



# Facilitating Impact Evaluations: Recommendations from the LAC Reads Evaluations

*This is one in a series of briefs uncovering lessons learned from four impact evaluations of promising reading interventions funded by USAID as part of the Latin America and the Caribbean Reads (LAC Reads) project. The evaluations were conducted by Mathematica.*

## Background

Successful implementation of a randomized impact evaluation can be difficult, time consuming, and expensive, causing many researchers to avoid or not fully implement this rigorous approach. As a result, funders and implementers may not receive strong evidence on which to base their program decisions. The evaluations conducted under the LAC Reads project reveal several factors that can facilitate a successful randomized impact evaluation. These factors include: engaging stakeholders and partners on the ground, being nimble and creative, evaluating validated programs, using national data to improve cost-efficiency and relevance, and using multisite approaches and multiple comparisons to generate generalizable results.

## Start early, work inclusively, and be flexible

### Design impact evaluations before rolling out interventions to ensure a robust design.

Random assignment designs are considered the most rigorous method for estimating causal effects, but they require careful planning and stakeholders' buy-in. On all LAC Reads evaluations, we worked closely with the implementing partner and other stakeholders to determine the evaluation design before project rollout. A planning period allows evaluators and implementers to share information and discuss necessary adjustments to the evaluation design or program implementation before the intervention begins. Stakeholders unfamiliar with randomized designs may have concerns about allocating services to a subgroup of potential beneficiaries (such as a treatment group). Evaluators can address those concerns by engaging stakeholders and explaining to them how randomized designs work, what they do and do not require, and what policymakers can learn from them. Evaluators can address any ethical concerns stakeholders have about equity of service provision, and discuss the advantages of using transparent criteria for random assignment, especially when a program does not have resources to serve all eligible beneficiaries. Allowing time for coordination and collaboration before finalizing the implementation

## Lesson from LAC Reads

In Honduras, where we evaluated the impacts of providing support to schools using summative and formative instruction, we began working with the implementing partner and the USAID Mission on the evaluation design nearly a year before the program rolled out. This allowed us to present the rationale for the evaluation design to program staff during their initial training and gain their buy-in. As a result, when program staff learned that formative assessments were inadvertently being distributed to schools in the control group, they alerted us and the Ministry of Education immediately—which enabled us to take the assessments out of circulation and ensure the integrity of the evaluation design.

and evaluation approach also helps establish strong relationships that facilitate information exchange and data sharing. This is particularly true in education evaluations, where accessing pre-intervention student achievement data is important. Often, implementing partners or ministries of education can provide these data, which the evaluation team can use to measure balance between groups and strengthen the precision and credibility of the impact estimates.

## Be flexible and creative to transform challenges into opportunities.

Despite adequate planning, program activities or delivery may change due to unforeseen circumstances. For example, education projects may face teacher strikes, changes in national curriculum, or the rollout of other similar projects implemented in targeted schools. If the context or program changes, so too must evaluation strategies. Being flexible can enable evaluators to identify new learning opportunities. Throughout the evaluation period, regular communication about project implementation, such as whether implementers are meeting goals laid out in the logic model or are adjusting program activities, is crucial to allow proactive adjustment of evaluation activities. If unexpected changes occur, evaluators can incorporate additional or new contrasts into the study design, or tweak implementation to preserve the design's integrity.

### Lesson from LAC Reads

The *Amazonia Lee* evaluation was complicated by the national government's rollout of a similar intervention in the same region. Thanks to the collaborative relationship our local research partner had established with the regional educational authorities, when our partner recognized the threat to the integrity of the evaluation they were able to facilitate an agreement to implement the government intervention only in the control schools. This step preserved the randomly assigned groups for the *Amazonia Lee* evaluation. By maintaining the integrity of the random assignment, the research team turned this challenge into an opportunity to test *Amazonia Lee's* effectiveness against *Soporte Pedagógico*—the flagship education quality improvement program from Peru's Ministry of Education.

The implementer of *Espacios para Crecer (EpC)* initially asked that random assignment for the evaluation be at the child level, so the program could serve the maximum number of beneficiaries. Once the implementers began to assess the sizes of communities eligible for *EpC*, they realized that many were too small for that approach to work in small communities where all eligible beneficiaries would need to be served to make the *EpC* itself viable. We developed a hybrid design for random assignment, in which we used child-level assignment for large communities, and community-level assignment for smaller communities.

## Evaluate validated programs

**Evaluations of programs that have been validated in a specific context may proceed more smoothly and lead to clearer findings.**

When planning to evaluate an intervention, consider whether it has been adapted to the context, aligns with needs of prospective beneficiaries, and represents a meaningful contrast to prevailing practices or business-as-usual services. Programs that are not validated may run into challenges due to local context that constrain implementation and limit improvements in learning. If an intervention has not been validated in a new context, funders should consider including time to adapt and pilot it before fully rolling out the intervention and the evaluation.

Piloting interventions can flag risks related to their feasibility, or barriers to implementation and obtaining local stakeholders' buy-in. Pilots can test assumptions underlying the program's theory of change and improve the alignment between the needs of beneficiaries and the program's offerings. Findings from a pilot provide a foundation for the evaluation design that enables researchers to understand points of similarity and contrast to the intervention and select a counterfactual that is best suited to answer the research questions. Pilots of interventions can also facilitate smooth rollout and implementation fidelity. Pilots can be particularly successful when implementers with local presence, deep understanding of the education system, and established relationships with education authorities lead them. This knowledge of the local context is helpful for adapting the intervention to local conditions, leading to evaluations with clear, actionable results.

### Lesson from LAC Reads

The two evaluations of LAC Reads interventions that were previously tested in very similar contexts in the same country led to clearer, more actionable results than did the ones that had been piloted in different countries and required more adaptation.

The teacher training model tested in the *Amazonia Lee* evaluation had been designed and implemented in other regions of Peru by the Universidad Peruana Cayetano Heredia prior to the evaluation. In contrast, although the *Leer Juntos, Aprender Juntos* intervention had been implemented in other developing countries, it was not piloted in Guatemala and Peru before the implementation for the impact evaluation. Findings indicate that it did not have an impact on learning outcomes in either of these countries.

The intervention evaluated in Honduras used assessments that were already in widespread use by the Ministry and was implemented by a team experienced in providing similar support to schools. Building on this experience, the implementer was able to roll out the program quickly to the evaluation schools and maintain fidelity to the intended model and dosage, which can be a challenge when significant adaptation occurs during implementation.

## Multisite evaluations with multiple comparisons maximize learning opportunities

Multisite evaluation designs are particularly useful to examine the impact of an intervention across a range of contexts, build a knowledge base rapidly, and increase confidence that the findings are broadly applicable.

Multisite evaluations use a coordinated effort to address the core research questions in different contexts. They often involve a centralized study design and are conducted in multiple sites to increase the sample size, improve the representativeness of the study population, or ensure that the intervention is effective in multiple settings or contexts (Straw and Herrel 2002). They can be particularly useful when examining the impact of an intervention across regions and countries, where the counterfactual varies. Collectively, multisite evaluations may be suggestive of the value of key components or approaches and offer insights not gained through single site evaluations. Including different sites with different comparison groups contributes to the strength and generalizability of the findings and yields insights about the mechanisms and local conditions that affected the outcomes.

### Lesson from LAC Reads

The multisite evaluations of *Amazonía Lee* and *Leer Juntos, Aprender Juntos* examined the impact of an intervention across sites where the counterfactual varied. Collectively, the multisite evaluations suggest that offering teacher training on foundational reading skills instruction can be a useful mechanism to help improve student reading outcomes in LAC countries when it is not part of the teachers' pre-service or in-service training. Including different sites with different comparison groups contributed to the strength and generalizability of the findings and yielded insights about the mechanisms and local conditions that affected the outcomes.

*Amazonía Lee*, implemented in two Amazonian regions of Peru, was found to be effective in Ucayali, where children who received *Amazonía Lee* had better reading outcomes than children in schools that did not receive additional support beyond what the Ministry of Education typically provides. In San Martín, where children who received *Amazonía Lee* were compared with children receiving additional support services from *Soporte Pedagógico*, students had similar reading outcomes.

## Leverage National Reading Assessments to Measure Early Grade Reading Outcomes Whenever Possible

Funders and evaluators should consider using available data from national reading assessments to assess the impacts of early grade reading interventions, instead of collecting data in person.

Measuring reading outcomes with reliable national assessments, if available, can save time and resources compared to adapting and conducting an early grade reading assessment (EGRA) or another evaluation-specific reading measure. Leveraging available data and systems supports local capacity and better enables national stakeholders to use evidence to refine policies and programs. Findings based on a local test may resonate more with education stakeholders and policymakers and generate greater interest in applying what is learned from the evaluation to their decisions. Although EGRA has been instrumental in assessing and tracking early grade reading progress in the developing world, country-specific adaptations of EGRA for program evaluation are resource-intensive, particularly when multilingual populations are involved, and the data collection approach must be tailored to measure intermediate outcomes. Using administrative or education management information systems data to measure intermediate or final outcomes, when available, can be more efficient than in-person data collection.

### Lesson from LAC Reads

For the Honduras evaluation, using national student assessment data to assess the impacts of providing support to principals and teachers in the use of formative and summative assessments to improve reading instruction had two advantages. First, using the readily available pre-intervention assessment data—and using the Ministry of Education tests at the end of the intervention—reduced costs and saved significant time for instrument development. Second, findings could be expressed in metrics that were familiar to local stakeholders and policymakers. While national assessment data also existed in Peru, the timing of testing for students did not align with our measurement timeframe for the evaluations of *Leer Juntos, Aprender Juntos* or *Amazonía Lee*. In the first case, students whose mother tongue was not Spanish were not tested until fourth grade; in the second case, waiting for the release of the data for second graders would have significantly delayed the delivery of our report. Not only were significant time and resources required to adapt the EGRA to this population, in-country stakeholders wanted to understand the program impacts in terms of the national reading assessment (ECE). A long-term follow-up of the *Amazonía Lee* evaluation compared whether impacts varied with the type of reading assessment and found using ECE in place of EGRA data yields similar results in this context.