



USAID
FROM THE AMERICAN PEOPLE



**EARLY GRADE READING IN LATIN
AMERICA AND THE CARIBBEAN:
A SYSTEMATIC REVIEW
EXECUTIVE SUMMARY**



DECEMBER, 2016

This is an overview of the Systematic Review on Early Grade Reading evidence in the LAC Region carried out by the LAC Reads Capacity Program (LRCP). This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the American Institutes for Research and do not necessarily reflect the views of USAID or the United States Government.

The 5-year regional USAID/LAC Reads Capacity Program (LRCP), Cooperative Agreement No.: AID-OAA-A-14-00058), implemented by AIR and Juárez and Associates and national-level implementing partners, increases the impact, scale, and sustainability of early grade reading interventions in the LAC region. This is achieved through the development and dissemination of state-of-the-art knowledge resources and the provision of technical assistance to host country governments and other key stakeholders. The program, implemented from 2014-2019, enhances efforts to boost early grade reading outcomes.

INTRODUCTION

Educational policy around early grade reading (EGR)¹ in the Latin American and Caribbean (LAC) region has faced a disjuncture between school practice and research. Studies exist in the global literature on how pedagogical programs should be designed to promote gains in EGR outcomes but it is unclear whether findings in other regions can be extrapolated to the LAC region. Also, within the LAC region itself, research on EGR is fragmented and often lacks quality. There is no comprehensive or systematic overview of the EGR research literature specific to the LAC region. As a result, policy makers, pedagogy and curriculum specialists, and other stakeholders in the region are unable to determine what is relevant and are thus unable to shape policy, practice and programs in an evidence-driven manner.

The systematic review prepared by the LAC Reads Capacity Program aims to assist policy makers, international funders, nongovernmental organizations (NGOs), practitioners, researchers, and other relevant stakeholders in the LAC region by synthesizing the evidence on what works to improve reading outcomes in the LAC region and by identifying areas where further research is needed. This document is a brief executive summary of the much more extensive and detailed full systematic review report; both are available at <http://www.lacreads.org>.

We address several research questions through the systematic review. First, we examine the effectiveness of various programs implemented in the LAC region that aim to improve EGR outcomes, including teacher training, school feeding, computer-aided instruction, programs with an emphasis on nutrition, and Information and Communication Technology (ICT) programs. Second, we assess the fidelity of implementation of programs that aim to improve reading outcomes. Third, we examine the factors that predict EGR outcomes. Fourth, we examine the experiences and perspectives of various stakeholders about early grade reading in the LAC region. For this purpose, we use a mixed-methods systematic review, in which we synthesize the evidence from both quantitative and qualitative research.

OPERATIONAL DEFINITIONS

This section provides the operational definitions of key terms used in this review.

Early grade reading is defined by USAID as pertaining to Grades 1–3 of primary schooling, and we broadened the definition to include children from birth to grade 3, as there is a large evidence base on the importance of developing early language skills, exposure to print, and prereading and writing activities for improving later reading success. Early grade reading refers to both reading and writing.

Evidence refers to a research or empirically derived body of facts that can be used to make informed decisions about education interventions (e.g., policies, practices, or programs).

Systematic review is a review of the evidence around a particular topic that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to extract and analyze data from the studies that are included in the review.

¹See operational definitions

Narrative synthesis is an “approach to the synthesis of evidence relevant to a wide range of questions including but not restricted to effectiveness [that] relies primarily on the use of words and text to summarize and explain—to ‘tell the story’—of the findings of multiple studies (Popay et al., 2006).”

Meta-analysis is “the statistical pooling of information on study effect sizes” (Waddington et al., 2012) to determine the impact of programs on specific outcomes. Meta-analysis enables researchers to estimate the average effect size of specific programs on early grade reading outcomes and to assess the variation in effect sizes across different contexts.

PICO criteria stands for population, intervention, comparison, and outcome. The Campbell Collaboration and Cochrane, the foremost research networks promoting best practices in systematic reviews worldwide, recommend using the PICO categories for formulating questions and search strategies for systematic reviews.

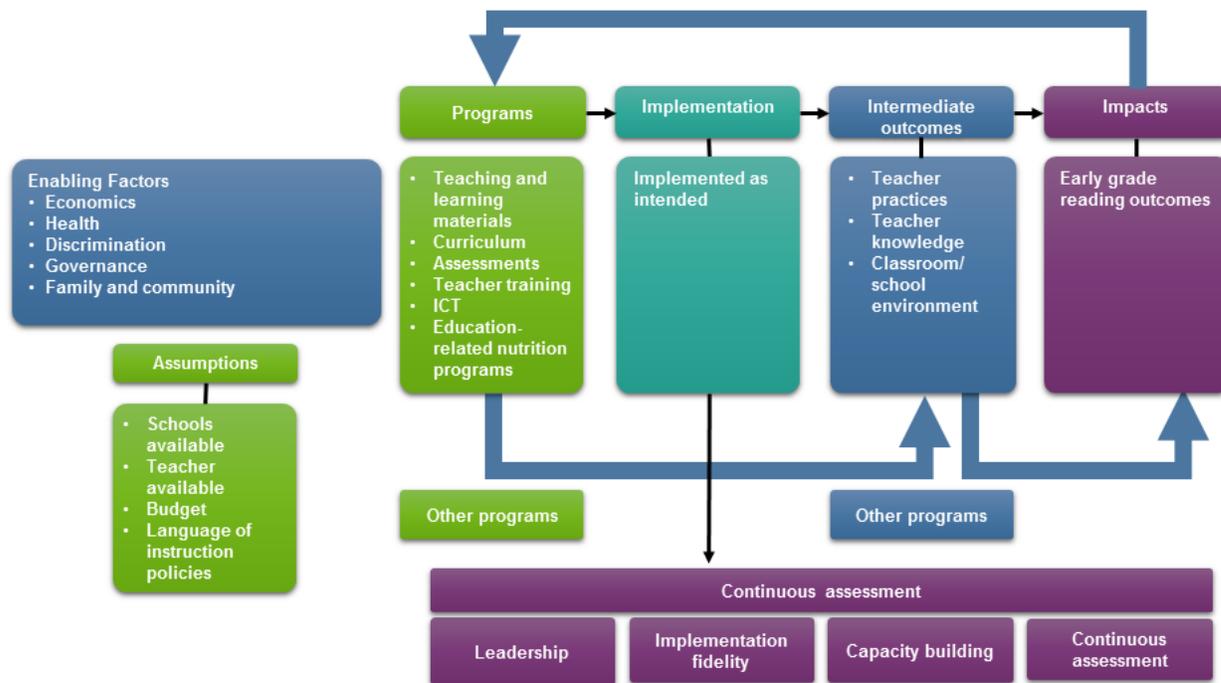
CONDUCTING THE SYSTEMATIC REVIEW

The systematic review included the following 13 phases described below:

I. ESTABLISHING A CONCEPTUAL FRAMEWORK

To synthesize the evidence, it is important to use a theory-based approach. Therefore, we developed a conceptual framework (Figure 1) that examines how various factors can influence early grade reading outcomes in the LAC region.

Figure 1. Conceptual Framework



EGR practice- and policy-relevant research should be built on a conceptual framework that maps out the linkages across enabling factors, education- and noneducation-related programs or initiatives that are

associated with literacy, intermediate outcomes, and reading outcomes, as well as the assumptions that underlie this framework. This conceptual framework explains how programs or initiatives can contribute to improving early grade reading outcomes in a sustainable manner.

2. DEVELOPING THE RESEARCH QUESTIONS

The research questions include both descriptive questions about what EGR programs are implemented in the LAC region and more analytically oriented questions related to what works to improve reading outcomes, how these programs work, and how enabling and implementation factors influence these relationships.

RESEARCH QUESTIONS ADDRESSED IN THE SYSTEMATIC REVIEW

1. What are the existing intervention- and nonintervention-based studies and what is the existing literature from or on the LAC region involving reading programs, practices, policies, and products focused on improving reading skills for children from birth through Grade 3?
2. What is the quality of the existing EGR evidence (quantitative intervention and nonintervention and qualitative intervention and nonintervention) in the LAC region and what is its practical use for varied LAC region stakeholders?
3. What are the gaps in the evidence base on EGR in the LAC region as compared to what we know globally about best practices in EGR?
4. What is the impact of reading programs, practices, policies, and products aimed at improving the reading skills of children from birth through Grade 3 on reading 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 effectively using evidence/knowledge to shape and/or improve EGR policy and practice in the LAC region?

3. DETERMINING THE RELEVANT POPULATION, INTERVENTION, COMPARISONS, AND OUTCOMES (PICO)

We determined the relevant population, interventions, comparisons, and outcomes on the basis of our research questions, knowledge about the LAC region, and our knowledge about experimental and quasi-experimental methods. We defined the relevant population as children in early grades in Latin America and the Caribbean. Furthermore, we decided to not restrict our sample on the basis of interventions because there are many interventions that can directly or indirectly influence early grade reading outcomes. We only determined appropriate comparisons for our synthesis of experimental and quasi-experimental studies. Other studies do not require a control or comparison group to enable a rigorous study. For the experimental and quasi-experimental studies, we included all randomized controlled trials (RCTs) and studies with multivariate analyses that included a comparison group. Finally, we included all quantitative studies that included a measure of early grade reading as an outcome variable. We did not determine appropriate outcomes for qualitative studies because high-quality qualitative studies do not require quantitative outcome measures.

4. DETERMINING THE RELEVANT STUDY TYPES

To answer our research questions, we included four study types. The first type encompasses experimental and multivariate nonexperimental studies that include a control or comparison group. We defined these studies as “quantitative intervention studies.” We included these studies to determine the impact of specific programs on early grade reading outcomes. The second study type consists of qualitatively oriented studies with a focus on interventions. These studies usually emphasize the process of program implementation or experiences of beneficiaries about the performance of the program. We defined these studies as “qualitative intervention studies.” The third type of study, quantitative studies, emphasizes predictors of reading outcomes and does not focus on the effects of a specific program. We defined these studies as “quantitative nonintervention studies.” We included these studies to increase our understanding of intermediate outcomes and their ability to predict reading outcomes. Fourth, we included qualitative studies that discuss literacy in the LAC region but do not include an emphasis on a specific program. We defined these studies as “qualitative non-intervention studies.” We included these studies to assess the experiences and perspectives of key stakeholders, including students, teachers, and policy makers, concerning literacy and reading.

5. DEVELOPING THE SEARCH STRATEGY

To develop and refine the search strategy, we relied on our PICO criteria and consultations with other researchers, librarians, computer scientists, and content experts. Through this process, we selected the most relevant databases for our review. We aimed to make the search strings as broad as possible to retrieve the maximum amount of potentially relevant items from all databases (Schuelke-Leech et al., 2015).²

6. SEARCHING FOR EVIDENCE

Following the development of the broad search strings, research associates at AIR used the search terms and strings (in each of the target languages)³ to conduct an initial search of online databases and development-focused websites, reviewed bibliographies of accepted articles to find other potentially relevant studies, and sent out emails to EGR experts in the LAC region and beyond in order to cast a broad net and capture as much of the evidence base as possible.

7. EXTRACTING DATA FROM IDENTIFIED SOURCES

We imported all citations found through the above search methods into the Mendeley reference management software (<http://www.mendeley.com/>). Mendeley automatically extracted bibliographic data from each book, article, or reference and removed all duplicates. At this stage, we were able to identify and export 9,696 unique documents.

² Full search strings and lists of all databases searched can be found in the full systematic review report.

³ Searches were conducted in English, Spanish, French, Portuguese and Dutch as appropriate. No documents were excluded because of language.

8. WIKILABELING

An AIR data scientist applied Wikipedia-based labelling and classification techniques to the extracted data to categorize text into meaningful categories and to increase the relevance of retrieved results using the well-known online encyclopedia, Wikipedia (Egozi, Markovitch, & Gabrilovich, 2011; Gabrilovich & Markovitch, 2006). Due to the broad and inclusive nature of our search strings, much of the initial evidence we captured was not actually relevant to our review. Therefore, we applied Wikipedia-based labelling to help us identify the most relevant pages. The process of identifying these pages is two-fold: first, experts need to share a list of potentially relevant categories. Next, we had to mine Wikipedia to find pages associated with exactly these or similar categories. We then validated the resulting list with the experts again. For example, “learning outcomes,” originally proposed by our experts, maps directly to “outcome-based education” within Wikipedia. Wikipedia’s innate hierarchical structure allowed us to make our categories less ambiguous and better organize them into a meaningful list.⁴

9. APPLYING INCLUSION CRITERIA AND RECORDING KEY INDICATORS

After narrowing down our list of articles through WikiLabeling, we imported all relevant citations back into the Mendeley reference manager software. We divided citations among reviewers, who applied the predetermined inclusion criteria (see Table 2) to each title and abstract. Our inclusion criteria were purposefully broad because we did not want to miss any relevant citations due to narrow inclusion criteria. Any article that did not meet one of the following five threshold criteria laid out in Table 2 was automatically excluded from further review.

Table 1. Initial Inclusion Criteria for EGR Evidence

#	Category	Criteria
1	Year of Publication	Include <i>literature</i> from the last 25 years, a time frame spanning 1990–2015. We will update the search in each subsequent year of the 5-year project.
2	Relevance to the Region	The evidence must be from or on the LAC region including any or all of the following: 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
3	Relevance to the Population	Boys or girls ages birth through Grade 3 in the LAC region. If the children are enrolled in Grade 3 or below, they fall within our population regardless of the age.
4	Relevance to the Topic	The literature must have a focus on reading or literacy (which includes reading and writing).
5	Is It Research?	There must be a research question or research objective and a methodology that matches that objective.

⁴ See the full systematic review document for more information on the process of WikiLabeling.

10. REVIEWING FULL TEXT USING QUALITY REVIEW PROTOCOLS

We compiled the full-text articles and books that met all inclusion criteria, as well as those that were still unclear after the title and abstract review, and assigned them to senior researchers based on language and type of study. The senior researchers reviewed the articles using separate quality review protocols based on the type of study as follows:

- Quantitative intervention studies: An adapted version of a risk of bias (RoB) assessment tool developed by Hombrados and Waddington (2012)
- Quantitative nonintervention studies: An adapted version of the RoB tool for quantitative intervention studies, which removed any questions regarding interventions.
- Qualitative intervention and nonintervention studies: An adapted version of the Critical Appraisal Skills Programme (CASP) Qualitative Research Checklist
- Mixed-methods studies: Both the RoB tool and the qualitative protocols were applied. Two or more reviewers read and rated all quantitative intervention studies to ensure consensus. For the other types of studies, pairs of reviewers rated the same studies at the outset to ensure a common understanding of the quality categories, but the remaining articles were reviewed by single reviewers because of time constraints.

11. ANALYZING DATA

We used different types of analyses for each type of research. First, we implemented a combination of meta-analysis and narrative synthesis to analyze the effectiveness of programs that could potentially influence reading outcomes. We calculated the standardized mean difference and the standard error for each of the studies included in the meta-analysis. Where possible, we used a stratified meta-analysis to differentiate the results of studies with a low, medium, or high risk of bias and to differentiate between the findings of RCTs and nonexperimental studies. Second, we used a narrative review to examine the main lessons from the included qualitative intervention and nonintervention studies. To identify these lessons, we relied mostly on the findings of high-quality studies. Third, we analyzed the main lessons about the predictors of early grade reading outcomes in the LAC region from quantitative nonintervention studies. For this purpose, we again relied on the studies that were identified as higher quality in the risk of bias assessment.

12. MAPPING THE GAPS IN THE EVIDENCE

To create the evidence-gap map for the quantitative intervention studies, we coded the intervention types and outcome measures and linked these to various characteristics of the evaluated programs. This process allowed us to create evidence-gap maps that demonstrate what evidence is available on the impact of teacher training, nutrition, ICT, preschool, school governance, teacher practices, and parental involvement on reading outcomes. In this process we differentiated evidence-gap maps by methodology (experimental versus nonexperimental), risk of bias, socioeconomic condition (high-income versus upper-middle income, lower-middle-income, and low-income country), and country. We created a separate evidence-gap map for quantitative nonintervention and qualitative intervention and nonintervention studies that shows the study topics, type of research, and country of the research for the medium and high quality studies.

13. TRIANGULATING FINDINGS

After conducting the quality review and synthesis of articles and mapping the gaps, reviewers triangulated the different syntheses by linking the evidence back to the conceptual framework. We examined the impact of the different programs on reading outcomes and triangulated these findings with the qualitative research articles to examine whether the fidelity of implementation or experiences and perspectives of different stakeholders may have influenced the impact of these programs. In addition, we assessed the predictors of reading outcomes to increase our understanding about the linkages between intermediate outcomes, such as teacher knowledge and behavior, and reading outcomes. Finally, we used the information from the qualitative research to examine whether and where any links in the conceptual framework broke down. Findings from the qualitative synthesis and the quantitative nonintervention synthesis helped describe, explore, and interpret how specific programs improve reading outcomes.



Miriam Martinez/LRCP-CIASES

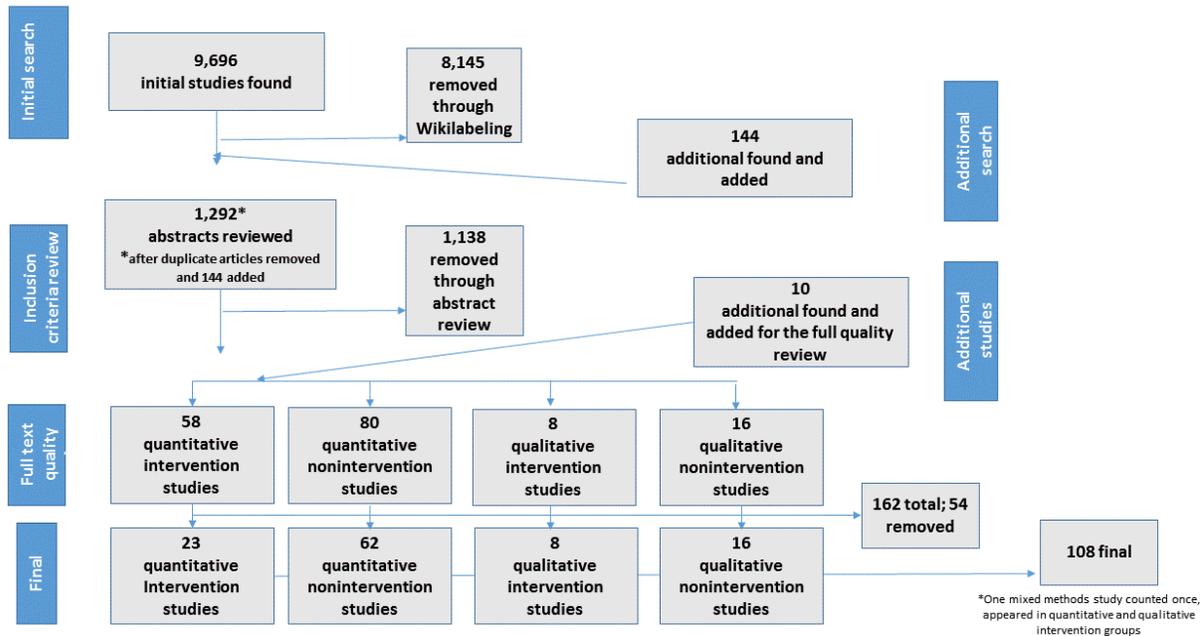
CATEGORIZING EVIDENCE

After completing the systematic review phases as described above, we included **108 studies with a focus on early grade reading outcomes in the LAC region in the final systematic review.** Our initial search resulted in 9,696 articles. We then applied WikiLabeling, and a manual review of the abstracts against our inclusion criteria. Following this review we were left with a total of 162 studies that underwent full text review. During this phase, an additional 54 articles were removed as not relevant, resulting in 108 studies included in the final review.⁵

⁵ One mixed-methods study was counted twice as it was reviewed by both qualitative and quantitative reviewers.

The 108 included articles were comprised of quantitative intervention research, quantitative nonintervention research, qualitative intervention research, and qualitative nonintervention research. The following graphic shows the stages and final results of the analysis:

Figure 2. Results per Stage



The vast majority of studies included in our review of evidence were published journal articles and came from either North or South America; significantly fewer articles were from Central America and the Caribbean. Most articles were published in English or Spanish.

More than 90% of the articles were focused on high- to upper-middle-income countries⁶. The disproportionate emphasis on high-income and upper-middle-income countries may be explained by the limited available resources and capacity for conducting high-quality research in low-income and lower-middle-income countries.

An analysis of the quantitative intervention studies indicates that impact evaluations with an emphasis on early grade reading outcomes only focus on a small portion of the intervention types that can influence early grade reading outcomes. We found only three topic areas with more than two impact evaluations that focus on early grade reading outcomes:

⁶ The World Bank classifies country economies into four income groupings: low, lower-middle, upper-middle, and high. Income is measured using gross national income (GNI) per capita. For the current 2017 fiscal year, low-income economies are defined as those with a GNI per capita, calculated using the World Bank Atlas method, of \$1,025 or less in 2015; lower middle-income economies are those with a GNI per capita between \$1,026 and \$4,035; upper middle-income economies are those with a GNI per capita between \$4,036 and \$12,475; high-income economies are those with a GNI per capita of \$12,476 or more. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

- (1) teacher training
- (2) nutrition intervention
- (3) Information and Communications Technology (ICT) programs

Although the majority of the included impact evaluations used a randomized controlled trial (RCT), only eight of the studies were rated as having a low risk of selection bias. Of the eight studies with a low risk of selection bias, two focus on child nutrition, three focus on ICT, one focuses on parental and community participation, one focuses on teacher practices for reading, and two focus on teacher training. These data show that there is little strong evidence regarding the impact of development programs on early grade reading outcomes in the LAC region.

Although we only found a very limited number of high-quality quantitative intervention studies, they did indicate examples of development programs that are likely to have positive effects on early grade reading outcomes in specific circumstances and contexts. For example we found evidence that:

- Teacher training programs can positively affect early grade reading outcomes in high-income economies when they are well implemented and complemented by the sustained coaching of teachers.
- Nutrition programs can have positive effects on early grade reading outcomes in contexts where stunting and wasting are high, such as Guatemala.
- The distribution of laptops to children can have adverse effects on early grade reading outcomes, particularly when the distribution of laptops was not complemented by additional programs.

Findings of the quantitative nonintervention studies indicate that:

- Phonemic awareness, phonics, fluency, and comprehension are associated with reading ability.
- Poverty and child labor are negatively correlated with early grade reading outcomes. This finding on the importance of poverty and socioeconomic factors for early grade reading outcomes supports the quantitative intervention result that nutrition programs may be effective in improving early grade reading outcomes.
- The quality of preschool is positively associated with early grade reading outcomes. Triangulating this result with the quantitative findings on the impact of teacher training suggests that teacher training combined with sustained coaching could possibly positively affect early grade reading outcomes through its influence on the quality of preschool.

Both qualitative and quantitative studies indicated that consideration of context is key to improving reading outcomes. This lends credence to the conceptual framework, which suggests that enabling factors and assumptions in part determine the potential for success of various programs or strategies. The most frequently discussed topic in qualitative nonintervention articles is the need to promote social learning to improve early grade reading.

However, we found strong evidence for publication bias in the studies that focus on the effects of teacher practices and parental involvement on early grade reading outcomes in the LAC region; that is, there are likely to be a large number of additional studies that have not been published on similar topics because they did not find statistically significant effects. Findings from statistically unsuccessful interventions are also important, and publishing only the results of programs that show positive and statistically significant effects on early grade reading outcomes impedes policy makers' ability to make evidence-based decisions.

EVIDENCE GAPS

The primary end goal of all activities within the LAC Reads Capacity Program is to enhance the capacity of key stakeholders (e.g., the Ministry of Education and the government, international funders and intergovernmental entities, international and national NGOs, academics, and researchers and practitioners) to use evidence to choose, develop, implement, and evaluate early grade reading strategies, programs, practices, and interventions. Through this systematic review of the EGR evidence from the LAC region we identified multiple gaps in the evidence-base. These gaps indicate that key stakeholders face significant challenges when attempting to make evidence-based decisions.

Next we present our recommendations based on the evidence gaps to highlight specific areas in need of further funding or research. This research could help to fill in the evidence gaps and provide more robust and comprehensive evidence on what works in early grade reading in the LAC region.



RECOMMENDATIONS BASED ON THE EVIDENCE GAPS:

- Ensure that language assessments include multiple reading constructs and differentiate between those constructs so it is easier to identify the effects of interventions on individual constructs.
- Fund long-term mixed-methods experimental or quasi-experimental research on the effects of preschool and early childhood education on early grade reading outcomes.
- Include several early grade reading constructs in administrative data to enable researchers to conduct high-quality research on the mechanisms underlying early grade reading using large sample sizes.
- Document ongoing research to minimize publication bias so that unpublished research is available to policy makers as well and to ensure that hypotheses are pre-specified.
- Register ongoing research on early grade reading in a central, publicly available location so that everyone can see what is being done and seek to complement and add to the research base.
- Develop more interdisciplinary mixed-methods research on early grade reading that includes more than one reading construct and large sample sizes.
- Fund rigorous research that allows for an examination of the causal effects of specific development programs on early grade reading outcomes. These studies include both experimental and quasi-experimental studies with a sufficient sample size. These studies also need to be supplemented with qualitative research.
- Pursue more research on EGR strategies for students with disabilities.
- Pursue more research on reading in indigenous languages.
- Conduct more research on the linkages between the development of prewriting and writing skills and early grade reading outcomes.

In addition to reviewing the EGR evidence from the LAC region, the LAC Reads Capacity Program also collects and catalogues EGR pedagogical resources (e.g., supplementary reading materials, assessments, instructional materials, videos), and other EGR documents that are neither research-based evidence nor resources (e.g., policy documents, project reports, best practices documents) from the LAC region. In order to support stakeholders to improve their practice, the program is developing a resource database making these resources available to a wide audience through the program's website at <http://www.lacreads.org>.

The systematic review examines the existing evidence about reading-related programs, practices, and policies in the LAC region. Over the remaining years of the project, the systematic review team will continue to search, evaluate and include new evidence on early grade reading in updated versions of the systematic review report. The evidence collected can be used to inform the ongoing work of practitioners, support evidence-based policy decisions, and provide direction and priorities for further research in this field. To further this process the LRCP team and national partners are also conducting a stakeholder analysis in the region to identify the key EGR stakeholders, determine their interests and needs, and how the evidence from this review and the resources collected can best be used to support EGR capacity and achievements in the region.



REFERENCES

- Egozi, O., Markovitch, S., & Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. *ACM Transactions on Information Systems (TOIS)*, 29(2), 8.
- Gabrilovich, E., & Markovitch, S. (2006, July). Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge (pp. 1301–1306). *AAAI'06 proceedings of the 21st national conference on artificial intelligence, Volume 2*. Menlo Park, CA: AAAI Press.
- Hombrados, J. G., & Waddington, H. (2012). Internal validity in social experiments and quasiexperiments: An assessment tool for reviewers. Mimeo: 3ie.
- Schuelke-Leech, B. A., Barry, B., Muratori, M., & Yurkovich, B. J. (2015). Big Data issues and opportunities for electric utilities. *Renewable and Sustainable Energy Reviews*, 52, 937–947.
- Shadish, W. R. (2002). Revisiting field experimentation: field notes for the future. *Psychological Methods*, 7(1), 3.
- 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.
- World Bank. World Bank Country and Lending Groups. Retrieved September 30, 2016, from <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>