.22235/pe.v15i2.2775>
- Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research*, 82(4), 351–395. <https://doi.org/10.3102/0034654312465472>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, 30(3), 727–744. <https://doi.org/10.1007/s1064801894377>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Crespo, P., Jiménez, J. E., Rodríguez, C., Baker, D., & Park, Y. (2018). Differences in growth reading patterns for at-risk Spanish-monolingual children as a function of a tier 2 intervention. *The Spanish Journal of Psychology*, 21(E4), 1–16. <https://doi.org/10.1017/sjp.2018.3>
- Defior, S., & Tudela, P. (1994). Effect of phonological training on reading and writing acquisition. *Reading and Writing: An Interdisciplinary Journal*, 6, 299–320. <https://doi.org/10.1007/BF01027087>
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, 56(S1), S25–S44. <https://doi.org/10.1002/rrq.411>
- The Education Endowment Foundation. (2018). *Annual report 2018*. Author. <https://educationendowmentfoundation.org.uk/news/eef-publishes-2018-annual-report>

- European Agency for Special Needs and Inclusive Education. (2018). *Country policy reviews and analysis: Spain*. Author. <https://www.european-agency.org/sites/default/files/agency-projects/CPRA/Phase2/CPRA%20Spain.pdf>
- European Commission. (2017). *Youth Guarantee country by country. Spain*. Author. <http://ec.europa.eu/social/main.jsp?catId=1161&langId=en&intPageId=3353>
- European Commission. (2018). *Country Report Spain. 2018 European Semester: Assessment of progress on structural reforms, prevention and correction of macroeconomic imbalances, and results of in-depth reviews under Regulation (EU) No 1176/2011*. Author. <https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX:52018SC0207>
- European Commission. (2019). *Education and Training. Monitor 2019*. Publications Office of the European Union. <https://education.ec.europa.eu/sites/default/files/document-library-docs/volume-1-2019-education-and-training-monitor.pdf>
- Favila, A., & Seda, I. (2010). Phonological awareness in children with reading difficulties: Effects of an intervention. *Journal for the Study of Education and Development*, 33(3), 399–411. <https://doi.org/10.1174/021037010792215064>
- *Flores, M., & Duran, D. (2016a). Tutoría entre iguales y comprensión lectora: ¿un tándem eficaz? Los efectos de la tutoría entre iguales sobre la comprensión lectora [Peer tutoring and reading comprehension: An effective tandem? The effects of peer tutoring on reading comprehension]. *Universitas Psychologica*, 15(2), 339–352. <https://doi.org/10.11144/Javeriana.upsy15-2.teic>
- Flores, M., & Duran, D. (2016b). Influence of a catalan peer tutoring programme on reading comprehension and self-concept as a reader. *Journal of Research in Reading*, 39(3), 330–346. <https://doi.org/10.1111/1467-9817.12044>
- Fonseca, L., Migliardo, G., Simian, M., Olmos, R., & León, J. A. (2019). Estrategias para mejorar la comprensión lectora: impacto de un programa de intervención en español [Strategies to improve reading comprehension: Impact of an intervention program in Spanish]. *Psicología Educativa*, 25, 91–99. <https://doi.org/10.5093/psed2019a1>
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS One*, 9(2), e89900. <https://doi.org/10.1371/journal.pone.0089900>
- García, E., Jiménez, J. E., González, D., & Jiménez, E. (2013). Reading comprehension difficulties among students in primary education and compulsory secondary education: A study of prevalence in Spanish. *European Journal of Investigation in Health, Psychology and Education*, 3(2), 113–123. <https://doi.org/10.30552/ejihpe.v3i2.43>
- Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, 13(2), 401–427. <https://doi.org/10.1080/19345747.2019.1689591>
- Gottheil, B., Fonseca, L., Aldrey, A., Lagomarcino, I., Pujals, M., Puyrredón, D., & Molina, S. (2011). *Programa LEE comprensivamente. Guía Teórica* [READ comprehensively program. Theoretical guide]. Paidós.
- González, J. E. J., & Valle, I. H. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities*, 33(1), 44–60. <https://doi.org/10.1177/002221940003300108>
- *Gutiérrez-Fresneda, R. (2017). Efecto de la lectura compartida y las habilidades prelectoras en el aprendizaje lector [The effect of shared reading and pre-reading

- skills on reading learning]. *Ocnos*, 16(2), 17–26. https://doi.org/10.18239/ocnos_2017.16.2.1356
- *Gutiérrez-Fresneda, R. (2018). Habilidades favorecedoras del aprendizaje de la lectura en alumnos de 5 y 6 años [Facilitating skills of reading learning on students 5 and 6 years old]. *Revista Signos*, 51(96), 45–60. <https://doi.org/10.4067/S0718-09342018000100045>
- *Gutiérrez-Fresneda, R., de Vicente-Yagüe Jara, I. M., & Jiménez-Pérez, E. (2021). Efectos de la conciencia suprasegmental en el aprendizaje de la lectura en los primeros cursos escolares [Effects of suprasegmental awareness on learning to read in the first school years]. *Revista de Psicodidáctica*, 26(1), 28–34. <https://doi.org/10.1016/j.psicod.2020.10.001>
- *Gutiérrez-Fresneda, R., Díez, A., & Jiménez-Pérez, E. (2017). Estudio longitudinal sobre el aprendizaje lector en las primeras edades [Longitudinal study on reading learning at early ages]. *Revista de Educación*, 378, 30–51. <https://doi.org/10.4438/1988-592X-RE-2017-378-360>
- Hall, C., Dahl-Leonard, K., Cho, E., Solari, E. J., Capin, P., Conner, C. L., Henry, A. R., Cook, L., Hayes, L., Vargas, I., Richmond, C. L., & Kehoe, K. F. (2023). Forty years of reading intervention research for elementary students with or at risk for dyslexia: A systematic review and meta-analysis. *Reading Research Quarterly*, 58(2), 285–312. <https://doi.org/10.1002/rrq.477>
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668. <https://doi.org/10.1257/jel.46.3.607>
- Hedges, L. V. (1981). Distribution theory for Glass’s estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370. <https://doi.org/10.3102/1076998606298043>
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-Analysis. *Psychological Methods*, 9(4), 426–445. <https://doi.org/10.1037/1082-989X.9.4.426>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Instituto Cervantes [Cervantes Institute]. (2021). *El español en el mundo. Anuario del Instituto Cervantes 2021* [The Spanish in the world. Cervantes Institute Yearbook 2021]. Author. https://cvc.cervantes.es/lengua/anuario/anuario_21/
- Jiménez, J. E., & García, C. R. H. (1995). Effects of word linguistic properties on phonological awareness in Spanish children. *Journal of Educational Psychology*, 87(2), 193–201. <https://doi.org/10.1037/0022-0663.87.2.193>
- Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English Language Learners who struggle with reading: Language acquisition or learning disabilities? *Journal of Learning Disabilities*, 39(2), 108–128. <https://doi.org/10.1177/00222194060390020101>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, 95(3), 482–494. <https://doi.org/10.1037/0022-0663.95.3.482>

- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.
- Ministerio Educación, Cultura y Deporte [Ministry of Education, Culture and Sports]. (2016). *PISA 2015. Programa para la evaluación internacional de los alumnos. Informe español* [PISA 2015. Program for international student assessment. Spanish report]. Author. <https://www.educacionyfp.gob.es/dctm/inee/internacional/pisa-2015/pisa2015preliminarok.pdf?documentId=0901e72b8228b93c>
- Ministry of Foreign Affairs, European Union and Cooperation. (2018). *Progress report. Implementation of the 2030 Agenda in Spain*. Author. https://sustainabledevelopment.un.org/content/documents/25394Progress_Report_2019_Spain.pdf
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/pirls2011/international-results-pirls.html>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center, Boston College. <http://timssandpirls.bc.edu/pirls2016/international-results/>
- *Muñoz, K., Valenzuela, M. F., & Orellana, P. (2018). Phonological awareness instruction: A program training design for low-income children. *International Journal of Educational Research*, 89, 47–58. <https://doi.org/10.1016/j.ijer.2017.02.003>
- Myers, J. A., Brownell, M. T., Griffin, C. C., Hughes, E. M., Witzel, B. S., Gage, N. A., Peyton, D., Acosta, K., & Wang, J. (2021). Mathematics interventions for adolescents with mathematics difficulties: A meta-analysis. *Learning Disabilities*, 36(2), 145–166. <https://doi.org/10.1111/ldrp.12244>
- Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2008). A cross-linguistic investigation of English language learners' reading comprehension in English and Spanish. *Scientific Studies of Reading*, 12(4), 351–371. <https://doi.org/10.1080/10888430802378526>
- National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. The National Academies Press. <https://doi.org/10.17226/24677>
- National Reading Panel (NRP). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.
- *Núñez, J. C., Tuero, E., Fernández, E., Añón, F. J., Manalo, E., & Rosário, P. (2022). Efecto de una intervención en estrategias de autorregulación en el rendimiento académico en Primaria: estudio del efecto mediador de la actividad autorregulatoria [Effect of an intervention in self-regulation strategies on academic achievement in elementary school: A study of the mediating effect of the self-regulatory activity]. *Revista de Psicodidáctica*, 27(1), 9–20. <https://doi.org/10.1016/j.psicod.2021.09.001>
- Orellana, P. (2018). La enseñanza de la lectura en América Latina: desafíos para el aula y la formación docente [Teaching reading in Latin-America: Challenges for the classroom and teachers training]. *Revista Electrónica Leer, Escribir y Descubrir*, 1(3), 1–16. <https://digitalcommons.fiu.edu/led/vol1/iss3/2>
- Organization for Economic Co-operation and Development (OECD). (2013). *Teaching and learning international survey TALIS 2013*. International Association for the Evaluation of Educational Achievement, Statistics Canada, and OECD. https://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf
- Organization for Economic Co-operation and Development (OECD). (2017). *Education at a glance 2017. OECD indicators*. Author. <https://doi.org/10.1787/eag-2017-en>

- Organization for Economic Co-operation and Development (OECD). (2019). *OECD Skills Strategy 2019: Skills to shape a better future*. <https://doi.org/10.1787/9789264313835-en>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*, 89. <https://doi.org/10.1186/s13643-021-01626-4>
- *Pallante, D. H., & Kim, Y. S. (2013). The effect of a multicomponent literacy instruction model on literacy growth for kindergartners and first-grade students in Chile. *International Journal of Psychology, 48*(5), 747–761. <https://doi.org/10.1080/00207594.2012.719628>
- Pearson, P. D., Palincsar, A., Biancarosa, G., & Berman, A. (2020). *Reaping the rewards of the Reading for Understanding initiative*. National Academy of Education. <https://doi.org/10.31094/2020/2>
- Piggot, T., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*(1), 24–46. <https://doi.org/10.3102/0034654319877153>
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*(1), 207–236. <https://doi.org/10.3102/0034654315582067>
- Pustejovsky, J. (2020). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (Version R package version 0.4.1) [Software]. <https://CRAN.R-project.org/package=clubSandwich>
- R Core Team. (2020). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ripoll, J., & Aguado, G. (2014). La mejora de la comprensión lectora en español: un meta-análisis [Reading comprehension improvement for Spanish students: A meta-analysis]. *Revista de Psicodidáctica, 19*(1), 27–44. <https://doi.org/10.1387/RevPsicodidact.9001>
- Roberts, G. J., Cho, E., Garwood, J. D., Goble, G. H., Robertson, T., & Hodges, A. (2020). Reading interventions for students with reading and behavioral difficulties: A meta-analysis and evaluation of co-occurring difficulties. *Educational Psychology Review, 32*, 17–47. <https://doi.org/10.1007/s10648-019-09485-1>
- *Rolla, A., Arias, M., Villers, R., & Snow, C. (2006). Evaluating the impact of different early literacy interventions on low-income Costa Rican kindergarteners. *International Journal of Educational Research, 45*, 188–201. <https://doi.org/10.1016/j.ijer.2006.11.002>
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11. <https://doi.org/10.3102/0013189X015009005>
- Slavin, R. E. (2020, March 26). Science of reading: Can we get beyond our 30-year pillar fight? *Robert Slavin's Blog*. <https://robertslavinsblog.wordpress.com/2020/03/26/science-of-reading-can-we-get-beyond-our-30-year-pillar-fight/>
- Solari, E. J., & Gerber, M. M. (2008). Early comprehension instruction for Spanish-speaking English language learners: Teaching text-level reading skills while maintaining effects on word-level skills. *Learning Disabilities Research & Practice, 23*(4), 155–168. <https://doi.org/10.1111/j.1540-5826.2008.00273>

- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393. <https://doi.org/10.1037/met0000011>
- Turner, R. M., Bird, S. M., & Higgins, J. P. T. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. *PLoS One*, 8(3), e59202. <https://doi.org/10.1371/journal.pone.0059202>
- United Nations Educational, Scientific and Cultural Organization. (2019). *Meeting commitments. Are countries on track to achieve SDG 4?* United Nations Educational, Scientific and Cultural Organization, United Nations Educational, Scientific and Cultural Organization Institute for Statistics, and Global Education Monitoring Report. <https://unesdoc.unesco.org/ark:/48223/pf0000369009>
- Vaughn, S., Linna-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., Cardenas-Hagan, E., & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, 39(1), 56–73. <https://doi.org/10.1177/00222194060390010601>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435. <https://doi.org/10.1007/BF02294384>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Vignoles, A. (2016). *What is the economic value of literacy and numeracy?* IZA World of Labor. <https://doi.org/10.15185/izawol.229>
- Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, 51(6), 612–624. <https://doi.org/10.1177/0022219418775110>
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades k-3. *Educational Psychology Review*, 28(3), 551–576. <https://doi.org/10.1007/s10648-015-9321-7>
- What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook* (Version 4.1). Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf
- World Economic Forum. (2017). *The global human capital report. Preparing people for the future of work*. Author. <https://weforum.ent.box.com/s/dari4dktg4jt-2g9xo2o5pksjpatvawdb>

Authors

JOSÉ L. ARCO-TIRADO is a full professor at the University of Granada, Faculty of Education, Campus de Cartuja s/n. Granada, Spain; jlarco@ugr.es. His research focuses on evidence-based education and policy, systematic reviews and meta-analysis, public program evaluation, self-regulation, English as medium of instruction, learning disabilities, peer mentoring, and service learning.

FRANCISCO D. FERNÁNDEZ-MARTÍN is an associate professor of developmental and educational psychology at the University of Granada, Faculty of Education, Campus de Cartuja s/n. Granada, Spain; fdfernan@ugr.es. His research focuses on evidence-based education and policy, program evaluation, non-cognitive skills, civic-community

engagement and service learning, peer learning, mentoring and tutoring, inclusive education, and response to diversity.

MIRIAM HERVÁS-TORRES is a permanent professor of developmental and educational psychology at the University of Granada, Faculty of Education, Campus de Cartuja s/n. Granada, Spain; miriamhervas@ugr.es. Her research focuses on program evaluation, civic-community engagement and (e-)service learning, peer learning, (e-)mentoring and tutoring, inclusive education and response to diversity, emerging technologies, and digital competences.

GRACIA JIMÉNEZ-FERNÁNDEZ is an associate professor at the University of Granada (Department of Developmental and Educational Psychology, Faculty of Education) in Spain; gracijf@go.ugr.es. Her research interests include factors associated with literacy acquisition, evidence-based reading comprehension strategies, and learning disabilities such as dyslexia.

NURIA CALET is an associate professor at the University of Granada (Department of Developmental and Educational Psychology) in Spain, Faculty of Education, Campus de Cartuja s/n. Granada, Spain; ncalet@ugr.es. Her research focused on literacy acquisition, fluency reading, prosody skills, learning disabilities, oral language abilities, reading comprehension, and evidence-based interventions to improve literacy skills.

SYLVIA DEFIOR is a professor at the University of Granada (UGR; Department of Developmental and Educational Psychology) in Spain; sdefior@ugr.es. Defior is currently retired, although active in the field of research and its dissemination, trying to mitigate the gap between theory and practice and promote an education/intervention based on scientific evidence. The main interest in research is the study of the processes of acquisition and development of literacy skills, as well as learning disabilities, in particular dyslexia and dysgraphia, from a cognitive and crosslinguistic perspective and from a theoretical and applied approach.

AMANDA J. NEITZEL is an assistant professor and deputy director of evidence research at the Center for Research and Reform in Education at the School of Education, Johns Hopkins University, Baltimore, MD; e-mail: aneitzel@jhu.edu. Her research interests include research synthesis, evidence-based education, and research use.

ROBERT E. SLAVIN, a noted education researcher, the first-ever Distinguished Professor at the Johns Hopkins School of Education and director of the Johns Hopkins Center for Research and Reform in Education. Slavin was a preeminent researcher at the School of Education and a globally recognized figure in the field. His personal mantra was a continual emphasis on “evidence-based” research as the driver of school reforms across the country—a phrase he often simplified as “what works” in education. Slavin was among a handful of education experts known by name worldwide. He was sought out regularly to testify before Congress and to weigh in on education reform in the national media.