WHAT WORKS TO IMPROVE EARLY GRADE LITERACY IN LATIN AMERICA AND THE CARIBBEAN: 2016 UPDATE OF A SYSTEMATIC REVIEW

LAC READS CAPACITY PROGRAM

What Works to Improve Early Grade Literacy in Latin America and the Caribbean: 2016 Update of a Systematic Review

LAC Reads Capacity Program

Authors

Rebecca Stone

Thomas de Hoop

Andrea Coombes

Mariela Goett

Mitchell Morey

Kevin Kamto

Table of Contents

Introduction

This update to the document "Early Grade Reading in Latin America and The Caribbean: A Systematic Review", published in 2017, seeks to inform the work of the Latin America and Caribbean Reads Capacity Program (LRCP), which is led by the American Institutes for Research (AIR) in partnership with Juárez and Associates. The LRCP aims to increase the impact, scale, and sustainability of early grade literacy (EGL) interventions in the Latin America and Caribbean (LAC) region by developing state-of-the-art knowledge resources and providing technical assistance to governments and other selected key stakeholders in the LAC region. Such resources and technical assistance should enable and enhance the efforts of governments in the LAC region to boost EGL outcomes, particularly for disadvantaged children. A key aim of the LRCP is to facilitate stakeholder understanding and application of evidence-based, context-appropriate approaches to improving EGL.

The LRCP conducted the first systematic review of the evidence on EGL specific to the LAC region in 2017. The review covered both quantitative and qualitative EGL evidence produced between 1990 and 2015. The original systematic review was based on studies that met the following five criteria: (1) focused on literacy (reading and/or writing), (2) produced between 1990 and 2015, (3) focused on children birth through Grade 3, (4) focused in or on the LAC region, and (5) had a research question and corresponding methodology. Previously, systematic reviews on EGL tended to focus on evidence from the United States or Europe, which is a very different context from the LAC region, and on English literacy acquisition which is linguistically different from Spanish, French, Portuguese, and the indigenous languages of the LAC region. The LRCP team conducted the initial systematic review to better understand the research on EGL specific to the LAC region and the gaps in that evidence.

The original systematic review ultimately included 108 articles, comprising quantitative and qualitative research on the effects and fidelity of implementation of specific interventions, as well as quantitative and qualitative non-intervention research. Most of the studies included in the review were published journal articles from either Mexico or South America; significantly fewer articles were from Central America and the Caribbean. Most articles were published in English or Spanish, and more than 90% of the articles focused on high-income to upper-middle-income countries as classified by the World Bank.

The original systematic review identified many gaps in the evidence produced outside Mexico and South America; in languages other than English, Spanish, and Portuguese; and income levels of the countries where evidence was collected (mostly high-income to upper-middle-income countries). In addition, it identified gaps in the evidence from the region on topics such as EGL strategies for students with disabilities, literacy in indigenous languages, and prewriting and writing skills, among others. Finally, the original systematic review also identified gaps in the type and quality of research being produced in the region. The review did not find much longitudinal mixed-methods experimental or quasi-experimental research focused on EGL, and the sample sizes of experimental or quasi-experimental studies often were too small to detect small but meaningful effects of interventions on EGL outcomes with sufficient precision.

One of the objectives of the LRCP is to build the capacity to use and continue to produce high-quality research on EGL within the LAC region. Therefore, the research team will continue to search for new evidence to build on the original systematic review and in the future will summarize new evidence (and compare and contrast the new evidence with existing evidence) on a biannual basis to increase understanding of the status of research on how to improve the reading and writing skills of children in the LAC region. The research team also intends to transfer this work to the Red para la Lectoescritura Inicial de Centroamerica y el Caribe (RedLEI, Early Grade Literacy Network of Central America and the Caribbean), which is based at the Universidad del Valle in Guatemala, to continue when program

funding ends. RedLEI is tasked with generating new evidence on EGL throughout the LAC region and promoting the use of evidence to inform policy and practice.

The original systematic review collected EGL evidence produced from 1990 to 2015, while this systematic review update compares and contrasts evidence published in 2016 with evidence published in the original systematic review for quantitative nonintervention and qualitative studies. It also includes new quantitative intervention studies published in 2016 to contribute to the synthesis from the original systematic review. Finally, the update compares new evidence with the evidence produced before 2016 to see whether new evidence has contributed to addressing gaps in evidence identified during the original review. Download the original systematic review, which contains copies of the search strings and quality review protocols in the annexes, at www.lacreads.org.

Research Questions

Both the original systematic review and this synthesis examined a broad range of study types and designs, all focused on EGL in the LAC region. Both were defined by the following research questions:

1. What are the existing intervention- and nonintervention-based studies, and what is the existing literature from or on the LAC region involving literacy programs, practices, policies, and products focused on improving literacy skills for children from birth through Grade 3?

2. What is the quality of the existing EGL evidence (quantitative intervention and nonintervention and qualitative intervention and nonintervention) in the LAC region, and what is its practical use for various LAC region stakeholders?

3. What are the gaps in the evidence base on EGL in the LAC region compared to what we know globally about best practices in EGL?

4. What is the impact of literacy programs, practices, policies, and products aimed at improving the literacy skills of children from birth through Grade 3 on literacy outcomes in the LAC region?

5. What strategies have been successful and what is the evidence for this success? Which strategies were unsuccessful and why?

6. What are examples of using evidence and knowledge effectively to shape and/or improve EGL policy and practice in the LAC region?

Methods

This section describes the process that the LRCP team followed to conduct the systematic review update for 2016. This update employed the same research design and methodology as the original systematic review (LAC Reads Capacity Program, 2016) and included the following phases:

1. Searching for evidence

2. Extracting data from identified sources

3. Applying inclusion criteria and recording key indicators

4. Reviewing full text using quality review protocols

5. Analyzing data

6. Triangulating findings

The following sections introduce these phases. For more detailed information, please see the methods section of the original systematic review (LAC Reads Capacity Program, 2016).

1. Searching for Evidence

The LRCP team relied on the same search strings from the original systematic review. These strings were developed by identifying search terms using the **PICO criteria** (population, intervention, comparison, and outcomes), which are the standard criteria used in the Cochrane and Campbell systematic reviews (http://handbook.cochrane.org/). The terms below represent the keywords and phrases that were identified for our English search. Their equivalents in the other target languages are not listed here but are available upon request.

Population: Birth to grade 3, 0-10, early childhood, pre-school, pre-primary, primary, kindergarten, grade 1, grade 2, grade 3, day care, early-grade, elementary

Latin America, Caribbean, Central America, South America, Antigua and Barbuda, Argentina, Aruba, Bahamas, Barbados, Belize, Bermuda, Bolivia, Brazil, British Virgin Islands, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Curacao, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Mexico, Mont Serrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Barthelemy, Saint Kitts and Nevis, Saint Lucia, Saint-Martin, Saint Vincent and the Grenadines, Sint Maarten, Suriname, Trinidad and Tobago, Turks and Caicos, Islands, Uruguay, US Virgin Islands, Venezuela

Intervention: We did not search for terms such as "program" or "intervention." We accepted all evidence-based research about literacy.
Comparison: We did not search for a control or comparison group. We included all evidence-based research about literacy regardless of the use of a control or comparison group.

Outcomes: We included no search terms associated with outcomes. We included any study that had an early grade literacy-related outcome. This outcome could focus on students, teachers, or parents/community.

The LRCP team searched the same databases from the original systematic review and added two more journals at the recommendation of EGL experts in the LAC region: the *Journal of Education and Development in the Caribbean* (https://www.mona.uwi.edu/soe/publications/jedic) and the *Caribbean Journal of Education* (https://www.mona.uwi.edu/soe/publications/cje). These new journals were searched using the same protocols as for the original systematic review, and the time frame was expanded from articles published from 1990–2015 to include articles published in 2016. Only the two new journals were searched for the period 1990–2016.[1]

2. Extracting Data

The LRCP team used Mendeley Reference Management software to import all citations identified in the search. Mendeley automatically extracted bibliographic data from each book, article, or reference and removed all duplicates. At this stage, the review team collected and exported 772 unique citations to Excel.

3. Applying Inclusion Criteria

Because the number of articles collected for the 2016 search was significantly smaller than that of the original systematic review, the LRCP team reviewed each abstract manually and eliminated any that were clearly unrelated to EGL in the LAC region. After two rounds of review, and erring on the side of inclusion, approximately half the articles were eliminated. Researchers identified an additional seven articles from the *Journal of Education and Development in the Caribbean* and the *Caribbean Journal of*

---

[1] The list of searched websites, databases, and the associated search strings can be found on pg. 16 and 17 of the original systematic review document, and the search strings can be found in Appendix B of the original document.

Education and added them to the spreadsheet. The remaining 395 articles were then evaluated against the five inclusion criteria:

- Year of publication: 2016
- Relevance to the region: Studies must be from or on the LAC region.
- Relevance to the population: Studies must focus on boys and girls from birth through Grade 3.
- Relevance to the topic: Studies must focus at least partly on reading and/or writing, even if multiple topics were addressed.
- Is the study research? Studies must have a research question or objective and a methodology that matches that objective.

Reviewers rated each inclusion criteria in all studies as "yes," "no," "unclear," or "not rated" on an Excel spreadsheet. Reviewers categorized studies that met all inclusion criteria according to a set of key indicators and flagged them for further review. Any article that did not meet all inclusion criteria was removed from the review process.

4. Reviewing Full Text Against Quality Review Protocols

The LRCP team identified 55 articles and books for full text review. The articles were categorized according to the study's methodological approach. Researchers reviewed the articles using the same quality review protocols as for the full systematic review:

- **Quantitative intervention studies:** An adapted version of a risk of bias (RoB) assessment tool developed by Hombrados and Waddington (2012).[2]
- **Quantitative nonintervention studies:** An adapted version of the RoB tool for quantitative intervention studies, which removed any questions regarding interventions.[3]
- **Qualitative intervention and nonintervention studies:** An adapted version of the Critical Appraisal Skills Programme (CASP) Qualitative Research Checklist.[4]
- **Mixed-methods studies:** One of the RoB tools (depending on whether the study focused on an intervention or not) and the qualitative protocols.

Most articles were reviewed by a single reviewer, except for quantitative intervention studies, which were reviewed by two people. Reviewers resolved any disagreements through discussion or by third-party adjudication. After the full-text review, reviewers eliminated an additional 19 articles, which left 36 articles for inclusion in the data analysis.

5. Data Analysis

The LRCP team used different analyses for each type of research. First, they implemented a combination of meta-analysis and narrative synthesis to analyze the effects of programs on EGL outcomes. Second, they used a narrative review to examine the main lessons from the qualitative intervention and nonintervention studies. To identify the main lessons of the qualitative intervention and nonintervention studies, the LRCP team relied most heavily on the findings from what were determined to be medium to high-quality studies based on: rigor of research methods, appropriateness of research design, ethical research practices, and rigor of data analysis among others. Third, the main lessons about predictors of

---

[2] The full tool can be found in Appendix C of the original systematic review.
[3] The full tool can be found in Appendix D of the original systematic review.
[4] The full tool can be found in Appendix E of the original systematic review.

EGL outcomes in the LAC region from quantitative nonintervention studies were analyzed. Again, only studies identified as medium to high quality in the quality assessment were analyzed for this purpose.

Quantitative Intervention Studies

To analyze the quantitative intervention studies, the LRCP team relied on a combination of a narrative review and a meta-analysis. The LRCP team first calculated effect sizes for each quantitative study eligible for inclusion in the meta-analysis, both the studies published between 1999 and 2015 and the studies published in 2016. They then conducted a meta-analysis, which is a way of statistically pooling effect sizes from different studies to identify patterns among study results, sources of disagreement among those results, or other relationships that could come to light in the context of multiple studies. They conducted separate meta-analyses to determine the impact of teacher training, ICT, and nutrition programs on EGL outcomes (for more information, see the section on the meta-analysis). Other meta-analyses were not feasible because there were too few high-quality and relatively homogeneous studies focused on the effects of other programs on EGL outcomes. The meta-analyses also examined differences in effect sizes across studies with different risks of bias. Furthermore, the reviewers conducted a narrative synthesis for studies that could not be included in the meta-analysis but that showed evidence concerning the impact of other programs on EGL outcomes. The narrative synthesis included studies that emphasized the effects of preschool, school governance, specific teacher practices, and parental involvement on EGL outcomes.[5]

Qualitative Studies and Quantitative Nonintervention Studies

The LRCP team coded the quality review protocols for full-text qualitative and quantitative nonintervention articles using NVivo qualitative data analysis software (QSR International Pty Ltd., Version 10, 2012). NVivo is used to manage and code empirical data (Bhattacharyya, 2004; Caldeira & Ward, 2003; Patashnick & Rich, 2004).

Separate NVivo files were created for qualitative intervention research, qualitative nonintervention research, and quantitative nonintervention research to enable coding and analysis of the **quality ratings** and justifications for each article. The quality rating categories were adapted from a risk of bias (RoB) assessment tool developed by Hombrados and Waddington (2012) and the Critical Appraisal Skills Programme (CASP) Qualitative Research Checklist. After coding the quality criteria and justifications in NVivo, reviewers then synthesized the findings for each quality criterion for each research type.

The reviewers also used NVivo to synthesize **study findings** for each qualitative research type. As with the quality review ratings, researchers coded statements of findings into topic nodes predetermined by literacy experts as covering the main areas of EGL. Analysts compared the coded findings for each topic area and wrote analysis and implications by topic area.

6. Triangulating Findings

Reviewers triangulated the syntheses of findings from all research types. They examined the impact of programs on literacy outcomes and triangulated findings with the qualitative research studies to examine whether the fidelity of implementation or experiences and perspectives of stakeholders could have influenced the impact of these programs. In addition, the researchers assessed the predictors of literacy outcomes to increase their understanding of the linkages between intermediate outcomes, such as teacher knowledge and behavior, and literacy outcomes. Findings from the qualitative synthesis and the quantitative nonintervention synthesis helped describe, explore, and interpret how specific programs improved literacy outcomes. This approach enabled the LRCP team to capture the state of the evidence
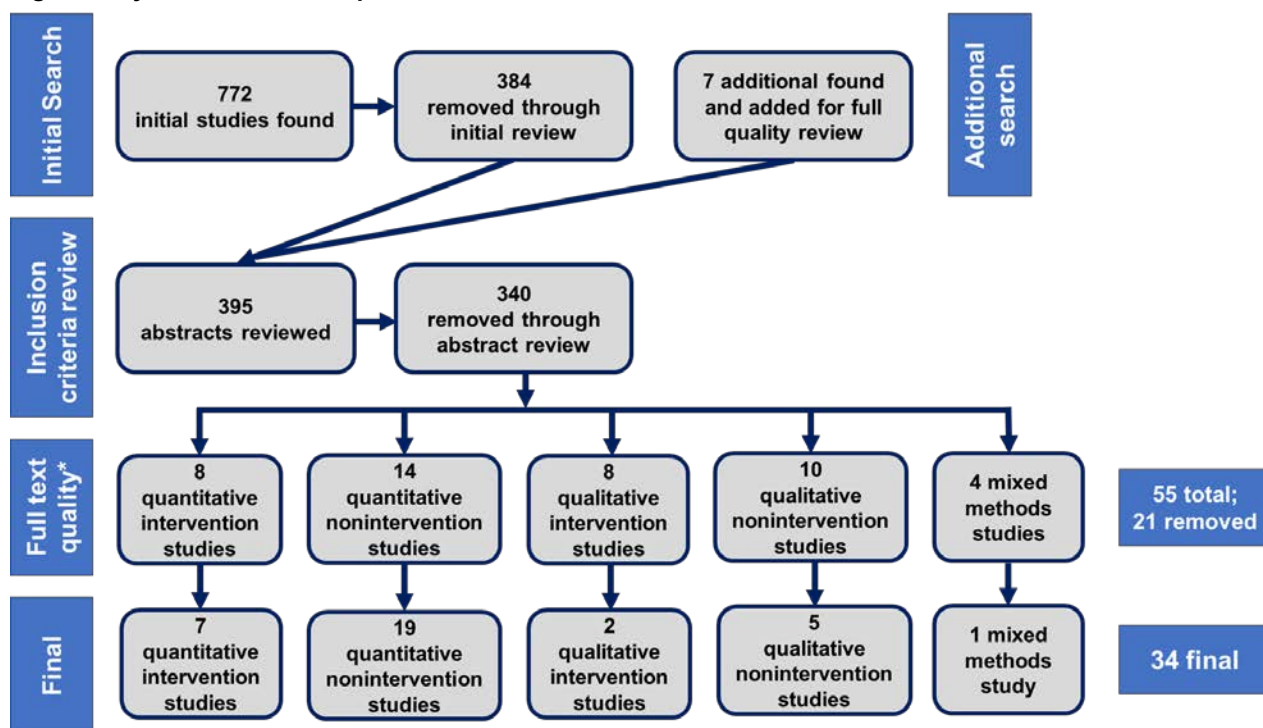
---

[5] For more detailed information on (1) calculation of effect sizes, (2) meta-analysis, and (3) narrative review, please see pp. 26–29 of the original systematic review.

on to what extent and in what ways specific programs improved literacy outcomes in Latin America, as well as gaps in the evidence.

Results of the Analyses

Using search strings and modified strings for all online sources, the LRCP team found 772 studies. Researchers manually reviewed these articles' titles and abstracts against the inclusion criteria and rejected an additional 384 articles. The full text of the remaining 395 articles were then reviewed in more detail against the inclusion criteria and, of these, 55 moved to the next phase, full-text review with quality protocols. In the final full-text article review stage, an additional 19 articles were rejected after quality reviewers determined that they did not fully meet one of the inclusion criteria. Figure 1 depicts the systematic review update phases from initial search through quality review. It indicates the number of studies that moved into each subsequent phase of review, as well as the number of studies that were eliminated at each phase.

Figure 1. Systematic Review Update Phases

Initial Search				Additional search
772 initial studies found	384 removed through initial review		7 additional found and added for full quality review	

Inclusion criteria review		
395 abstracts reviewed	340 removed through abstract review	

Full text quality*					
8 quantitative intervention studies	14 quantitative nonintervention studies	8 qualitative intervention studies	10 qualitative nonintervention studies	4 mixed methods studies	55 total; 21 removed

Final					
7 quantitative intervention studies	19 quantitative nonintervention studies	2 qualitative intervention studies	5 qualitative nonintervention studies	1 mixed methods study	34 final

*Could not immediately identify methodology for 11 studies during the full text quality review

Characteristics of Included Studies

Table 1 summarizes the characteristics of all studies included in the final review update by publication type, region and country of focus, language of publication, research type, and income level of the country in focus (World Bank, n.d.).

Table 1. Characteristics of Included Studies (2016)

Characteristic	*N*	Percentage
Publication type		
Book chapter	1	3%
Dissertation/thesis	4	12%
Journal article	27	79%
Technical report	1	3%
Other	1	3%
Region and country of focus		
Caribbean		6%
Jamaica	2	
Central		6%
Guatemala	2	
North		6%
Mexico	2	
South		79%
Brazil	13	
Chile	13	
Colombia	1	
Multiple countries	1	3%
Language of publication		
English	23	64%
Spanish	6	19%
Portuguese	5	17%
Type of research		
Quantitative nonintervention	19	56%
Quantitative intervention	7	21%
Qualitative nonintervention	5	15%
Qualitative intervention	2	6%
Mixed methods	1	3%
Country of focus income level (per World Bank)		
Lower middle income	2	6%
Upper middle income	18	53%
High income	13	38%
NA/multiple countries	1	3%

As can be seen in Table 1, most studies included in the systematic review update were published journal articles. This aligns with the original systematic review, for which 90% of the studies were published journal articles. In addition, 79% of the articles included in the systematic review update were from South America, primarily from Brazil and Chile—also the two highest producers of research from South America in the original systematic review. Central America and the Caribbean continue to produce the fewest studies on EGL. English also continues to be the main language of publication, followed by Spanish and then Portuguese. We again found no articles in any of the indigenous languages of the region. Finally, almost all research was conducted in high-income to upper-middle-income countries (91% compared to 90% in the original systematic review).

The following sections summarize the results of the systematic review of the literature for the 2016 review. The LRCP team describes the quality parameters and present the results of quality review for the four types of research including quantitative intervention, quantitative nonintervention, qualitative intervention, and qualitative nonintervention. Findings are triangulated across research types.

Quantitative Intervention Research

Analyses

The LRCP team included and reviewed 28 experimental and quasi-experimental papers that focused on determining the effects of various programs on EGL outcomes in Latin America. Of the 28 papers included, three estimated the impact of more than one program (Cardoso-Martins, Mesquita, & Ehri, 2011; Larrain, Strasser, & Lissi, 2012; Vivas, 1996). Of the 28 papers reviewed, five were published in 2016 and had thus not been included in the original systematic review (LAC Reads Capacity Program, 2016). In the meta-analyses, each unique program is treated as one observation. Seven papers focused on the effect and implementation of the Buen Comienzo program in Chile (Arbour, Yoshikawa, Willett et al., 2016 Bowne, 2014; Bowne, Yoshikawa, & Snow, 2016; Gomez Franco, 2014; Mascareño, Snow, Deunk, & Bosker, 2016; Mendive, Weiland, Yoshikawa, & Snow, 2016; Yoshikawa et al., 2015). Of the seven papers that focused on this program, only the paper with the lowest risk of selection bias was included in the meta-analysis to determine the effects of education programs on EGL outcomes (Yoshikawa et al., 2015). In addition, researchers included a narrative synthesis to examine the mechanisms of how the education program in Chile has affected EGL outcomes because several of the papers that focused on this program included a richer analysis with an emphasis on intermediate outcomes and the estimation of heterogeneous effects (Arbour, Yoshikawa, Willett et al., 2016; Bowne, 2014; Bowne et al., 2016; Mascareño et al., 2016; Mendive et al., 2016).

The studies included were diverse in terms of program characteristics, outcome measures, sample size, evaluation design, context, and type of analysis. For this reason, the LRCP team weighted the narrative synthesis of the experimental and quasi-experimental studies the same as the meta-analysis. This equal emphasis aligned with the recommendations of Waddington et al. (2012) in their toolkit for systematic reviews in international development. Importantly, however, researchers were unable to include in the meta-analysis any of the five experimental or quasi-experimental studies that had not been included in the original systematic review because the meta-analysis treats each unique program as one observation. Each of the newly included studies either focused on the same program (e.g., the Buen Comienzo program in Chile) as in the previous systematic review (LAC Reads Capacity Program, 2016) and had a higher or equal risk of bias than the previously included studies, and/or focused on an intermediate outcome variable as opposed to EGL outcomes (Arbour, Yoshikawa, Willett, et al., 2016; Bowne, 2014; Bowne et al., 2016; Mascareño et al., 2016).

The main characteristics of the included studies—program characteristics, outcome measures, sample size, study design, and analysis—can be accessed through the following link: https://lacreads.org/sites/default/files/documents/summary_of_quant_int_characteristics_final.pdf

Program Characteristics

The 28 included papers focused on programs with an emphasis on teacher training, nutrition, ICT, preschools, specific teacher practices inside and outside the classroom, and specific parental practices outside the classroom. Of the included evaluations, six focused on the impact of specific teacher practices, such as reading aloud or the explicit instruction of new words (Bowne, 2014; Larrain et al., 2012; Neugebauer & Currie-Rubin, 2009; Viramontes, Morales-Sifuentes, & Delgado-Nájera, 2016; Vivas, 1996); three focused on the effects of parental involvement, for example by paired reading (Murad & Topping, 2000; Tapia & Benitez, 2013; Vivas, 1996); five examined the impact of programs with an emphasis on nutrition, such as food supplements (Adrogue & Orlicki, 2013; Ismail, Jarvis, & Borja-Vega, 2014; Maluccio et al., 2009; Powell, Walker, Chang, & Grantham-McGregor, 1998; Simeon, Grantham-McGregor, & Wong, 1995); four estimated the impact of ICT programs, such as the distribution of laptops to children and computer-aided instruction (Beuermann, Cristia, Cueto, Malamud, & Cruz-Aguayo, 2015; Cristia, Ibarrarán, Cueto, Santiago, & Severín, 2012; Ferrando, Machado,

Perazzo, & Vernengo, 2011; Osorio & Linden, 2009); two focused on the impact of preschool (Campos et al., 2011; Felicio, Terra, & Zoghbi, 2011); six estimated the impact of teacher training (Arbour, Yoshikawa, Willett, et al., 2016; Bowne et al., 2016; Gomez Franco, 2014; Mendive et al., 2016; Pallante & Kim, 2013; Yoshikawa et al., 2015); and two emphasized the effects of school governance reforms (Bando, 2010; Lockheed, Harris, & Jayasundera, 2010).

Outcome Measures

This section presents the outcome measures used to determine program impact in the experimental and quasi-experimental studies. The included studies estimated program impacts on literacy outcomes such as letter–word identification, emergent writing, and picture vocabulary as well as letter-sound recognition, but also on constructed indices based on different elements of literacy outcomes. Five other studies measured impacts on intermediary outcomes such as literacy practices (Beuermann et al., 2015; Tapia & Benitez, 2013) and literacy instruction (Bowne, 2014; Bowne et al., 2016; Gomez Franco, 2014).

Each literacy outcome can be measured as part of a larger construct. Because literacy is a broad concept that can be subdivided into many separate constructs, researchers can use a variety of working definitions to measure literacy outcomes and practices. Several studies constructed indices based on different elements of literacy outcomes, whereas others were more specific in their definition of literacy outcomes and practices. Both approaches have their benefits. Relying on an index addresses the so-called "indicator soup" problem, which refers to the difficulty of organizing and interpreting results with many outcome variables (King, Samii, & Snilstveit, 2010). However, using indices can also lead to a loss of detail, for instance when interventions have positive effects on picture vocabulary but not on decoding. Of the 28 included evaluations, 20 contained more than one outcome measure. Furthermore, 12 of the included studies relied exclusively on existing or administrative data to determine the impact of the program, and the remaining studies collected their own data.

Context

The experimental and quasi-experimental studies focused on a wide range of countries in the LAC region, but high-income economies were overrepresented. Of the 28 included studies, 15 focused on high-income countries, 11 focused on upper middle–income countries, only two focused on lower middle–income countries, and none on the only low-income country in the region (Haiti). Chile is particularly overrepresented among the high-income economies; of the included experimental and quasi-experimental studies, 10 focused on Chile. These findings indicated that we might not be able to generalize the findings of our synthesis of the quantitative intervention studies to low- or middle-income economies. The overrepresentation of high-income countries might be due to limited resources and researchers' capacity in middle-income and low-income economies to conduct experimental and quasi-experimental evaluations. Most studies conducted in high-income and upper-middle-income countries are authored or coauthored by researchers based in the countries of interest, whereas most studies conducted in lower income countries are implemented by researchers based in the United States or Europe.

## *Evaluation Design and Sample Size*

The quantitative studies included had to focus on program effectiveness by using either an experimental or a quasi-experimental design. Of the 31 program evaluations included, 23 conducted a randomized controlled trial (RCT) and eight used quasi-experimental design. Furthermore, seven of the 31 included evaluations had a sample size of fewer than 100 students to determine the impact of the program, 11 evaluations determined the impact of the programs with a sample size between 100 and 1,000 students, and 13 evaluations had a sample size of more than 1,000 students. As discussed in the original systematic review, most of the small sample studies focused on a program that was implemented for academic

purposes, while studies with sample sizes of more than 1,000 students often focused on the effectiveness of government-supported programs (LAC Reads Capacity Program, 2016).

Critical Appraisal

For a critical appraisal of the studies, the LRCP team relied on a risk-of-bias assessment tool with 71 questions, with which they could accurately determine four types of risk of bias. The tool was an adapted version of a risk-of-bias assessment tool developed by Hombrados and Waddington (2012). The LRCP team examined the risk of selection bias and confounding, performance bias, outcome and analysis reporting bias, and other biases. The complete risk-of-bias assessment tool and a detailed assessment of the risk of bias of each individual study can be accessed through the following link: https://lacreads.org/sites/default/files/documents/risk_of_bias_for_quant_int_studies_final.pdf. Figure 2 shows the distribution of low-, medium-, and high-risk bias across the 31 included evaluations for each risk-of-bias category.

Figure 2. Risk-of-Bias Assessment of Quantitative Intervention Studies
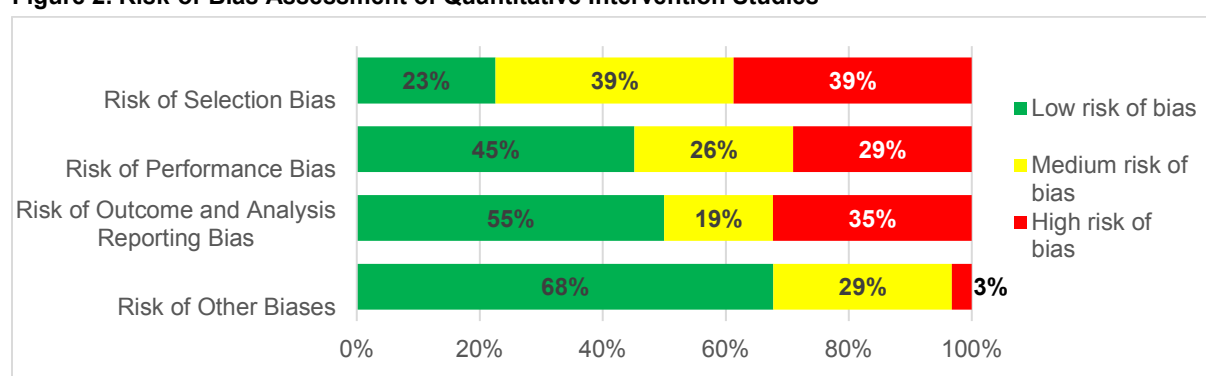

Synthesis of Quantitative Intervention Studies

This section presents results from the meta-analysis and narrative review of the effects of different types of programs on literacy outcomes. A separate analysis is presented for each of the program types evaluated in the primary studies, including teacher training, ICT, nutrition, and school governance programs; programs with an emphasis on teacher practices; and programs with an emphasis on parental involvement. Detailed descriptions of the meta-analyses can be found in the original systematic review (LAC Reads Capacity Program, 2016). This updated systematic review does not include any experimental or quasi-experimental studies in the meta-analysis that were not included in the original systematic review (LAC Reads Capacity Program, 2016). However, new narrative syntheses are included that examine how programs can improve EGL outcomes in the LAC region.

Teacher Training Programs

A meta-analysis of the effects of teacher training programs on EGL outcomes (based on two studies) found no statistically significant effects, but a narrative synthesis suggested that comprehensive teacher training programs complemented by coaching and sustained follow-up could have positive effects on EGL outcomes. The researchers found no evidence that, on average, teacher training had a positive effect on literacy outcomes (SMD = 0.16, 95% confidence interval (CI) = -0.17, 0.48; evidence from two cluster-RCTs). However, Pallante and Kim (2015) found positive and statistically significant effects of a teacher training program in Chile that included a focus on coaching and sustained follow-up. Studies that focused on the effects of the Buen Comienzo program in Chile found positive program effects of teacher training on the emotional and instructional support of teachers and the probability that teachers will focus on literacy in their classes (Bowne et al., 2014; Yoshikawa et al., 2015), but the results suggested that these behavioral changes did not translate to positive effects on EGL outcomes, possibly

because of challenges in program implementation (Mendive et al., 2016). In addition, Arbour, Yoshikawa, Willett, and colleagues (2016) showed that the Buen Comienzo program had positive effects on children with the lowest likelihood of absenteeism, which suggests that student absenteeism moderates the effects of teacher training on EGL outcomes. Finally, more time spent on literacy instruction may come at the expense of time spent teaching other topics. Bowne and colleagues (2014) showed that the Buen Comienzo program led to more time spent on literacy instruction by teachers but less time spent on vocabulary support during other-topic instruction. The meta-analysis is presented in Figure 3.

Figure 3. Impact of Teacher Training Programs on Early Grade Literacy Outcomes
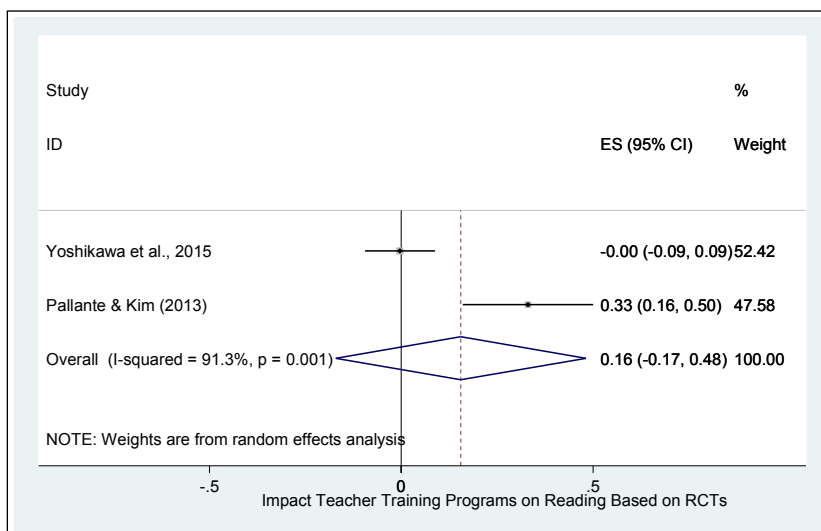
These results suggest that relatively large program effects on teacher practices may be insufficient for achieving positive program effects on EGL outcomes, especially when school attendance is relatively low. Several of the newly included studies that focused on teacher training (and especially the Un Buen Comienzo program in Chile) found relatively large positive program effects on teacher practices, such as time spent on literacy instruction, and time spent on vocabulary support during literacy instruction, but these changes were insufficient to achieve statistically significant effects of the Un Buen Comienzo program on EGL outcomes (Bowne et al., 2014). We present the relatively large effect sizes on teacher practices in Table 2 (including the unintended negative effects on time spent on other-topic instruction). Care must be exercised in interpreting this result, however, because it is based solely on evaluations of a single program in a high-income country. Changes in teacher practices could have different effects on EGL outcomes in low- and middle-income countries.

Table 2. Effects of the Un Buen Comienzo Program on Teacher Practices

Outcome	Effect size	Confidence interval
Time spent on literacy instruction	0.86 SMDs	0.27, 1.46
Time spent on other-topic instruction	−0.56 SMDs	−1.15, 0.02
Time spent on vocabulary support in literacy instruction	0.49 SMDs	−0.16, 1.13
Time spent on vocabulary support during other-topic instruction	−0.67 SMDs	−1.33, −0.01

ICT Programs

A meta-analysis on the effects of the one-laptop per child program on EGL outcomes found evidence of negative and statistically significant effects on early literacy outcomes when findings from two RCTs in Peru and a quasi-experimental study in Uruguay (SMD = -0.06, 95% confidence interval (CI) = -0.11,

0.00; evidence from three studies) were pooled, but a computer distribution program in Colombia had positive effects on EGL outcomes (Osorio & Linden, 2009). The first result showed that ICT programs did not consistently have positive effects on EGL outcomes and may have had negative effects in some cases. However, a cluster-RCT for a computer distribution program in Colombia found evidence of positive effects of approximately 0.14 SMDs on the EGL outcomes of third-grade children even in the presence of implementation problems (Osorio & Linden, 2009). Figure 4 depicts the meta-analysis for the one laptop per child program.
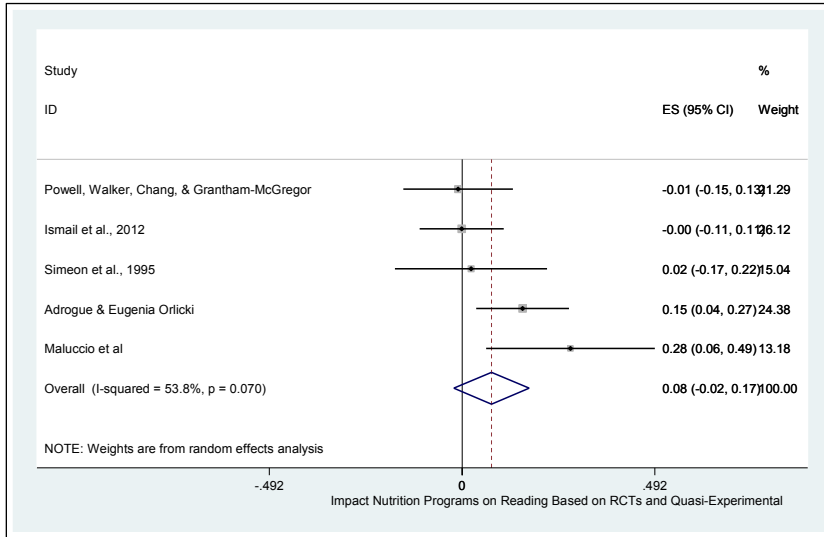
Figure 4. Impact of One Laptop per Child Program on Literacy Outcomes Based on RCTs and Quasi-Experimental Studies

Study ID	ES (95% CI)	% Weight
Cristia et al. (2012)	-0.08 (-0.15, -0.01)	59.82
Ferrando et al., (2012)	-0.05 (-0.15, 0.05)	30.01
Beuermann et al. (2015)	0.05 (-0.12, 0.22)	10.17
Overall (I-squared = 0.0%, p = 0.378)	-0.06 (-0.11, -0.00)	100.00

NOTE: Weights are from random effects analysis

Impact ICT Programs on Reading Based on RCTs and Quasi-Experimental
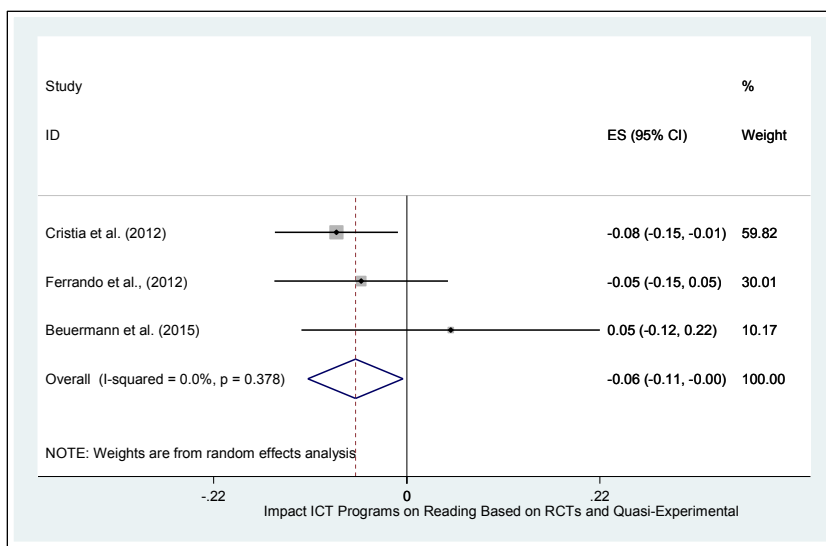
Nutrition Programs

A meta-analysis of the effects of nutrition programs, such as school feeding programs and micronutrient supplements, found no evidence of positive average effects on EGL outcomes, but the results suggest that nutrition programs may be more effective in low-income country settings, such as Guatemala. Specifically, a meta-analysis of impact evaluations of deworming and a school breakfast program in Jamaica, an impact evaluation of a program that included the distribution of supplementary nutritious drinks in Guatemala 25 years after the start of the intervention, and school feeding programs in Guyana, found no evidence that school feeding programs had positive and statistically significant effects on EGL outcomes in the LAC region (SMD = 0.08, 95% CI=-0.02, 0.17; evidence from five studies). However, the quasi-experimental study in Guatemala found evidence for positive program effects of micronutrient supplements of 0.28 standardized mean differences on literacy outcomes 25 years after the start of the program (Maluccio et al., 2009). Figure 5 depicts the meta-analysis for the impact of nutrition programs on EGL outcomes.

Figure 5. Impact of Nutrition Programs on Literacy Outcomes in the LAC Region Based on RCTs and Quasi-Experimental Studies

Study ID	ES (95% CI)	% Weight
Powell, Walker, Chang, & Grantham-McGregor	-0.01 (-0.15, 0.13)	21.29
Ismail et al., 2012	-0.00 (-0.11, 0.11)	26.12
Simeon et al., 1995	0.02 (-0.17, 0.22)	15.04
Adrogue & Eugenia Orlicki	0.15 (0.04, 0.27)	24.38
Maluccio et al	0.28 (0.06, 0.49)	13.18
Overall (I-squared = 53.8%, p = 0.070)	0.08 (-0.02, 0.17)	100.00

NOTE: Weights are from random effects analysis

Impact Nutrition Programs on Reading Based on RCTs and Quasi-Experimental

School Governance Programs

Two quasi-experimental studies found mixed evidence that school governance programs had positive effects on EGL outcomes in the LAC region. A matching grant program in Mexico, in which parents were given information and decision-making power to spend the matching grant, showed positive effects on EGL outcomes (Bando, 2010). However, a study in Jamaica did not find evidence that a school improvement program had positive effects on literacy outcomes for students in Grade 4, possibly because the program had only small positive effects on school inputs, such as supplementary reading materials, and functioning computers (Lockheed et al., 2010).

Preschool Programs

Two quasi-experimental studies found evidence for positive effects of preschool programs on EGL outcomes, but both studies suffered from a high risk of selection bias (Campos et al., 2011; De Felicio et al., 2011). Hence, these studies do not present convincing evidence that participation in preschool leads to improvements in EGL outcomes. Participation in preschool might have these effects in the LAC region, but more rigorous research is needed to demonstrate these effects.

Teacher Practices

The included studies on teacher practices presented interesting findings about how specific teacher practices could influence literacy outcomes, but they generally suffered from high selection bias. Most studies with an emphasis on teacher practices found evidence for positive effects on EGL outcomes despite relatively small sample sizes ($n < 100$ in most studies). However, these statistically significant effects were unlikely an indicator of the effectiveness of teacher practices. Instead, evidence suggests that published studies with small sample sizes and statistically significant effects are disproportionally affected by publication bias (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Nonetheless, the included studies on teacher practices presented some interesting hypotheses about how specific teacher practices could influence literacy outcomes. The studies highlighted how word elaboration during shared story book reading, using simpler definitions of words, teaching the names of letters, training children in phonological awareness,[6] reading aloud, and listening to teachers reading stories aloud could result in improvements in EGL outcomes (Cardoso-Martin et al., 2011; Larrain et al.,

---

[6] Phonological awareness is the ability to orally manipulate the sounds of a language.

2012; Vivas, 1996). These hypotheses should be tested in larger scale studies, however, because the combination of a small sample size and statistically significant findings could have resulted in publication bias (Borenstein et al., 2009).

Parental Involvement

Similar to studies on teacher practices, studies on parental involvement showed how parental involvement could influence literacy outcomes, but the studies were generally not credible because of a high risk of selection bias. For example, studies highlighted how teaching mothers about joint reading of stories and puppet play, listening to stories read aloud by parents, and paired reading with parents could positively influence EGL outcomes (Murad & Topping, 2000; Tapia & Benitez, 2013; Vivas, 1996). However, because of the small sample size, these studies might be vulnerable to publication bias (Borenstein et al., 2009). For this reason, the research team recommends studying these hypotheses in larger scale studies.

Publication Bias

As indicated in the original systematic review, an analysis of publication bias suggested that the unusually high statistical significance in studies with a smaller sample size in the review was a strong indication for publication bias in the included studies with a focus on teacher practices and parental involvement. It was likely that these studies presented a biased overview of the impact of specific teacher practices and parental involvement on EGL outcomes. In addition, the research team found some evidence for publication bias in the studies that could be included in the meta-analysis, but care is required in interpreting this result because statistical tests for publication bias can only be considered indicative of publication bias and not be used as definitive proof.

Synthesis of Findings

The results suggest that teacher training programs can have positive effects on EGL outcomes when they are complemented by coaching, and that nutrition programs can be effective in improving EGL outcomes when they are implemented in countries with high rates of stunting and wasting. But overall, the LCRP team finds little evidence for research that can convincingly demonstrate causal effects of interventions on EGL outcomes. There is a clear evidence-gap on the effects on EGL outcomes of preschool programs, school governance programs, teacher practices, and parental involvement. In addition, more mixed-methods evidence is needed on how ICT programs can positively affect EGL outcomes. Research published in 2016 primarily focused on programs that had already been included in the original systematic review (LAC Reads Capacity Program, 2016). The LCRP did not find major new evidence on the effects of specific interventions on EGL outcomes, so important evidence gaps remain. Importantly, however, the new research shows that large improvements in teacher practices are not necessarily sufficient to improve EGL outcomes. Research from Chile demonstrates that programs may not have positive effects on EGL outcomes—even if they show large and positive effects on teacher practices.

Evidence-Gap Maps

The analysis in the original systematic review showed several evidence gaps related to the intervention type, geography, and quality of the included studies (LAC Reads Capacity Program, 2016). First, the research team found only three topic areas with more than two impact evaluations with a sufficient sample size that focus on EGL outcomes in the LAC region: teacher training, nutrition programs, and ICT programs. Furthermore, the evidence predominantly focused on high-income and upper-middle-income countries in the LAC region. In addition, only eight of the included studies qualified as studies with a low risk of selection bias. Of these studies, only three qualified as studies with a low risk of performance bias. These studies were the only included studies that could make credible causal claims about the impact on EGL outcomes without significant risks of spillovers and contamination. These three studies focused on

a teacher training program in Chile, the distribution of one laptop per child in Peru, and the provision of computers for computer-aided instruction in Colombia. This finding indicated that an evidence gap existed on what works to improve literacy outcomes in the LAC region; this question could be credibly addressed only in studies that qualify as having a low risk of selection bias and a low risk of performance bias.

Tables 3–6 present evidence-gap maps that focus only on study designs that enabled credible causal claims about the impact of development programs on EGL outcomes. These maps include only those studies having a low risk of selection bias and a low risk of performance bias. The tables show that all studies that enabled the demonstration of causal claims about the impact of development programs on EGL outcomes were cluster RCTs that focused on high-income or upper-middle-income economies.

Table 3. Evaluations by Country

	Literacy skills	Student literacy practices	Parental literacy practices
Chile	Yoshikawa et al. (2015)	N.A.	N.A.
Colombia	Osorio & Linden (2009)	N.A.	N.A.
Peru	Cristia et al. (2012)	N.A.	N.A.

Table 4. Evaluations by Study Design

	Literacy skills	Student literacy practices	Parental literacy practices
RCT	Yoshikawa et al. (2015); Osorio & Linden (2009); Cristia et al. (2012)	N.A.	N.A.

Table 5. Evaluations by Country Type

	Literacy skills	Student literacy practices	Parental literacy practices
High income	Yoshikawa et al. (2015)	N.A.	N.A.
Upper middle income	Osorio & Linden (2009); Cristia et al. (2012)	N.A.	N.A.

Table 6. Evaluations by Intervention Type

	Literacy skills	Student literacy practices	Parental literacy practices
ICT	Osorio & Linden (2009); Cristia et al. (2012)	N.A.	N.A.
Teacher training	Yoshikawa et al. (2015)	N.A.	N.A.

Quantitative Nonintervention Research

Quantitative nonintervention studies allow for learning about the predictors of EGL outcomes and the mechanisms that explain changes in EGL outcomes. Systematic reviews typically do not include quantitative nonintervention studies because these study types are unable to address counterfactual questions. It is important to include these studies, however, because they often examined the specifics of literacy acquisition mechanisms and trajectories. In addition, these studies could uncover predictors of literacy success, as part of the larger story of evidence of EGL development in the LAC region. The research team believes that these studies can guide curricular and standards development; entangle specific aspects of how a "bundled" EGL program may impact literacy; and help develop more targeted, language- and country-specific literacy measures.

The quantitative nonintervention studies comprised the largest number of studies in the systematic review update. The review included 19 articles that were published in 2016 from the following countries: Brazil ($N = 11$), Chile ($N = 4$), Guatemala ($N = 1$), Mexico ($N = 1$), and Colombia ($N = 1$). One study

that analyzed data from both Peru and Guatemala also was included. This is a proportionally large increase in the number of quantitative nonintervention studies from the original review, which included only 62 studies since 1990. The included studies covered a range of topics on predictors of literacy skill development in the LAC region, although they were highly concentrated in South America, particularly in Brazil.

The following section presents the quality criteria that determined which studies were included in the review, the main findings that emerged from those studies, and implications for stakeholders and relevance to the field at large.

Quality Criteria

Reviewers rated all nonintervention studies on a range of questions pooled together to target the following categories of quality: outcome measures, sample, data collection, data analyses, and external validity.

Outcome Measures

The research team reviewed each study to determine if the main outcome being measured was a reading, writing, or some reading- or writing-related subskill. Two main questions were used to determine whether a study was included:

1. Did the outcome measure include some measure of literacy or a literacy subskill (e.g., fluency, phonological awareness, language decoding, letter knowledge, or comprehension)?

2. If the study did not include a measurement of literacy or a literacy subskill, was literacy measured in a different manner?

All 19 studies included an outcome measure of literacy. Seventeen of the 19 studies included an outcome measure of reading or a reading subskill and the remaining two studies had an outcome measure focused on writing. All the studies except for dos Santos Santana, de Oliveira, and dos Reis (2016), which used a writing excerpt, used literacy assessments to measure reading and writing outcomes, which reduced the risk of measurement error. No study relied on information from self-reports.

Sample

For each study, the research team reviewed whether the sample selection criteria were justified and whether the sample was appropriate for answering the research question. Fourteen of the 19 studies provided sample selection criteria or justification of the sample selection process. Samples were generally described by age, grade, gender, economic level, country, and geographical region. Five studies did not justify their sample or provided only partial information. Sargiani and Albuquerque (2016), for example, stated that "thirty-eight children enrolled in the last-year classes of Early Childhood Education of a private school ($n = 17$) and a public school ($n = 21$) were included in the sample" (p. 594; English translation of original Portuguese), but they did not justify the size of the sample or how those schools or grades were chosen.

Data Collection

The research team aimed to determine the quality of various aspects of data collection, including the training of test administrators, data collection procedures, and consideration of potential data collection implementation challenges. Because the research team had to rely on study authors to report this information, we were cautious in interpreting these results. In some cases, study authors may not have reported on the training of test administrators, data collection procedures, and data collection implementation challenges even if the quality of data was high.

Fourteen studies reported on data collection procedures. The procedures ranged from individual to group administration of tests in the school. For example, Barbosa, Medeiras, & Vale (2016) stated, "The tests were applied in quiet rooms designated by the school board in two moments of the 2008 school year: during the months of March and April and during the month of December" (p. 670; English translation of original Portuguese). Five studies did not report any information on the data collection procedures. Only six studies reported considering data collection implementation failures, such as applying a test originally designed for older students (Rosal, Andrade Cordeiro, da Silva, & de Queiroga, 2016) and possible recall bias by participants on a questionnaire (Jorge, Botelho, Silva, & Moi, 2015).

Data Analysis

The research team asked the following questions to determine the quality of the analysis:

1. **Was there a description of the analytic method(s) used?** 18 of the 19 studies described the analysis methods to some degree. Some studies provided a full description of the statistical analyses (Paez, Tabors, & Lopez, 2007), whereas others gave brief descriptions and used simple analyses such as histograms (Bandini, Oliveira, & Souza, 2006). One study did not describe the analytic method (Vilhena, Sucena, Castro, & Pinheiro, 2016).
2. **Did the study justify the analysis method (was the analysis method appropriate for the research question/objective)?** Only 10 of the 20 studies justified the analysis method selected. However, all studies used analysis methods that were appropriate for the research question they were attempting to answer.
3. **Were any participants not included in the analysis? If so, was there justification for why?** Students were excluded for various reasons in 10 of the 19 studies. Four studies excluded students who did not provide signed consent forms (Cadavid-Ruiz, Quijano-Martínez, Escobar, Posas, & Tenorio, 2016; Jorge et al., 2016; Lopes-Silva et al., 2016; Sargiani & Albuquerque, 2016); others excluded children with learning disabilities (Lopes-Silva et al., 2016; Mariángel & Jiménez, 2016), and others excluded those who did not meet a predetermined threshold on a pretest (Lugo-Gil et al., 2016; Nobre & Salles, 2016). Nine studies did not specify whether participants were excluded or why.
4. **Was data reported on covariates?** Information on covariates was reported in 14 of the 19 studies. Covariates included measures of characteristics similar to those for sample descriptions (e.g., age, grade, gender, economic level, country, and geographical region). Four studies included additional covariates, such as parents' educational levels and socioeconomic status (Rosal et al., 2016; Silva Pacheco, Gabbard, Ries, & Bobbio, 2016; Strasser, Vergara, & Río, 2016) and their intelligence quotient (Nobre & Salles, 2016).
5. **Were appropriate reliability scores available for all tests?** Twelve of the 19 studies reported reliability scores for the tests. These studies commonly used Cronbach's alpha to calculate reliability scores (Lopes-Silva et al., 2016; Mariángel & Jiménez 2016; Strasser et al., 2016; Vergara, Strasser, & del Río, 2016). Studies with tests having reasonable reliability scores were rated as high quality.

External Validity

For external validity, the research team determined whether authors generalized their findings only to the relevant population of study. Fourteen of the studies appropriately generalized the study outcomes only to the population in the study. One study generalized the study findings across school types in Chile (Helman, Delbridge, Parker, Arnal, & Jara Mödinger, 2016), one to the entire country (Sargiani & Albuquerque, 2016), and one as applicable to all children (Vergara et al., 2016).

Synthesis of Quantitative Nonintervention Studies (by Topic)

Multiple themes emerged from the nonintervention quantitative studies. Nine of the 19 studies involved a measure of one or more literacy skills and looked at correlations between skills. Others focused on

assessment validation, disabilities, smoke exposure, teacher training and community interventions, and writing.

Literacy Skills

Nine of the 19 studies focused on some measure of literacy skills: Silva Pacheco et al. (2016); Lopes-Silva et al. (2016); Mariángel & Jiménez (2016); Nobre & Salles (2016); Strasser et al. (2016); Ureta, Fortin, & Cruz (2006); Vergara et al. (2016); Barbosa et al. (2016); and Guimarães & Mota (2016).

Silva Pacheco et al. (2016) examined gross motor and academic skills in 100 Brazilian children using the Bruininks–Oseretsky Test of Motor Proficiency and the Academic Performance Test and found a positive relationship between gross motor behavior—particularly for activities involving interlimb coordination[7]—and academic performance. The authors suggested that early testing on interlimb coordination might be useful as a measure to identify children at risk of academic difficulties but this requires further testing to validate the results of this study.

Lopes-Silva et al. (2016) investigated shared and nonshared mechanisms involved in reading and writing words and Arabic numerals in 172 second through fourth grade children in Brazil. The authors found that phonemic awareness was the only cognitive variable that systematically predicted all the academic skills investigated, both numerical and word tasks. Intelligence and the phonological and visuospatial components of working memory (WM) were not found to significantly predict the literacy and numeracy skills. Research is clear that phonemic awareness is an important predictor of reading acquisition across languages (Ehri et al., 2001) but this research suggests that it is also important for numeracy development.

Mariángel and Jiménez (2016) compared the development of syntactic awareness[8] and phonological awareness in 234 Chilean children from Grades 3 and 5. The main finding was that phonological and syntax comprehension evolve across time, even as children pass through elementary school. This suggests that these skills might need to be continually developed and supported for a longer time frame than was previously understood.

Nobre and Salles (2016) investigated whether word reading and reading comprehension correlated with semantic priming effects[9] (SPE) and if lexical semantic processing, as measured by SPE, predicts word reading or reading comprehension among 68 7- to 12-year-old students in Brazil. The study found a relationship between SPEs and word reading measures, which was different from what Hoover and Gough (1990) posited in the Simple View of Reading (SVR). In SVR, semantic skills contribute to listening comprehension but not to decoding. This study suggested that both decoding and comprehension benefited from semantic processing, at least in the case of real words. More research is needed to see if these results hold true for other languages.

Strasser et al. (2016) examined the contribution of print exposure to oral language (expressive vocabulary and listening comprehension) and literacy (word reading and reading comprehension) in 281 first- and second-grade Chilean children. Study results indicated that exposure to print was an important factor in explaining young children's oral language and reading comprehension. The longitudinal results showed that print exposure measured in first grade had an indirect correlation with listening comprehension, expressive vocabulary, word reading, and reading comprehension in second grade. The study also found evidence that word reading should be considered a mediator in the link between print exposure and

---

[7] Interlimb coordination primarily involves movements requiring sequential and simultaneous use of both sides of the body.
[8] Syntactic awareness is a metalinguistic skill that involves applying grammatical rules to manipulate or judge word order within the context of a sentence.
[9] SPEs occur when exposure to one semantic stimulus influences a response to a subsequent stimulus, without conscious guidance or intention. For example, the word *nurse* is recognized more quickly following the word *doctor* than following the word *bread*.

reading comprehension in first grade in Spanish, which was similar to previous results obtained in English. However, household income may have served as a confounding factor in the relationship between print exposure and reading comprehension (Hart & Risley, 2003).

Vergara et al. (2016) attempted to identify what basic skills and abilities children have when successful at reading and listening comprehension. Multiple regression models revealed that higher order cognitive skills and knowledge (such as monitoring one's own understanding and verbal reasoning) were as important, and in some cases more so, than decoding in explaining the reading comprehension of 202 first-grade children although they did not accurately predict listening comprehension. These results suggested the importance of teaching children skills such as self-monitoring, self-correcting, and reasoning around words and concepts early on and explicitly. In addition, the study showed that children who could identify the cover page of a book (showing familiarity with print) had stronger reading comprehension scores. These results highlight the need for strategies to maximize access to books especially for vulnerable populations that do not possess these resources in their homes, and to make better use of resources of existing school libraries.

Ureta et al. (2006) analyzed secondary data from literacy and mathematics tests, interviews, and questionnaires for 19,530 first and 16,770 third-grade children in Guatemala. The main finding of this study was that children's literacy and mathematics scores varied across demographic characteristics. The authors found that urban students outperformed rural students, nonindigenous students outperformed indigenous students, and boys slightly outperformed girls but not enough to be statistically significant. They also found that the infrastructure and pedagogical support that a community provided was associated with significantly higher test scores. For first-grade children in bilingual schools (Mayan language plus Spanish), their Spanish scores were higher, on average, than children attending Spanish-only schools, but this did not remain true for third-grade children. Items that students struggled most with, both on literacy and mathematics components, were items that required higher order thinking skills. Finally, the authors found a positive relationship between having breakfast prior to school and improved performance on literacy and mathematics in both first and third grades. This last finding is consistent with the finding from the quantitative intervention research that nutrition programs could positively affect EGL outcomes in countries with high rates of stunting and wasting, such as Guatemala.

Barbosa et al. (2016) investigated the relationship between spelling levels, phonological awareness, and letter knowledge in 60 first-grade children in Brazil. The study found significant correlations between phonological awareness, letter knowledge, and spelling level; those who had better phonological awareness skills and letter-name knowledge at the beginning of the school year significantly improved their spelling skills by the end of the school year. This study suggests that phonological awareness instruction combined with letter knowledge might support the development of good spelling skills in subsequent school years but further research is needed as this study was based on a relatively small sample size.

Finally, Guimarães and Mota (2016) investigated the hypothesis that morphological awareness[10] was independently associated with reading when cognitive and phonological awareness controls were applied. The results showed that after controlling for age, cognitive abilities, and phonological awareness, derivational morphology[11] was still correlated with word reading in the sample of 114 Brazilian children although it did not significantly correlate with reading comprehension. These results indicated that morphological awareness could be relevant to reading acquisition in Brazilian Portuguese.

---

[10] Morphological awareness: The recognition, understanding, and use of word parts that carry significance.
[11] Morphology: The process of forming a new word from an existing word, often by adding a prefix or a suffix.

Synthesis of Literacy Skills Studies

Three studies examined the importance of phonological awareness as a predictor of other reading skills. Lopes-Silva et al. (2016) identified the importance of phonemic awareness as a predictor of both numeracy and literacy tasks, Barbosa et al. (2016) found that children who had better phonological awareness skills and letter-name knowledge at the beginning of the school year significantly improved their spelling skills by the end of the school year, and Mariángel and Jiménez (2016) showed that phonological and syntax comprehension evolve across time, even as children pass through elementary school and that these skills are still developing even in grade five.

Two studies examined exposure to print as a predictor of children's reading comprehension and found positive correlations. Strasser et al. (2016) indicated that exposure to print was an important factor in explaining both children's oral language and reading comprehension in first grade and Vergara et al. (2016) also showed that children familiar with print had stronger reading comprehension scores.

Several studies examined other factors that contributed to word reading and reading comprehension. Strasser et al. found that higher order cognitive skills and knowledge (such as monitoring one's own understanding and verbal reasoning) were as important, and in some cases more so, than decoding in explaining reading comprehension. Nobre and Salles (2016) suggested that both decoding and comprehension benefited from semantic processing and Guimarães and Mota (2016) showed that morphological awareness was significantly correlated with word reading in Brazilian Portuguese.

One study, Ureta et al. (2006) found that children's literacy and mathematics scores varied across demographic characteristics, such as gender and language spoken at home. The study also showed that students struggled most, both on literacy and mathematics components, with items that required higher order thinking skills. Finally, the authors found a positive relationship between having breakfast prior to school and improved performance on literacy and mathematics in both first and third grades.

One study, (Silva Pacheco et al., 2016) suggested that interlimb coordination could help predict academic performance.

Assessment Adaptation and/or Validation

Three studies focused on adapting or validating various literacy assessments (Vilhena et al., 2016; Cadavid-Ruiz et al., 2016; Helman et al., 2016).

The study by Vilhena et al. (2016) described the procedure for adapting Lobrot's L3 literacy test to a Brazilian cultural-linguistic context and how this adaptation resulted in a new test deemed the Teste de Leitura: Compreensão de Sentenças (TELCS) for students from ages 7 to 11. The TELCS was validated through another study (Vilhena & Pinheiro, 2015), which showed the instrument was reliable to assess the literacy ability of students up to fourth grade. The study provided detailed descriptions of the rigorous seven-step test adaptation process: (a) classification of the response alternatives of the L3 test, (b) adaptation of the original test (sentences and target words) into Brazilian Portuguese, (c) back translation (from Brazilian Portuguese to French), (d) adaptation of the distractors for the TELCS, (e) configuration of the Brazilian version, (f) a pilot study, and (g) validation and standardization of the final instrument.

Cadavid-Ruiz et al. (2016) focused on the validation of the ABCDeti, a computerized test for initial literacy in Colombian schoolchildren ages 5–8 years old. The study found that the ABCDeti was a valid and reliable test (comparable to the ENI, which was the instrument most often used to measure literacy in Colombia) to evaluate initial literacy competence in children in the early grades. In addition, the computerized format provided an alternative to the traditional test that also reduced application times and errors associated with its correction and rating.

Helman et al. (2016) examined the reliability of the Spanish Developmental Spelling Inventory (SDSI), a tool for assessing orthographic development in Spanish in Grades 1–5. Results from the study demonstrated that the SDSI could reliably measure the development of orthographic knowledge in Spanish; previously, it had been validated only for English. This tool could support teachers in identifying student progress on orthographic knowledge in Spanish to then tailor instruction to student's needs.

Disabilities

Kida, De Ávila, and Capellini (2016) assessed reading comprehension through a task of oral retelling after reading with children who had dyslexia as well as those who had a language-based learning disability (LBLD) and their peers. The reading comprehension assessment indicated that children who had dyslexia or an LBLD showed difficulty in making sense of an expository text read to them. However, students who had dyslexia or an LBLD presented varying levels of text processing and different coverage of reading comprehension deficits. Children who had dyslexia could use their linguistic competence and background knowledge to minimize the effects of their decoding deficit, whereas children with an LBLD showed broader difficulties that impaired every level of text processing.

Smoke Exposure

Jorge et al. (2016) conducted a cross sectional study with 785 students from the second to the fifth year of elementary school in Brazil that demonstrated an association between passive exposure to tobacco smoke and learning difficulties at school. The findings also indicated that learning difficulties were strongly associated with the economic status of both children and households.

Teacher Training and Community Interventions

Lugo-Gil et al. (2016) presented the results of a baseline study of an RCT that evaluated a teacher training component and a community involvement component for EGL instruction in Guatemala with large linguistically diverse populations. The baseline report did not provide findings as to the impact of the programs but included reports on the baseline literacy assessment scores of each treatment and control arm. The final study results are expected to be published in 2018 and will be included in the next update to this systematic review.

Writing

Three studies focused on writing, including Rosal et al. (2016); dos Santos Santana, Oliveira, and Reis (2016); and Sargiani and Albuquerque (2016).

Rosal et al. (2016) investigated the correlation between phonological awareness and rapid serial naming of letters, numbers, objects, and colors and the initial learning of writing with 2–6-year-old children in Recife, Brazil. The findings showed significant correlations of phonological awareness skills with rapid serial naming and writing. Children in more advanced writing phases also had greater ease in phonological awareness skills. This study suggested that focusing on phonological awareness as early as possible could support later writing development.

Dos Santos Santana et al. (2016) evaluated how "monophthongization"[12] manifests itself in the textual writing production of children in Brazilian Portuguese. The authors found that the youngest children with the least schooling had the highest percentage of deletion of the diphthongs studied in written form. Also, children more frequently added the glides in their writing as they got older and had more years of schooling. These findings indicated that children learned more about the differences between oral and

---

[12] Monophthongization is a sound change by which a diphthong (a combination of two adjacent vowel sounds within the same syllable) becomes a monophthong (a pure vowel sound whose articulation is relatively fixed).

written language as they age and continue in their schooling and gained more exposure to the practice of reading and writing.

Sargiani and Albuquerque (2016) investigated the strategies that preschoolers use to write words in Brazilian Portuguese and developed explicit criteria for classifying children's writing. The authors stated that letter knowledge and phonemic awareness were the best predictors for word writing, although the correlation between phonemic awareness and children's writing was not statistically significant.

Qualitative Intervention and Nonintervention Research

Overall, the research team found that qualitative intervention and nonintervention studies followed similar patterns of quality ratings as in the original systematic review, although some individual examples of stronger articles provided high-quality explanations of study design, methods, and analysis approaches. The qualitative studies performed particularly low on measures of ethical practices, including examining the researcher/participant relationship and getting ethical approval through a committee, and all could also clarify better how their analysis led to the conclusions they presented. Two qualitative nonintervention research studies (de Morais, de Albuquerque, & Brandão, 2016; and Wajskop, 2016) received consistently low ratings on the quality dimensions; as a result, these studies were not included in the content review summary. The following subsections present the findings for the quality of articles on research design, ethics and reflexivity, and relevance to the field.

Research Design

Statement of Research

A clear statement of purpose forms the basis for how a researcher decides on methods, measurement, and analysis of a problem (Ford, 2009). Our review assumed the purpose of the research (i.e., problem statement) "may be phrased as statements of research purpose, as specific research questions, or as research hypotheses, depending on the purpose of the study and selected design" (McMillan & Schumacher, 2001, p. 86). A research statement introduces the reader to the research, provides context, and creates a framework in which to report results that, in the end, guide the entire exercise (Bryman, 2007). The quality of the research statement was rated on two parameters: the goal of the research and why it was important. Table 7 shows the ratings for the qualitative articles on each measure.

Table 7. Quality Review Ratings on Statement of Research

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Statement of Research	Goal of Research	High	High	Low	High	High	High	Med.
	Importance of Research	High	Med.	Low	NM	High	High	Med.

Qualitative Interventions. Both of the studies included clear research goals and clearly explained the purpose of the research. Both studies also established the importance of their research for the field. Effective statements of importance not only explained why the research was necessary but also showed why the findings would be important within the research context as well as within the larger community of stakeholders.

Qualitative Noninterventions. Three of the five qualitative nonintervention studies clearly stated the goal of the research, which was consistent with the original systematic review that showed most studies had clearly communicated the goal of the research. Two studies also had a high rating for effectively communicating the importance of the stated goal, whereas the three remaining articles received ratings of medium, low, and not mentioned. The highly rated studies commonly stated the specifics of the population as well as the literacy skill being tested. The study that received low ratings in this dimension (de Morais et al., 2016) said the goal of the research was to "discuss" rather than demonstrate, but the statement of importance was not backed up with literature or elaborated beyond stating that the authors' stance differed from other literature.

Methodology

Clear explanations of how research was conducted and what methods were used affects the way the reader will interpret the findings derived from that research. Strong research methodologies are guided by clearly presented research questions or hypotheses, an explanation of the research methods, and a justification for why particular methods are used. The research team assessed the quality of the papers' methodologies to the extent they were described according to the following criteria:

- **Appropriateness of Qualitative Methodology.** Did the research interpret or illuminate the actions and/or subjective experiences of the research participants?

- **Relationship of the Research Design to the Aims of the Research.**
    - Was the research guided by research questions or hypotheses?
    - Did the researcher justify the research design (i.e., did they discuss how they decided which methods to use)?

Table 8 shows the ratings for the qualitative articles on each measure.

Table 8. Quality Review Ratings on Methodology

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Methodology	Appropriateness of Methodology	High	Med.	NM	Med.	High	High	Low
	Guided by Research Questions or Hypotheses	High	Med.	NM	High	High	High	Med.
	Justification of Research Design	High	Low	NM	NM	High	High	High

Qualitative Interventions. The Moodie-Reid (2016) study scored consistently high for all categories. The study included clear research questions and adequately justified the research design and discussed how they decided which methods to use. The Apén-Son (2016) study scored medium except for the justification of research design which received a low rating because it identified the research approach but did not explain why that approach was chosen to answer the research questions.

Qualitative Noninterventions. Two of the five qualitative nonintervention studies scored high on appropriateness of methodology, which was consistent with the original systematic review findings that showed approximately half of the studies had used an appropriate methodology to answer the research questions. Only one study (de Morais et al., 2016) did not include clear research questions or a hypothesis to guide the study and due to its consistently low ratings, its results were not included in the content review. Finally, the researchers sufficiently justified the research design in three of the five studies.

Data

Describing methodologies also entails detailing the setting, justification, process, and form of the data collected. This information is particularly important to include as data in qualitative research, where context can determine participants' level of comfort with participation, especially in cases where participant observation is part of the methodology. In addition, a discussion of saturation—the point at which researchers gain no new information from an additional data point—typically serves as the justification for a study's numerical sample. Reviewers accounted for the following elements when rating a study on data quality:

- **Were the data collected in a way that addressed the research issue?**
    - Was the setting for data collection justified?
    - Was it clear how data were collected (e.g., focus group, interview)?
    - Did the researcher justify the methods chosen?
    - If the researcher made the methods explicit (e.g., for interview method), was there an indication of how interviews were conducted? Did the researcher use a topic guide?
    - If methods were modified during the study, did the researcher explain how and why?
    - Was the form of data clear (e.g., tape recordings, video material, notes)?
    - Did the researcher discuss saturation of data?

Table 9 shows the ratings for the qualitative articles on each measure.

Table 9. Quality Review Ratings on Data

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Data	Setting for Data Collection Justified	High	NM	NM	NA	High	High	NM
	Clear How Data Were Collected	Med.	High	Med.	High	High	High	High
	Justification of Methods?	High	NM	NM	NM	Low	High	NM
	Methods Made Explicit	Med.	Med.	NM	NM	NM	High	Med.
	Methods Modified and Explained	NM	NM	NM	NA	High	NA	NM
	Form of Data Are Clear	Med.	High	Med.	High	High	High	Low
	Data Saturation Discussed	NM	NM	NM	NA	NM	NM	NM

Qualitative Interventions. The Moodie-Reid (2016) study provided a strong justification for the data collection site, whereas the Apén-Son (2016) study did not justify the choice. Apén-Son (2016) rated high for clarity on how the data were collected while Moodie-Reid (2016) was rated medium for listing the data collection protocols but not fully indicating how they were used. Only one study attempted to justify the data collection methods used for the study, and the other did not. Both studies attempted to make the methods explicit, but the Moodie-Reid (2016) study was much clearer than the other in explaining how each instrument was applied. The Apén-Son (2016) study only mentioned the data collection instruments and the results without giving any indication as to how those instruments were applied. Finally, neither of the studies included a discussion of data saturation.

Qualitative Noninterventions. Only two of the five studies described the site of the data collection, which was a slightly lower ratio than that in the original systematic review. In contrast to the original systematic review, three studies did not include a discussion of the explicit methods. Three studies received a high rating for clarity on the type and form of data collected; however, only two studies justified the chosen method—one received a high rating for doing so (Orellana-Garcia & Sailors, 2016) and another received a low rating (Newman et al., 2016) as justifications had to be inferred. Only one study discussed having modified the original methods. Finally, as with the original systematic review, none of the studies discussed data saturation.

Data Analysis

Qualitative data analysis is by nature not as systematic as quantitative data analysis, but it can become systematic through methodical coding with an iterative process that promotes consistency in all facets of data collection, analysis, and reporting. The research team reviewed the quality of qualitative data analysis for the included articles on the following criteria:

- **Was the data analysis sufficiently rigorous?**
  - Was the analysis process thoroughly described?
  - If thematic analysis was used, was it clear how themes were derived from the data?
  - Did the researcher explain how the data presented were selected from the original sample to demonstrate the analysis process?
  - Were sufficient data presented to support the findings?
  - Were any contradictory data considered?
  - Did the researcher critically examined his or her own role, potential bias, and influence during analysis and selection of data for presentation?
  - Did the researcher consider contextual factors that may have influenced the results?

Table 10 shows the ratings for the qualitative articles on each measure.

Table 10. Quality Review Ratings on Data Analysis

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Data Analysis	**Analysis Process Thoroughly Described**	High	NM	NM	Low	High	High	Low
	Thematic Analysis Clear	High	NM	NM	NM	Med.	Med.	Low
	Explanation of Presented Data	High	NM	NM	NM	Low	High	Low
	Sufficient Data to Support Findings	High	Low	NM	High	Med.	High	Low
	Contradictory Data Considered	Med.	NM	NM	Med.	Med.	NM	NM
	Critical Examination of Researcher's Role	High	NM	NM	NM	NM	NM	Low
	Consideration of Contextual Factors	Low	Med.	NM	High	Med.	Med.	Med.

Qualitative Interventions. The Moodie-Reid (2016) study scored high for having a thorough description of the analysis process and for presenting sufficient data to support the findings. The Apén Son (2016) article did provide a description of the data analysis process, nor did it present sufficient data to support the findings. Moodie-Reid (2016) was the only study that provided a high-quality description of its use of thematic analysis and presented data to demonstrate the analysis process. Overall, the studies did not score very high in the data analysis quality review category because they did not consider contextual factors that could influence their results, only one examined potential researcher bias, and only one explained how presented data were selected from the body of available data. Beyond the process of describing the analysis, the researchers did not provide many additional details about factors that may have influenced the data analysis process.

Qualitative Noninterventions. The de Morais et al. (2016) and Wajskop (2016) studies received consistently low ratings for data analysis measures due to low quality descriptions lacking the necessary details or because they did not mention the data analysis procedures at all in the study. These two studies were not included in the final content synthesis. The remaining three studies mostly scored in the medium to high range for most measures. One area lacking across all these studies was any mention of the researcher's own role, potential bias, or influence during analysis and selection of data for presentation.

Statement of Findings

A primary goal of research is to translate data into accessible findings and practice among nonresearcher audiences—especially in the case of the LAC Reads Capacity Program. Findings should explicitly relate back to the purpose of the study and directly answer the research questions, while also phrasing them in a way that makes the level of reliability and transferability explicit. The research team rated articles' statements of findings on the following parameters:

- **Was there a clear statement of findings?**
    - Were the findings explicit?

- Was there adequate discussion of the evidence both for and against the researcher's interpretations?

- Did the researcher discuss the credibility of his or her findings (e.g., triangulation, respondent validation, more than one analyst)?

- Were the findings discussed in relation to the original research questions?

Table 11 shows the ratings for the qualitative articles on each measure.

Table 11. Quality Review Ratings on Statement of Findings

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Methodology	**Findings Made Explicit**	High	High	Low	High	High	High	Low
	Adequate Discussion of All Evidence	NM	NM	NM	NM	NM	High	Low
	Discussion of Credibility of Findings	High	NM	NM	NM	Med.	Med.	NM
	Findings Related to Research Questions	High	NM	NA	Med.	Med.	Med.	High

Qualitative Interventions. All the selected papers clearly presented findings, but only the Moodie-Reid (2016) study showed the relationship of the findings to the original research questions. None of the studies discussed any evidence for or against the researchers' own interpretations. Moodie-Reid (2016) was also the only study to discuss the credibility of their findings as well as the use of triangulation, member checking of the results and interpretations, and other study limitations.

Qualitative Noninterventions. Three studies received a high rating for explicitly discussing the findings, and two received a low rating. Only two studies discussed evidence both for and against the findings, and only two discussed the credibility of the findings. Four of the five studies discussed findings in relation to the original research questions.

Ethics and Reflexivity

Reviewers assessed the quality of an article's transparency on ethics based on its described recruitment strategy, its recognition of potential bias in the researcher-participant relationship, and its attention to the protection of human subjects in research.

Recruitment Strategy

Researchers should explain the criteria they used for participant recruitment and how they selected and excluded potential participants. Recruitment should have explained why the informants were best suited to inform the research using explicitly stated criteria. In cases where the intervention research involved selecting some potential participants over others, the researcher should have explained why the selected participants were appropriate over others who may have benefitted from the intervention. The research team rated articles' statements of findings on the following parameters related to appropriateness:

- Did the researcher explain how the participants were selected?

- Did the researcher explain why the participants selected were the most appropriate to provide access to the type of knowledge sought by the study?

Table 12 shows the ratings for the qualitative articles on each measure.

Table 12. Quality Review Ratings on Recruitment Strategy

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Recruitment Strategy	Explanation of Participant Selection	High	Low	NM	Med.	Med.	High	High
	Explanation of Why Selected Participants are Most Appropriate	High	NM	NM	NM	Med.	High	Med.

Qualitative Interventions. The Moodie-Reid (2016) study sufficiently described the sample procedure to select the target study population and why researchers selected certain participants over others. The Apén-Son (2016) study included limited information on recruitment strategies and did not mention why selected participants were most appropriate.

Qualitative Noninterventions. All studies except for de Morais et al. (2016) explained the recruitment strategy, and three studies explained why selected participants were the most appropriate for the study.

Researcher-Participant Relationship

Researchers' positionality may affect the formulation of research questions or the interpretation of data. This might not necessarily affect the research, but outlining and recognizing the potential for the researcher to influence the process reflects an attempt at neutrality. The research team evaluated the assessment of researcher-participant bias using the following criteria:

- **Was the relationship between the researcher and participants adequately considered?**
  - Did the researcher critically examined his or her own role, potential bias, and influence during
    - formulation of the research questions and research instruments (e.g., asking leading questions)?
    - data collection, including sample recruitment and choice of location?

Table 13 shows the ratings for the qualitative articles on each measure.

Table 13. Quality Review Ratings on Researcher-Participant Relationship

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Researcher-Participant Relationship	Examination of Bias in Formulation of Research Questions and Instruments	High	NM	NM	NM	NM	NM	Low
	Examination of Bias in Data Collection	High	NM	NM	NM	NM	High	Med

Qualitative Interventions. Only the Moodie-Reid (2016) study included a discussion of subjectivity and positionality in the formulation of research questions and described potential researcher bias in data collection.

Qualitative Noninterventions. None of the studies discussed the researcher-participant relationship, and only two studies (Orellana-Garcia & Sailors, 2016; Wajskop, 2016) discussed potential bias in data collection. None of the authors discussed potential researcher bias in data analysis. A critical discussion would have been particularly useful in Wajskop (2016), a study for which the implementers also participated in the research.

Ethics

Ethics review committees review research to ensure the protection of human subjects. Ethical committees (sometimes known as Institutional Review Boards) standardize the requirements that "fully inform and protect" participants and serve as an "affirmation of our commitment to treat all people with respect" (Patton, 2015, p. 341). Although no overarching ethical review board covered the entire LAC region, individual countries, institutions, universities, and publications have their own ethical review boards and ethics codes with similar standards that researchers should follow. As a standard protection for human subjects, the Critical Appraisal Skills Programme Qualitative Research Checklist recommends assessing ethics on the following dimensions:

- **Were ethical issues considered?**
  - Were sufficient details included of how the research was explained to participants for the reader to assess whether ethical standards were maintained?
  - Did the researcher discuss issues raised by the study on sensitive issues?
  - Was approval sought from an ethics committee?

Table 14 shows the ratings for the qualitative articles on each measure.

Table 14. Quality Review Ratings on Ethical Issues

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Ethical Issues	Sufficient Details of How Research was Explained	High	NM	NM	NM	NM	NM	NM
	Discussion of Sensitive Issues	NA	NM	NM	NM	NM	NM	NM
	Ethics Committee Approval Obtained	High	NM	NM	NM	High	NM	NM

Qualitative Interventions. Only the Moodie-Reid (2016) study was approved by an ethics committee and obtained informed consent from participants. They also provided details on how the study was explained to participants and how consent was obtained. The other study did not mention any ethical considerations taken. In addition, neither of the studies discussed the potential sensitivity of issues raised, but they may have deemed a discussion not applicable due to the type of data collected.

Qualitative Noninterventions. One of the five studies (Newman, Arthur, Staples, & Woodrow, 2016) mentioned being approved by an ethics committee. No study described the process of obtaining consent from parents to engage with students or how they obtained consent from students during the data collection process.

Relevance to the Field

Finally, raters reviewed qualitative intervention and nonintervention studies on two dimensions that described how the study was relevant to the field. First, the study should have added new information to the existing body of literature and policy documents. Second, the research should have discussed the potential for study replicability and how replicability could add to the current findings—that is, how the study could be improved or adapted to further gain insight on the topic of research.

Situating the Research

Situating research within the existing body of literature entails discussing existing knowledge, understanding, and views on the topic. Researchers should describe how the research question adds to what knowledge already exists and why the study is necessary. The research team assessed this discussion based on the following criteria:

- **How valuable was the research?**
  - Did the researcher discuss the contribution the study made to existing knowledge or understanding?
  - Did the researcher identify new areas where research was necessary?
  - Did the researcher discuss whether or how the findings could be transferred to other populations or consider other ways the research could be used?

Table 15 shows the ratings for the qualitative articles on each measure.

Table 15. Quality Review Ratings on Value of Research

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Value of Research	Discussion of Contribution of the Study to Existing Knowledge	High	High	NM	High	High	High	Med.
	Identification of New Areas of Research	Med.	NM	NM	Med.	High	High	High
	Discussion of Transfer of Findings	Low	NM	NM	Low	Med.	High	NA

Qualitative Interventions. Both studies discussed their contribution to existing knowledge in the field, but only Moodie-Reid (2016) identified new areas of research, and discussed how the findings could be used in different contexts.

Qualitative Noninterventions. All studies except for de Morais et al. (2016) discussed their contribution to existing literature and identified areas for further research. Three studies discussed how findings could be used in different contexts.

Replicability

Replicability was assessed on two dimensions: first, whether stakeholders could replicate the program; and second, whether researchers provided sufficient information for other researchers to replicate the study in different contexts. The research team assessed replicability based on the following criteria:

- **Information for stakeholders to assess replicability**
    - Does the paper provide adequate details on the design and implementation of the intervention to enable replication, such as: Length of training, Monitoring tools, and Training materials?

Table 16 shows the ratings for the qualitative articles on each measure.

Table 16. Quality Review Ratings on Replicability

Quality Review Categories		Qualitative Intervention Studies		Qualitative Nonintervention Studies				
		Moodie-Reid, 2016	Apén-Son, 2016	de Morais et al., 2016	Jennings, 2016	Newman et al., 2016	Orellana-Garcia & Sailors, 2016	Wajskop, 2016
Replicability	Information for Stakeholders to Assess Replicability	NM	Med.	NM	NA	NA	NA	Low

Qualitative Interventions. Only the Apén-Son study received a medium rating on its description of the study and intervention, and the remaining study, (Moodie-Reid, 2016), did not provide a sufficient description of the program so that it could be replicated; however, the article did provide detail on the study design, which could provide the details needed to replicate the approach in another setting on a similar program.

Qualitative Noninterventions. Only one of the five studies discussed potential replicability for the observed practices. Replicability was not applicable for three of the studies.

Synthesis of Qualitative Studies

In this section, the results of studies rated as high and medium quality by topic area are presented. The synthesis includes both qualitative intervention and nonintervention research and highlights the main findings across the studies within topic areas. Findings from three of the five qualitative nonintervention studies with medium to high ratings are discussed here (Jennings, 2016; Newman et al., 2016; & Orellana-Garcia & Sailors, 2016) as well as both of the qualitative intervention studies (Apén Son, 2016; Moodie-Reid, 2016). Studies focused on the following topic areas: phonics, parent and community participation, motivation to read, curriculum, and the print environment. Each of the topics was only covered in one study, so in these categories we describe the findings from that study alone instead of a synthesis.

Phonics

A dissertation published by Moodie-Reid (2016) reviewed teachers' perceptions of the impact of the Jolly Phonics Program on Grades 1–3 children's literacy in two primary schools in Western Jamaica. For this qualitative intervention study, data were collected in the form of interviews with eight teachers from two selected primary schools, with at least two teachers from each grade level. Because the sample is small and from only two schools in one region of Jamaica, the study was not representative and cannot be generalized. Study teachers perceived the Jolly Phonics Program as effective in teaching literacy and facilitated children's participation and enjoyment of the learning process; however, the study did not provide any evidence of actual impacts of the intervention on literacy skills.

Parent and Community Participation

Newman et al.'s (2016) qualitative nonintervention study examined how low-income families perceived their contribution to children's literacy learning in the home learning environment, as well as the impact of teachers' pedagogical documentation on families' engagement with and connection to their children's literacy learning. The researchers studied 15 schools within the *Futuro Infantil Hoy* program in Antofagasta and Calama in northern Chile and engaged the teachers in iterative cycles of action research focused on literacy learning pedagogies, with attention to developing learning partnerships with families.

The researchers found that parents from low socioeconomic backgrounds, who often were presumed to be providing inadequate language and literacy experiences for their children, were interested in supporting their children's education. The researchers argued that policymakers, researchers, and educators should strengthen relationships with families with the goal of benefitting children's literacy learning, increasing chances of school success, and contributing to social change as desired by families. Instead of seeing these parents as "receivers" of information, they should be viewed as sources of "giving" information, and their role in their children's literacy development should be valued. In summary, "the data demonstrates the potential of sociocultural literacy pedagogies to form a basis for literacy learning partnerships which are currently under-utilized by educators working with poor and/or socially excluded families" (Newman et al., 2016, p. 73).

Motivation to Read

Apén Son (2016) conducted qualitative intervention research examining the motivation of children to engage in literacy activities. The study focused on 23 four-year-old children and eight preschool teachers in Guatemala. The author developed a guide for teachers to use with the children to help motivate and develop a love of reading. Apén Son conducted a pre- and postdiagnosis to collect data on students' level of motivation as well as the knowledge of their teachers on the subject. Study findings recommended that teachers and parents should promote activities that motivate children to love reading at home and in school and that teachers need more innovative methods and strategies to motivate children to enjoy reading. However, Apén Son did not provide a thorough description of the analysis process. This finding coupled with the small sample size meant that the results were not generalizable and need further validation.

Curriculum

Jennings (2016) conducted qualitative nonintervention research consisting of a review of research studies conducted between 2004 and 2012 on the implementation of the Literacy 1–2–3 (L1–2–3) primary literacy curriculum in Jamaica. The review included nine studies either by Jennings or graduate students supervised by Jennings. L1–2–3 is a constructivist teaching/learning literacy methodology that replaced the former early grades literacy curriculum in the early 2000s.

The findings from this review indicated that the L1–2–3 curriculum was not implemented with fidelity, partly because of resource constraints, inadequate training of the teachers and principals, and unilateral decision making in the top-down approach to curriculum change. The author also emphasized the importance of consistent training for teachers to help them move from the traditional "chalk and talk" and choral recitation method to a more constructivist approach as mandated in the L1–2–3. "In a nutshell, the studies reviewed in this paper underscore the fact that top-down approaches to curriculum change inevitably lead to what Sarason (1999) described as 'self-defeating processes of implementation'" (Jennings, 2016, p. 55).

Print Environment

Orellana-Garcia and Sailors (2016) conducted a descriptive case study (using cross-case analysis) to examine classroom text environments in Chile in relation to literacy practices. The authors conducted the case study in four classrooms within two private schools in a large urban area in Chile. The participating teachers were selected based on their reputation for successful literacy instruction. The findings from the study indicated that classroom texts were underused and underappreciated by both teachers and students. The authors posited that this was caused by a lack of awareness about the role of nontextbook reading materials in literacy development. The authors also suggested that these findings pointed to a need to focus teacher professional development on the role of the text environment in raising literacy achievement.

Synthesis Across Research Types

In this section, the research team seeks to tie together the key findings from each research type as well as the research identified in the original systematic review to present a coherent picture of the EGL evidence emerging from the LAC region and identify gaps in that evidence. LAC is a large region composed of more than 40 countries with multiple languages, diverse populations, and unique education systems. Thus, the findings presented here are not necessarily generalizable to the entire LAC region.

Similarities to Previous Findings

The 2016 update to the original systematic review found not only many similarities in terms of characteristics of the included studies as well as findings but also some important differences. In addition, the synthesis suggests that researchers have started addressing some of the gaps identified in the original

systematic review. Nonetheless, the research continues to focus primarily on South America and most studies continue to be published in English. The majority of the studies included in this update (as also in the original systematic review) were published journal articles, and the majority were from high-income countries in South America (primarily Brazil and Chile). Central America and the Caribbean continue to be underrepresented in the production of EGL studies. English continues to be the main language of publication, followed by Spanish and then Portuguese. No articles were published in any of the indigenous languages of the region, just like in the original systematic review. The type of research produced was almost the same as in the original systematic review. Most of the research was quantitative nonintervention, followed by quantitative intervention, qualitative nonintervention, and qualitative intervention. Finally, most of the research was conducted in high-income to upper-middle-income countries. In addition, most of the research conducted continues to have a very small and limited sample size, making it difficult to generalize the findings across the country or region.

The new evidence produced in 2016 builds on many of the content-related findings in the original systematic review. For example, one study identified in this summary (Jorge et al., 2016) supported the finding from the original systematic review that socioeconomic factors were important predictors of EGL outcomes. Another finding from the original systematic review was about considering student input in the learning process, and Apén Son (2016) identified something similar in his study with preschool children in Guatemala. His study suggested that teachers and parents should use more innovative methods and strategies to motivate children to enjoy reading, including engaging them more in the process.

One research gap identified in the original systematic review was a dearth of research on literacy in indigenous languages. The 2016 update also found very little research focusing on this topic, although one study that analyzed secondary data in Guatemala (Ureta et al., 2006) found that nonindigenous students outperformed indigenous students on mathematics and literacy in Grades 1 and 3. The study also found that for first-grade children in bilingual schools (Mayan language plus Spanish), their Spanish scores were higher, on average, than children attending Spanish-only schools, but this did not remain true for third-grade children. Lugo-Gil et al. (2016) also presented the results of a baseline study of an RCT that evaluated a bilingual EGL program in Guatemala with large linguistically diverse populations. The full study is expected to be published in 2018 and will be included in future updates to this systematic review.

In terms of quantitative intervention studies, the LRCP team still found only limited mixed-methods evidence of the effects of specific programs on EGL outcomes. There remains an important evidence gap on the effects on EGL outcomes of preschool programs, school governance programs, teacher practices, and parental involvement. In addition, more mixed-methods evidence is needed regarding how ICT programs can positively affect EGL outcomes.

Nonetheless, the evidence suggests that teacher training programs can be effective in improving EGL outcomes when they are complemented by coaching, and nutrition programs can be effective in improving EGL outcomes when they target geographies with high rates of stunting and wasting. Importantly, however, the large majority of the evidence on the impact of specific programs on EGL outcomes continues to be produced in high-income or higher middle–income countries. There remains an important evidence gap on the impact of specific development programs on EGL outcomes in low-income and lower middle–income countries.

Gaps Addressed

The 2016 update did include some evidence that started addressing gaps identified during the original systematic review related to the topics of EGL for children with disabilities, prewriting and writing, and reading comprehension development.

One gap identified in the original systematic review was a lack of research on EGL for children with disabilities. Between 1995 and 2015, no articles were found related to EGL and disabilities from the LAC region. This systematic review update identified one quantitative nonintervention research study (Kida et al., 2016) that focused on children with an LBLD and children with dyslexia. The addition of one new study in 2016 is an important step forward in addressing the evidence-gap on EGL for children with disabilities.

Another gap identified during the original systematic review related to the development of prewriting and writing skills and the relationship to literacy outcomes. In the 2016 update, three studies were identified that focused on writing (dos Santos Santana et al., 2016; Rosal et al., 2016; Sargiani & Albuquerque, 2016). Each study focused on different aspects of writing, although two examined the relationship between phonological awareness and writing.

In the original systematic review, there was a clear lack of studies focused on reading comprehension. This was interesting given the fact that comprehension is the ultimate goal of literacy and is something that students in the LAC region struggle to master, as evidenced by scores on national literacy assessments. In the 2016 update, more studies were found that focused on reading comprehension, particularly studies that looked at the role of various literacy skills as predictors of reading comprehension (Guimarães & Mota, 2016; Nobre & Salles, 2016; Strasser et al., 2016; Vergara et al., 2016).

Finally, newly included quantitative intervention research shows that large improvements in teacher practices are not necessarily sufficient for any measurable improvements in EGL outcomes. Research from Chile demonstrates that programs may not have measurable positive effects on EGL outcomes even if they show large and positive effects on teacher practices. More research from other contexts will be needed to further examine how much teacher practices need to improve to produce any measurable positive effects on EGL outcomes.

Appendix A. References

Adrogue, C., & Orlicki, M. E. (2013). Do in-school feeding programs have an impact on academic performance and dropouts? The case of public schools in Argentina. *Education Policy Analysis Archives, 21,* 50.

Apén Son, J. A. (2016). Motivación a la lectura para niños de cuatro años, del nivel preprimario a través de grupos cooperativos. *Universidad Rafael Landívar Facultad de Humanidades Licenciatura en Educacion Inicial y Preprimaria,* 1–130.

Arbour, M., Yoshikawa, H., Willett, J., Weiland, C., Snow, C., Mendive, S., . . . & Treviño, E. (2016). Experimental impacts of a preschool intervention in Chile on children's language outcomes: Moderation by student absenteeism. *Journal of Research on Educational Effectiveness, 9*(suppl 1), 117–149.

Bandini, H. H. M. (2003). Um programa para a promoção de consciência fonológica em pré-escolares aplicado em sala de aula (Dissertação Mestrado em Educação Especial). Universidade Federal de São Carlos, São Carlos, Brazil.

Bandini, H. H. M., Oliveira, C. L. A. C., & Souza, E. C. (2006). Habilidades de leitura de pré-escolares deficientes auditivos: Letramento Emergente. *Paidéia, 16*(33), 51–58.

Bando, R. (2010). *The effect of school based management on parent behavior and the quality of education in Mexico* (Dissertation). University of California–Berkeley. Retrieved from http://digitalassets.lib.berkeley.edu/etd/ucb/text/Bando_berkeley_0028E_10483.pdf

Barbosa, M. R., Medeiras, L. B. D. O., & Vale, A. P. S. D. (2016). Relação entre os níveis de escrita, consciência fonológica e conhecimento de letras. *Estud. psicol. (Campinas), 33*(4), 667–676.

Barby, A. A. de O. M., & Guimarães, S. R. K. (2016). Desenvolvimento de Habilidades Metafonológicas e Aprendizagem da Leitura e da Escrita em Alunos com Síndrome de Down. *Revista Brasileira de Educação Especial, 22*(3), 381–398.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., & Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in Peru. *American Economic Journal: Applied Economics, 7*(2), 53–80.

Bhattacharyya, S. (2004, November 24–26). If you build it they will come: Creating a field for project based approach in learning science through qualitative inquiry and technology integration. In *Proceedings of QualIT2004: The Way Forward,* Griffith University, Brisbane, Australia.

Borenstein, M. H., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Publication bias. In: Introduction to meta-analysis (pp. 277–294). Chichester, England: Wiley.

Bowne, J. B. (2014). *The vocabulary instruction of Chilean kindergarten teachers: Impacts of* Un Buen Comienzo *and relationships with children's vocabulary development.* Cambridge, MA: Harvard University.

Bowne, J. B., Yoshikawa, H., & Snow, C. E. (2016). Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum. *Early Childhood Research Quarterly, 34,* 27–39.

Bryman, A. (2007). Effective leadership in higher education: A literature review. *Studies in Higher Education, 32*(6), 693–710.

Burgoyne, K., Duff, F. J., Clarke, P. J., Snowling, M. J. & Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*(10), 1044–1053.

Cadavid-Ruiz, N., Quijano-Martínez, M. C., Escobar, P., Rosas, R., & Tenorio, M. (2016). Validación de una prueba computarizada de lectura inicial en niños escolares colombianos. *Ocnos: Revista de estudios sobre lectura*, *15*(2), 98–109.

Caldeira, M. M., & Ward, J. M. (2003). Using resource-based theory to interpret the successful adoption and use of information systems and technology in manufacturing small and medium-sized enterprises. *European Journal of Information Systems, 12*(2), 127–141.

Campos, M. M., Bhering, E. B., Esposito, Y., Gimenes, N., Abuchaim, B., Valle, R., . . . Unbehaum, S. (2011). The contribution of quality early childhood education and its impacts on the beginning of fundamental education. *Educação e Pesquisa, 37*(1), 15–33.

Capovilla, A. G. S. (1999). Leitura, escrita e consciência fonológica: Desenvolvimento, intercorrelações e intervenções (Tese Doutorado em Psicologia Experimental). Instituto de Psicologia Universidade de São Paulo, São Paulo, Brazil.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., . . . Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23,* 678–686. https:/doi.org/10.1177/0956797611434536.

Caravolas, M., Lervåg, A., Defior, S., Málková, G.S. & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science, 24,* 1398–1407. https://doi.org/10.1177/0956797612473122.

Cardoso-Martins, C., Mesquita, T. C. L., & Ehri, L. (2011). Letter names and phonological awareness help children to learn letter–sound relations. *Journal of Experimental Child Psychology, 109*(1), 25–38.

Carvalho, L. M. de M. (2010). Consciência fonológica e sucesso na aprendizagem da leitura e da escrita: melhor prevenir do que remediar (Tese Doutorado em Psicologia). Pontifícia Universidade Católica de São Paulo, São Paulo, Brazil.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2012). *Technology and child development: Evidence from the one laptop per child program* (Discussion Paper No. 6401). Bonn, Germany: Institute for the Study of Labor.

de Morais, A. G., de Albuquerque, E. B. C., & Brandão, A. C. P. A. (2016). Refletindo sobre a língua escrita e sobre sua notação no final da educação infantil. *Revista Brasileira de Estudos Pedagógicos, 97*(247).

dos Santos Santana, J. H., de Oliveira, I. B., & dos Reis, M. S. (2016). Monotongação e ensino: Quando a variação linguística chega à escrita. *Entrepalavras, 5*(2), 65–85.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly. 36*, 250–287. doi:10.1598/RRQ.36.3.2

Felício, F. D., Terra, R., & Zoghbi, A. C. (2012). The effects of early childhood education on literacy scores using data from a new Brazilian assessment tool. *Estudos Econômicos* (São Paulo), *42*(1), 97–128.

Ferrando, M., Machado, A., Perazzo, I., & Vernengo, A. (2011). *Aprendiendo con las XO: El impacto del Plan Ceibal en el aprendizaje*. Serie Documentos de Trabajo/FCEA-IE; DT03/11.

Flores Mancilla, A. N. (2012). Propuesta de Evaluación de la Implementación del Programa "Estimulación Temprana a la Lectura" y el Componente de Nutrición, que beneficia a niños y niñas de 0 a 5 años en el municipio de Chiché, departamento del Quiché. *Universidad Rafael Landívar Facultad de Humanidades Licenciatura en Educacion Inicial y Preprimaria.*

Ford, E. D. (2009). The importance of a research data statement and how to develop one. *Annales Zoologici Fennici, 46*(2), 82–92.

Gomez Franco, L. E. (2014). *Exploring teachers' read-aloud practices as predictors of children's language skills: The case of low-income Chilean preschool classrooms* (Unpublished doctoral dissertation). Boston College, Lynch School of Education, Chestnut Hill, MA.

Guimarães, S. B., & Mota, M. M. P. E. D. (2016). Qual a contribuição da consciência morfológica das crianças na precisão de leitura de palavras e compreensão de texto no português? *Estudos de Psicologia (Natal), 21*(3), 239–248.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3." *American Educator, 27*(1), 4–9. Retrieved from www.aft.org/pdfs/americaneducator/spring2003/TheEarlyCatastrophe.pdf

Helman, L., Delbridge, A., Parker, D., Arnal, M., & Jara Mödinger, L. (2016). Measuring Spanish orthographic development in private, public and subsidised schools in Chile. *Assessment in Education: Principles, Policy & Practice, 23*(3), 327–352.

Hombrados, J. G., & Waddington, H. (2012). Internal validity in social experiments and quasiexperiments: An assessment tool for reviewers. London: International Initiative for Impact Evaluation.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160.

Ismail, S. J., Jarvis, E. A., & Borja-Vega, C. (2014). Guyana's hinterland community-based school feeding program (SFP). *Improving Diets and Nutrition,* 124.

Jennings, Z. (2016). A review of studies on the implementation of Literacy 1–2–3 in Jamaican primary schools. *Journal of Education and Development in the Caribbean, 13*(1–2), 84–109.

Jorge, J. G., Botelho, C., Silva, A. M. C., & Moi, G. P. (2016). Influence of passive smoking on learning in elementary school. *Jornal de Pediatria (Versão em Português), 92*(3), 260–267.

Kida, A. D. S., De Ávila, C. R., & Capellini, S. A. (2016). Reading comprehension assessment through retelling: Performance profiles of children with dyslexia and language-based learning disability. *Frontiers in Psychology, 7,* Article 787.

King, E., Samii, C., & Snilstveit, B. (2010). Interventions to promote social cohesion in subsaharan Africa. *Journal of Development Effectiveness, 2*(3), 336–370.

Larraín, A., Strasser, K., & Lissi, M. R. (2012). Lectura compartida de cuentos y aprendizaje de vocabulario en edad preescolar: un estudio de eficacia. *Estudios de Psicología, 33*(3), 379–383.

Lockheed, M., Harris, A., & Jayasundera, T. (2010). School improvement plans and student learning in Jamaica. *International Journal of Educational Development, 30*(1), 54–66.

Lopes-Silva, J. B., Moura, R., Júlio-Costa, A., Wood, G., Salles, J. F., & Haase, V. G. (2016). What is specific and what is shared between numbers and words? *Frontiers in Psychology, 7,* Article 22.

Lugo-Gil, J., Contreary, K., Murray, N., Glazerman, S., Campuzano, L., Fernandez, C., & Place, K. (2016). *Latin American and the Caribbean (Lac) reading evaluation contract evaluation of Leer Juntos, Aprender Juntos early grade intervention in Guatemala: Final baseline report.* Washington, DC: United States Agency for International Development.

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., & Stein, A. D. (2009). The impact of improving nutrition during early childhood on education among Guatemalan adults. *The Economic Journal, 119*(537), 734–763.

Mariángel, S. V., & Jiménez, J. E. (2016). Desarrollo de la conciencia sintáctica y fonológica en niños chilenos: un estudio transversal. *Revista Latinoamericana de Psicología, 48*(1), 1–7.

Mascareño, M., Snow, C. E., Deunk, M. I., & Bosker, R. J. (2016). Language complexity during read-alouds and kindergartners' vocabulary and symbolic understanding. *Journal of Applied Developmental Psychology, 44,* 39–51.

McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction.* Boston, MA: Little, Brown.

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology, 108*(1), 130.

Moodie-Reid, L. (2016). *Teachers' perceptions of the impact of the Jolly Phonics Program on students' literacy* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.

Murad, C. R., & Topping, K. J. (2000). Parents as reading tutors for first graders in Brazil. *School Psychology International, 21*(2), 152–171.

Neugebauer, S. R., & Currie-Rubin, R. (2009). Read-alouds in Calca, Peru: A bilingual indigenous context. *The Reading Teacher, 62*(5), 396–405.

Newman, L., Arthur, L., Staples, K., & Woodrow, C. (2016). Recognition of family engagement in young children's literacy learning. *Australasian Journal of Early Childhood, 41*(1), 73–81.

Nobre, A. D. P., & Salles, J. F. D. (2016). Lexical-semantic processing and reading: Relations between semantic priming, visual word recognition and reading comprehension. *Educational Psychology, 36*(4), 753–770.

Orellana-Garcia, P., & Sailors, M. (2016). Examining text environments in elementary Chilean classrooms. *Literacy Research and Instruction, 55*(2), 163–181.

Osorio, F. B., & Linden, L. L. (2009). *The use and misuse of computers in education: Evidence from a randomized experiment in Colombia* (Policy Research Working Paper 4836). Washington, DC: The World Bank Human Development Network Education Team.

Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *Journal of Applied Developmental Psychology, 28*(2), 85–102.

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.

Patashnick, J., & Rich, M. (2004). Researching human experience: Video intervention/prevention assessment (VIA). *Australian Journal of Information Systems, 12*(2), 103–111.

Patton, M. Q. (2015). Qualitative evaluation and research methods. Thousand Oaks, CA: SAGE.

Piedra Santa, M. (2004). Estimulación temprana y su relación en el desarrollo de la lectura y escritura en niños del nivel pre-primario (Tesis inedita). Universidad Rafael Landívar, Guatemala.

Powell, C. A., Walker, S. P., Chang, S. M., & Grantham-McGregor, S. M. (1998). Nutrition and education: A randomized trial of the effects of breakfast in rural primary school children. *The American Journal of Clinical Nutrition, 68*(4), 873–879.

Rosal, A. G. C., de Andrade Cordeiro, A. A., da Silva, A. C. F., Silva, R. L., & de Queiroga, B. A. M. (2016). Contribuições da consciência fonológica e nomeação seriada rápida para a aprendizagem inicial da escrita. *CEP, 50670,* 420.

Sargiani, R. D. A., & Albuquerque, A. (2016). Análise das estratégias de escrita de crianças pré-escolares em Português do Brasil. *Psicologia Escolar e Educacional, 20*(3), 591–600.

Sás, R.M. (2009). *Efeitos de um programa de remediação fonológica nas habilidades de leitura e escrita em alunos com Síndrome de Down* (Dissertação Mestrado em Educação Especial). Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, Brazil.

Silva Pacheco, S. C., Gabbard, C., Ries, L. G. K., & Bobbio, T. G. (2016). Interlimb coordination and academic performance in elementary school children. *Pediatrics International, 58*(10), 967–973.

Simeon, D. T., Grantham-McGregor, S. M., & Wong, M. S. (1995). Treatment of Trichuris trichiura infections improves growth, spelling scores and school attendance in some children. *Community and International Nutrition, 125*(7), 1875–1883.

Strasser, K., Vergara, D., & Río, M. F. (2016). Contributions of print exposure to first and second grade oral language and reading in Chile. *Journal of Research in Reading, 40*(S1). https://doi.org/10.1111/1467-9817.12086

Tapia, J. P. R., & Benítez, Y. G. (2013). Desarrollo de habilidades conductuales maternas para promover la alfabetización inicial en niños preescolares. *Acta Colombiana de Psicología, 16*(1), 81–90.

Ureta, F., Fortin, A., & Molina, R. (2006). Rendimiento en lectura y matemática de estudiantes de primero y tercero primaria. *Revista de la Universidad del Valle de Guatemala, 15,* 9–37.

Vergara, D., Strasser, K., & del Río, M. F. (2016). Más que palabras por minuto: Las otras habilidades que afectan la comprensión en 1º básico. *Calidad en la Educación,* (44), 46–67.

Vilhena, D. A., & Pinheiro, A. M. V. (2016). Validation and standardization of the Brazilian reading test – Sentence comprehension. *Dyslexia, 22*(1). https://doi.org/10.1002/dys.1521

Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, Â. M. V. (2016). Reading test—Sentence comprehension: An adapted version of Lobrot's Lecture 3 test for Brazilian Portuguese. *Dyslexia, 22,* 47–63.

Viramontes Anaya, E., Morales-Sifuentes, L. A., & Delgado-Nájera, M. D. J. (2016). La comprensión lectora: Una evaluación clínica y cuantitativa. *Ra Ximhai, 12*(6), 71-88.

Vivas, E. (1996). Effects of story reading on language. *Language Learning, 46*(2), 189–216.

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., . . . & Valentine, J. C. (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness, 4*(3), 359–387.

Wajskop, G. (2016). Are teaching practices suitable for classes of six-year-old children? In K. González & R. Frumkin (Eds.), *Handbook of Research on Effective Communication in Culturally Diverse Classrooms* (pp. 183–195). IGI Global.

World Bank, The. (n.d.). *World Bank country and lending groups*. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51*(3), 309.