IARTEM · e-journal

Vol. 14 no. 1
Peer-reviewed research article

DOI: https://doi.org/10.21344/iartem.v14i1.976
ISSN: 1837-2104

Teaching by the Book: Teacher Decision-making while using Structured Lesson Plans

Yasmin Sitabkhan, RTI International
Karon Harden, RTI International
Timothy Slade, RTI International

Abstract

The purpose of this paper is to present a methodology for understanding materials usage in primary classrooms in Sub-Saharan Africa that centres teachers' actions and voices. The United Nations Sustainable Development Goal 4 focuses on improved primary education around the world. To meet this goal, many large, donor-funded interventions aim to improve education through provision of teacher's guides and student textbooks. However, what many of these interventions lack is a systematic way to understand how and why teachers make pedagogical decisions while using materials. A large body of work seeks to understand how teachers make decisions as they teach, and how these decisions are influenced by their knowledge and beliefs. Drawing from this work, we describe a methodology and set of tools that uses observations and interviews to identify key decisions that teachers make in the classroom and why they make those decisions. We piloted and iteratively refined this methodology over the course of three studies and use examples from these studies to illustrate the methodology. By closely observing and listening to teachers, we gain insights that allow us to continually refine and improve materials to ultimately improve the quality of classroom instruction.

Keywords: Teacher's Guides, Sub-Saharan Africa, Primary school, Teacher decision-making

Introduction

Mrs Banda is teaching a Standard 2 Chichewa reading lesson in Malawi. The lesson plan that Mrs Banda is using asks her to introduce new vocabulary words and to ask the students to share the meanings of the words; if they do not know, Mrs. Banda should tell them the meaning of the word. When Mrs Banda arrives at two words, *mkeka* and *mphasa*, she anticipates that students will be confused because the words refer to mats made of different materials and used for different purposes. Mrs Banda is prepared for this and takes out examples of the two types of mats and explains the difference between them. After the lesson, when asked why she used the mats, Mrs. Banda replies that when students see the actual objects, it helps them understand and remember the meaning of the word.

Mrs Banda and other teachers in government schools in Malawi were participating in a large-scale reading intervention programme. They were continually making modifications to the structured lesson plans provided by the programme. Mrs Banda drew upon her knowledge of her students to anticipate a misunderstanding that might arise and apply a pedagogical strategy to address this misunderstanding. Since this use of realia was not explicitly prescribed in the teacher guide, it can be considered a modification—in this case, an enhancement—to the existing lesson plan.

Goal 4 of the Sustainable Development Goals calls for access and quality education for all (United Nations, 2015), and a growing number of interventions in middle- and low-income countries, such as the one Mrs. Banda participated in, recognise the need to increase the quality of education through improved classroom instruction. To promote better student outcomes, recent donor-funded projects offer a package of interventions that include teacher training, classroom materials, and support systems. One common component of classroom materials is a teacher's guide with structured lesson plans, based on the local curriculum and evidence-based pedagogical practices. The lesson plans may or may not be "scripted" to varying degrees (see Smart, 2021, for a discussion), but at a minimum offer guidance to teachers attempting a new methodology.

In any context, providing a teacher's guide with lesson plans is one thing, and its execution in the classroom is another. Many researchers have highlighted the fact that teachers are constantly making decisions about the lesson and adjusting their instruction midstream (Cuban, 2021); a large body of literature, both in high-income and low- and middle-income contexts, is dedicated to documenting and understanding teacher decision-making before, during, and after lessons (Chachage, 2020; Cohen, 1990; Serbessa, 2006; Schoenfeld, 2008). This literature has established that teachers' spur-of-the-moment decisions, as well as their post-hoc analysis of events, stem directly from their knowledge of students and content, attitudes about learning, and beliefs about good practices in instruction. All of these factors are rooted in cultural systems that cannot be ignored when promoting education reform.

The decisions which teachers make about the lesson plans, and their reasoning behind these decisions, should be a valuable source of information for an intervention (Smart, 2021). We argue that for a teacher's guide to be of maximum utility, it should be well calibrated to where the teachers are, taking into account the skills they already have, what they believe and value; and how their existing pedagogical worldview interacts with the new approaches promoted in the guide. It is important to stress that all teachers make decisions about lesson plans. We fully expect teachers to do so; modifying does not constitute an implementation "failure". How, then, does one identify the changes teachers are making to existing guides, and use that information to improve the guides for future revisions? We need a systematic methodology to identify the decisions and voices of the implementers and then present patterns in these decisions that provide actionable information for teacher guide development.

In this paper, we describe the development of a methodology and tools used to centre teacher actions and voices in the development and improvement of pedagogical materials. The methodology focuses on the analysis of teacher decision-making through the lens of their modifications to a prescribed lesson plan. After teachers have been given a set of pedagogical materials and trained on their use, we capture the modifications they make to the lessons as they implement them in real time. We then probe the teachers for insights into why they made these modifications. The lens of modifications allows us to better understand which instructional moves teachers are inclined to make and why, and how these moves are supportive or not supportive of the target learning outcomes compared to the original lesson plan. We thereby gain actionable insights that inform improvements to the lesson plans.

This paper draws on our experience across three studies to show how we developed and refined the lens of modifications across different contexts: first, through an exploratory study in Liberia to understand how teachers were using new mathematics teacher guides; second, through a comprehensive study conducted in Malawi on teachers' implementation of a reading programme; and finally, a systematic, multicountry study examining the use of teacher guides in various projects.

Literature review

Many interventions aimed at improving teachers' classroom practice are premised on a simplistic theory of change that if a teacher's guide is pedagogically sound and user-friendly, and if the teachers are adequately trained and supported in its execution, teachers will follow the guide and their teaching will improve. This oversimplification of teacher behaviour change does not consider the influence of the rich body of experience, knowledge and beliefs that teachers bring with them into the classroom. Instead, knowing what teachers who are freshly armed with new materials and training do upon their return to the classroom is key to understanding how to improve uptake and use, with the aim of higher quality instruction as well as continual adaptation of materials and training to better target the needs of teachers.

One of the most common tools for monitoring implementation is quantitative classroom observation (Hertz, Kochetkova, & Pflepsen, 2019). In this kind of observation, a visitor to the classroom—often an instructional coach, district education officer, project staff, or external evaluator—sits "unobtrusively" during a lesson and follows a predetermined checklist to record discrete behaviours that they do and do not observe during the lesson. For example, in Nigeria, observers ticked yes/no responses to items such as, among others, "Did the teacher say the letter sound? Did the teacher ask the pupils to predict what might happen in the story? Did the teacher balance opportunities given to boys and girls to speak?" (RTI International, 2015). Low-inference checklist items ask the observer to record data on "behaviors that are distinct, observable, and objective, requiring no evaluative judgment by the observer" (Hertz, Kochetkova, & Pflepsen, 2019, p. 8). Sometimes the observers use a Stallingstype timed observation (World Bank Group, 2017) to count the relative frequency of a type of behaviour of interest. These methods lend themselves relatively easily to high interrater reliability and can be easily rendered into quantitative data. At the individual level, the resulting data are also frequently used to identify teacher strengths and weaknesses to guide coaches in giving targeted support and to inform ongoing professional development needs. At the system level, this kind of data is often collected for monitoring the operationalisation and fidelity of implementation of interventions, evaluating intervention impact on teacher behaviour, and even measuring the quality of instruction at the system level. Education stakeholders have utilised this type of observation protocol across diverse country contexts such as Kenya (Freudenberger & Davis, 2017), Tanzania (Brombacher et al, 2014), Zambia (EnCompass, 2017), Senegal (Chemonics International, 2020), Morocco (Messaoud-Galusi et al, 2012), the West Bank (RTI International, 2018), Nepal (RTI International, 2014), the Philippines (Harden, Sowa & Punjabi, 2019), Guatemala (Lugo-Gil et al, 2016), and numerous others.

However, the data derived from checklist instruments have limitations: they do not give a full picture of classroom practice; they do not necessarily discriminate between high and low-quality instruction; they do not consider teacher agency or the why behind their classroom behaviours; and they do not always provide sufficiently nuanced information to inform refinements of the intervention components. A more detailed understanding is needed to identify which elements are working as intended or not, why they are or are not, and how to equip the intervention to better meet its objectives by addressing strengths and weaknesses in the intervention model itself.

Considering the shortcomings with quantitative data, in recent years there have been movements towards the collection of qualitative data to better understand teacher practice in the classroom. For example, when observing teachers in the USAID-funded Tusome Early Grade Reading Activity in Kenya, in addition to completing quantitative checklists, project staff completed brief forms with descriptive, qualitative responses to open-ended prompts such as "What went well?", "What went poorly?", and "Overall observations and recommendations?" Many interventions that offer structured coaching support incorporate similar prompts into the protocols that guide coaches' lesson observations

and teacher debriefs, often in combination with the checklist-type tools mentioned above, such as in Liberia (RTI International, 2019). While valuable for understanding an individual teacher's struggles, the resulting data are often an anecdotal collection of the observer's thoughts and feelings than evidence that can be analysed in a systematic manner. Systematically organising and analysing qualitative data also requires a lot of time and a highly specialised skill set. These factors make regular and timely qualitative data collection impractical for many organisations to implement.

The success of an intervention rests in large part on the ability of teachers to deliver improved instruction that leads to higher learning outcomes. Missing in both types of data mentioned above is the voice of the teacher. These data collection methods neglect the centrality of the teacher. Many studies have found that teachers' beliefs and knowledge influence their classroom practice (Nespor, 1987; Ernest, 1989; Wilkins, 2008; Stipek et al., 2001; Brickhouse, 1990). We know from prior research that teachers interpret new pedagogical methods through the lens of their existing knowledge and belief systems (Pajares, 1992). Schoenfeld (2008) describes a theoretical model that explains teacher behaviour in the classroom in terms of a series of decisions made in the moment, based on their knowledge, beliefs, and goals. Teacher knowledge includes knowledge about the subject matter and pedagogy (both general and content-specific); it also includes knowledge built from prior experience, knowledge of the class history and students' individual characteristics. Teacher beliefs include beliefs about how students learn, what students should learn, what the teacher's and students' respective roles are in the learning process, and much more. In Schoenfeld's model, as a teacher proceeds step by step through a lesson period, her knowledge, beliefs, and goals activate for her a "set of plausible options" for what to do next at any given moment in time (whether the lesson is going as expected or not); she likewise relies on her knowledge, beliefs, and goals to conduct an instantaneous cost–benefit analysis in order to decide from among those options. Any new pedagogical behaviour that she is being asked to implement will be subjected to this same momentby-moment analysis, grounded in her body of knowledge, beliefs, and goals, and enacted in response to the unfolding events. As a result, the teacher may decide to execute the action prescribed by the new metholodogy (e.g. as stated in the lesson plan) at that moment, or she may choose another option, which then becomes a "modification."

Cohen (1990) demonstrates this principle at work through the case of Ms Oublier. Ms Oublier was trained in a new approach to mathematics instruction in the mid-1980s as part of curriculum reform. She enthusiastically embraced the new materials but also, throughout a year, interpreted and used them through her existing frames of knowledge, such that in the end, her actual teaching practices did not look very different from before the reform curricula was introduced.

In this paper, we aim to describe the process we developed and refined to gather systematic, actionable evidence on how teachers are using materials in classrooms, with insight into the decisions they are making, thus centring the voice of the teacher. In doing so, we did not attempt to categorise teacher decision-making as insight into their knowledge versus their beliefs or attitudes. Instead, we focused our analysis on how teacher decision-making around the teachers' guide would inform the revision of the guide, and contribute to efforts to improve the overall programme. We illustrate how we developed and iteratively refined our methodology across three studies, all in different contexts in Sub-Saharan Africa, to be able to capture teacher actions and decisions in a way that can provide actionable results for improving materials, training, coaching, and all ways in which we strive to improve learning outcomes in low- and middle-income contexts.

Case Study 1: What is valuable to capture in classrooms to understand teachers' use of teaching and learning materials?

In 2013, as part of the Liberia Teacher Training Project (LTTP), a USAID-funded intervention targeting improved learning outcomes in early grade mathematics and reading, we conducted an exploratory study to better understand how teachers were using the project-provided mathematics teacher guides and accompanying student books. At the time, the project had an upcoming opportunity to revise the scripted teacher guides and we wanted to gather data to inform improvements.

A small sample of teachers (n=15) were identified in three counties who were observed to be implementing the materials with success after the training. We created a team comprised of an international mathematics education researcher, two Liberian teaching experts, and a representative from the Liberian Ministry of Education, who co-created a lesson observation protocol based on one already in use for project monitoring. The protocol consisted of a checklist with items such as "Did the teacher teach all elements in the guide?", "Was their pace adequate?", and "Were students engaged?"

Upon piloting the protocol, the research team quickly realised that it did not provide enough specific information to inform revisions to the materials. The checklist did not capture useful data about how teachers were interacting with the guides. For example, in the pilot, all observers checked "No" to "Was the pace adequate?", but this item provided us with little actionable information to inform revisions. We did not know why the pace was not adequate, or even what this meant across different classrooms. We did, however, notice that several teachers were modifying specific parts of the lessons. Therefore, the research team decided to have one observer focus on recording what teachers were modifying by closely comparing their actions against the scripted lesson plan.

The observers found a total of 35 teacher modifications, defined as the teacher deviating from the given lesson plan in any way. We then looked for patterns in the types of modification that teachers made and found three emerging types: student participation, elaboration of content, and sequence and timing. Student participation modifications consisted of a change in the amount or structure of children's participation in the lesson. Elaboration of content modifications consisted of increasing or decreasing explanation of a particular topic. Sequence and timing modifications consisted of a change in the steps of a lesson, or a change in the amount of time that the guide suggested being allocated to a part of the lesson. In addition to these patterns, each modification was coded as either helpful or hurtful to the overall pedagogical effectiveness of the lesson. These codes, though subjective, were based on the stated pedagogical goals of the project.

To illustrate, we present observation #16, a student participation modification which was judged to help the overall effectiveness of the lesson. In this lesson, the teacher's guide asked teachers to draw two triangles on the board and then tell students that each triangle has 3 sides and 3 angles. The teacher drew the triangles on the board as the guide asked. However, instead of then *telling* the students the scripted explanation, the teacher *asked* students to tell her how many sides and angles the triangles had. When one child answered correctly, the teacher followed up by asking the child to explain how they knew that. In this way, the teacher not only increased the students' participation but also asked a higher-order question to support students' conceptual understanding.

Now consider observation #03, another change to student participation that was judged as hurting effectiveness. In this lesson, the teacher's guide told teachers to ask the class for different ways to make 20 and then write the equations on the board. Instead, the teacher came up with different combinations herself and wrote them on the board without asking students first. This modification hurt effectiveness as students were not given an opportunity to construct equations themselves; the change in participation shifted students from active learners to passive receivers of information.

These two examples illustrate the type of detail that was acquired from the observations through the lens of modifications. From this data, we were able to better understand the circumstances around when teachers made each type of modification, whether or not it increased or reduced the lesson's effectiveness. We were able to make decisions around which modifications we wanted to encourage and support, as well as those we wanted to try to limit. These patterns were rich enough to guide revisions of the teaching and learning materials to maximise their usefulness[1].

As useful as the modification lens was, something was missing. As we analysed the data, the question that repeatedly came up was, "Why?" Why did the teacher ask a higher-order question at a certain moment? Why did the teacher skip that part of the lesson? Was it because something in the guide was unclear? How were teachers using their own beliefs and knowledge of students to interpret the guide and at times to chart their own course of action? These "Why?" questions would form the heart of a subsequent study in Malawi.

Case Study 2: How and why did teachers modify while teaching lessons?

In this section, we draw on data from a comprehensive qualitative study done under the USAID Malawi Early Grade Reading Activity (EGRA) (Mattos & Sitabkhan, 2016). Like the study above, the purpose of this study was to understand how teachers were using the existing teacher guides to inform revisions to both the guides and upcoming teacher trainings and coaching support. A team of two international researchers and three Malawian qualitative researchers developed observation protocols focused on capturing modifications, as well as interview protocols where teachers were asked after the lesson why they had made certain modifications. A small sample of grade 1 and 2 teachers (n=17) in six schools across three districts was observed over the course of two to three days each as they taught reading lessons. The schools were purposively selected by project staff to represent a crosssection of teacher performance in delivering the intervention; two schools were considered "high performing", one was "medium performing", and three were "low performing."

We observed a total of 886 modifications across the 17 teachers' 63 lessons. We then used an iterative process to code all the modifications, drawing from the types of modifications we found in Liberia. We found the codes as defined in Liberia to be too narrow for the expansive data set in Malawi and initially created many different codes that turned out to be too difficult to parse out. We therefore combined several codes together to be more inclusive of actions we viewed as similar and ended up with three overarching patterns in the data: structural modifications, where teachers modified the structure of the lesson in some way, such as changes to grouping structures (e.g., pair work, whole class, small groups, large groups) or to the gradual release model of the scripted lessons (i.e. "I Do, We Do, You Do" approach); content modifications, which consisted of additions or omissions to the content of each part of the lesson; and finally, classroom management modifications, such as singing a song to get students' attention, or redirecting students who were not paying attention. We found that content modifications were the most frequent type, accounting for 52 % of all modifications. The structural modification accounted for 23 % of all modifications, with classroom management accounting for the remaining 25 %. As in Liberia, we then added the subjective codes of "helped effectiveness," "hindered effectiveness," or "neutral" to each modification to better understand how the types of modifications that teachers were making were impacting the goals of the project as manifested in the lessons.

After coding all the modifications, we then turned to the teacher explanation data in order to understand why teachers were making the modifications we had recorded. Given the number of modifications, we

1 Revisions were started, but unfortunately due to the Ebola epidemic the project ended early and revisions were not able to be finished.

had not asked teachers to explain all of them after the lesson; of the 886 modifications, we collected teacher explanations for 188. After several rounds of coding, we identified eight patterns in teacher explanations of student justifications. Examples of some of these patterns can be seen in Table 1 below. Importantly, the pattern related to the teacher "forgetting" to do something was mentioned in fewer than 10 instances; the majority of modifications were the result of conscious decisions on the part of the teacher, suggesting that teachers were actively thinking about and modifying the lessons to enhance their teaching. For example, teachers added extra words to blend because they thought their class needed more practice with that skill, or they changed the structure of the lesson because they thought the content was too difficult and students were not ready to try something out on their own.

Table 1 *Teacher explanation exemplars*

Pattern	Description	Example
Too Difficult	Teachers said they modified the lesson because either the content and/or the structure were too difficult for the students.	Teacher 11 (Standard 1, Chichewa lesson, low-performing school) was teaching a lesson about segmenting words. She omitted the You Do section and explained that she thinks the You Do is too difficult for Standard 1 students to do on their own.
Inappropriate Content and/or Structure	Teachers said they modified the lesson because the content and/or structure were either too easy, or somehow inappropriate or missing an element.	Teacher 14 (Standard 1, English lesson, high-performing school) was teaching the activity "Naming objects beginning with the letter h." During the We Do section, instead of asking 2 learners to come to the front of the classroom and ask and answer questions, the teacher called on various students to point to words on the wall that start with the letter h. The teacher explained she thought learners should see the letter h in the context of a whole word.
Extra Practice	Teachers said they modified the lesson because students needed extra practice. These modifications were in both content (e.g., adding extra words) and structure (asking multiple students to demonstrate), and included checking for understanding.	Teacher 1 (Standard 2, Chichewa lesson, medium-performing school) was teaching a lesson on reading words that begin with the letters mf. The teacher added 5 words during the I Do section and explained that one example was not enough for students to learn.

The explanations of why teachers made modifications provided us with rich and detailed information on how teachers were using the lesson plans and provided actionable items to improve the design and content of the lesson plans and teacher training. For example, some recommendations that came out of the study were "through messaging in teacher training and coaching, support teacher reflection on how to interpret and respond to student errors and utilise student errors as teachable moments," as well as "when revising [scripted lesson plans], explicitly make allowance in activity-planning for reasonable transition times between activities" (Mattos & Sitabkhan, 2016, p. 47). These and other insights guided the subsequent revisions of the materials and training content.

This study allowed us to flesh out methods we had initially conceived based on the Liberia study and integrate teacher voices into the analysis. Would this more mature set of methods and instruments prove useful across multiple contexts using project-developed lessons, each of which was approaching reading instruction slightly differently? And would this be useful for a study which would not allow for indepth data collection and subsequent time-consuming analysis as done in Malawi? Answering those questions affirmatively would help validate the approach and provide insights into operationalizing it at lower cost under other programmes, as well as the types of insights that can be gleaned from a larger, more systematic study. For this, we now turn to our multi-country mixed methods study.

Case Study 3:
How can the modification lens be applied to a mixed methods study?

RTI implements education projects in many countries. These projects often develop teacher guides and lesson plans to support improved instruction. In 2018, we conducted a cross-country study to understand how those teacher guides were developed and used in each context (Piper et.al., 2018). Common approaches that appeared to have been successful across contexts might form the basis for developing higher quality teacher guides and instructional materials under future projects. The emphasis was not on evaluating a specific set of lesson plans; rather, it was on comparing lesson plans across contexts and identifying patterns that would provide insight for future projects.

While this study had multiple components, the section relevant to this paper consisted of classroom observations and teacher interviews in four countries using the modifications lens and methods refined in Malawi. Given that this study would include multiple observations across multiple countries, the openended approach used in Malawi would not be possible, due to the extensive process of coding and iteratively refining the codes. This study also had a shorter timeline, where results were expected within a six-month time frame. Because of this, the researchers decided to adapt the patterns in modifications that we found during the Malawi study and look specifically for evidence of them during observations. We reasoned that, although we may miss some novel modifications, our detailed analysis of the over 800 modifications in Malawi had identified some of the most important patterns for the purposes of understanding how teachers interact with the teacher's guide. In modifying the instruments from Malawi, a goal was to make the protocol easy to use without requiring either trained qualitative researchers or extensive training for data collectors. Instead, the researchers wanted a streamlined version that would not require extensive coding and analysis post-hoc and provide insight into patterns in classrooms by country.

Two changes were made to the Malawi instruments to meet the needs of this study. First, we added a *skipped* modification. The *skipped* modification referred to skipping one entire activity of a lesson. In Malawi, skipping an activity had been integrated as part of other modifications, but here it was decided to pull it out as it was a clear and stark way to note the extent to which lessons were being used in their entirety. Second, the classroom management modification was dropped due to the study's focus on how the provided lessons plans supported teachers in implementing the desired programmatic changes, and the classroom management modifications often were not directly related to this focus.

In the revised observation protocol, the observers recorded whenever there was either a content, structural, or skipped modification, and provided a short description of that modification. As in Malawi, each modification was subjectively described as *help*, *hinder*, or *neutral*, and teachers were asked why they made certain modifications after the lesson was finished.

A total of 27 observations were conducted across 4 countries (Ethiopia, Kenya, Malawi[2], and Uganda). All observations took place with teachers trained on using teachers' guides provided by the early grade reading project in country. Each guide had various levels of scripting, from highly to minimally scripted. In each country, the data collection team consisted of one member of the research team plus one project staff member. It is important to note that the coding of the modifications for this study took place during the observation itself. This process greatly streamlined the time for the analysis needed once all data was collected and enabled us to have more observations in more countries.

Again, after each observation, the teacher was interviewed and asked why they made some of the modifications that they did. All data was entered into an Excel table and the research team together analysed the frequency of patterns and the help/hinder modification code.

2 The observations in Malawi for Case Study 2 were conducted as part of the Malawi EGRA programme. The observations in Malawi for Case Study 3 were conducted under the Malawi MERIT programme. Each project had different sets of materials.

Because of the streamlining of the data collection and coding process, the research team was able to quickly analyse the data to provide new insights. We wanted to be able to compare results across countries, so we decided to create indices of "fidelity" and "quality", based on the modifications that were recorded. Each teacher was given a score for fidelity, defined as the number of modifications made per 30 minutes of lesson observed. The higher the score, the more modifications a teacher made and the lower the fidelity to the teacher's guide. For quality, we looked at the number of modifications per 30 minutes that were aligned with the intention of the guides, a subjective measurement given by the research team including the original curriculum designers, in many cases. The higher the quality score, the more helpful the modifications were in terms of supporting the lesson objectives. The researchers created cut scores to label lessons observed as either high quality (less than 34% of modifications coded as negative), medium quality (34–66% of modification coded as negative), and low quality (more than 66% of modifications coded as negative). The same was done for fidelity.

With these scores, we were able to see patterns in modification types and fidelity and quality scores. For example, teachers in Kenya, who averaged high quality and high fidelity scores, tended to make more structural modifications (64% of all modifications) and less partial and full activity skipped modifications (6%). In contrast, teachers in Uganda, who averaged low quality and low fidelity scores, tended to have less structural modifications (36%) and more partial and full activity skipped modifications (41%). These results suggest that certain modifications, such as structural, may still allow the teacher to maintain the quality of the lesson as it was intended, whereas other modifications have a larger impact on the overall quality. More research is needed, however to determine if this pattern can be confirmed.

For our purposes, the results provided an efficient way to compare how teachers across countries were approaching the materials. These results were paired with other analyses done as part of the study (detailed analysis of the teacher guides, understanding of current student learning outcomes in projects, and patterns in teacher explanations for modifications) to fuel a conversation about what types of revisions were recommended for each country, as well as to extract high-level "best practices" for teacher guide development in general (see Piper et.al., 2018 for more detailed results as well as fidelity and quality measure descriptions).

Conclusion

The three case studies described above illustrate how we iteratively developed methods that put teacher actions and voices at the centre of implementation. It provided a way to quantify what is happening in classrooms with new materials, instead of relying only on qualitative analysis, which is time-consuming to analyse, or on solely quantitative data, which tell us *whether* teachers are using the materials but not *how* they are using them or how their own beliefs about teaching interact with the materials. The results from this research are actionable. We ended up with very detailed information that has been used to inform revisions to materials, revisions to training strategies and other forms of teacher support such as coaching, as well as guidelines for the development of future teacher guides.

Our process of developing the methods and instruments spanned many years and several iterations. In Liberia, we had an inkling that focusing on general lesson criteria through a checklist-type form was not good enough: the resulting information was too vague and not actionable. We then pivoted to a lens of modifications. In Malawi, we focused on the need to understand what motivated teachers to make modifications. Importantly, the reasons *why* teachers modified provided us with a lens into the

classroom culture that otherwise remained hidden. In the multi-country case study, we were interested in using these instruments more efficiently on a larger scale and to be able to make comparisons across diverse contexts.

Throughout these iterations, we refined methods that we believe will work for both in-depth studies as well as shorter, more rapid, mixed method studies. Taken together, the instruments offer a way to continually adapt and improve project implementation.

Importantly, we placed teacher voices and actions at the centre of the intervention. If we are to address shortfalls in education around the globe, we must understand and value the essential role of teachers. We cannot assume teachers are robots that will do everything that is asked of them in exactly the way it is asked; rather, we should make every effort to understand their knowledge and beliefs, and develop materials that build on these assets, while also incorporating new knowledge.

The studies presented in this paper provide an example of rigorous research in the field of international education development that is not solely quantitative. Often, RCTs are the gold standard in international development in terms of providing evidence on what is working. However, studies such as those presented in this paper have the potential to answer "how" questions and provide insight in why a particular programme may be working or not working. Often, qualitative work that focuses on answering "how" questions is not undertaken in international education work because of the time and cost they require. The study in Malawi, for example, required extensive data collection from experienced researchers and the follow-up analysis was labour-intensive. However, our goal in refining the methodology over the course of three studies was to arrive at a place where in-depth studies with actionable information can be done in reasonable time frames and budget limitations. The multi-country study illustrates some of the accommodations we made to allow for faster data analysis, such as coding during the observations instead of open-ended coding after the observations. Future studies should aim to find methodologies and instruments that can be used to collect qualitative data that fits within the constraints of international development work.

Limitations and next steps

This methodology comes with some caveats. The case studies that we used to develop and refine in Liberia and Malawi were small and exploratory and required expert researchers to be part of the research team. The multi-country case study was not exploratory but also required at least one experienced researcher to oversee data collection and analyse the data. To get actionable information, researchers and data collectors need to be trained and highly skilled, so that they understand the classroom as well as the behaviours being targeted by an intervention.

As noted above, we did not separate out teacher insights by category (knowledge, beliefs, or attitudes). It would be beneficial in future work to tease out these categories as it would provide programmes with specific and detailed information for improving performance. Addressing a gap in teacher knowledge may require a different approach than addressing a belief about how students learn. Some beliefs are more malleable than others. A future study can use this methodology to conduct a detailed analysis of the categories of explanations that teachers give and link them to the decisions they make in class.

Alternatively, the methodology might be adapted to understand contexts before materials are developed or interventions are even designed. If we can better understand classroom culture and current classroom practices, and how teachers' knowledge and beliefs are manifested during instruction, we can design interventions that are truly specific to the context and that may have higher chances of uptake and sustainability, ultimately leading to better learning outcomes.

References

Brickhouse, N. W. (1990). Teachers' beliefs about the nature of science and their relationship to classroom practice. *Journal of Teacher Education, 41*(3), 53–62

Brombacher, A., Nordstrum, L., Davidson, M., Batchelder, K., Cummiskey, C., & King, S. (2014). *National baseline assessment for the 3Rs (reading, writing, and arithmetic) using EGRA, EGMA, and SSME in Tanzania.* Retrieved from https://pdf.usaid.gov/pdf_docs/pa00k5cn.pdf

Chachage, K. (2020). *Pedagogy as social practice and teachers' pedagogic choices in Tanzanian primary schools* [Doctoral dissertation, University of Minnesota]. University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/213097

Chemonics International. (2020). *Teacher knowledge, attitudes, and practices regarding early grade reading: Baseline study report – Version B.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00X338.pdf

Cohen, D.K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 311–329.

Cuban, L. (2021). *The Complexity of Teacher Decision-making.* (blogpost) https://larrycuban.wordpress.com/2021/04/28/THECOMPLEXITY-OF-TEACHER-DECISION-MAKING/

EnCompass, LLC. (2017). *Time to Learn Endline Evaluation Report.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00MK22.pdf

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching, 15*(1), 13–33.

Freudenberger, E. & Davis, J. (2017). *Tusome external evaluation – Midline report.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00MS6J.pdf

Harden, K., Sowa, P., & Punjabi, M. (2019). *2019 Language Usage Study in Bahasa Sug, Chavacano, Magindanawn, and Mëranaw Mother Tongue Schools. Prepared for USAID under the All Children Reading-Philippines Project, AID-OAA-TO-16-00017. Research Triangle Park, NC: RTI.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00WKGZ.pdf

Hertz, Ashley C., Kochetkova, E., & Pflepsen, A. (2019). *Classroom Observation Toolkit for Early Grade Reading Improvement. A Global Reading Network Resource.* Retrieved from www.globalreadingnetwork.net.

Lugo-Gil, J., Contreary, K., Murray, N., Glazerman, S., Campuzano, L., Fernandez, C. & Place, K. (2016). Latin America and the Caribbean (LAC) *Reading Evaluation Contract: Evaluation of Leer Juntos, Aprender Juntos early grade intervention in Guatemala. Final baseline report.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00MMTD.pdf

Mattos, M. & Sitabkhan, Y. (2016). *Malawi Early Grade Reading Activity: Scripting study report.* Retrieved from https://pdf.usaid.gov/pdf_docs/pa00mdzf.pdf

Messaoud-Galusi, S., Mulcahy-Dunn, A., Ralaingita, W., & Kochetkova E. (2012). *Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Doukkala Abda, Morocco.* Retrieved from https://pdf.usaid.gov/pdf_docs/pnadz048.pdf

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*(4), 317–328.

Pajares, M.F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.

Piper, B., Sitabkhan, Y., Mejía, J., and Betts, K. (2018). *Effectiveness of teachers' guides in the Global South: Scripting, learning outcomes, and classroom utilization.* RTI Press Publication No. OP-0053-1805. Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2018.op.0053.1805

RTI International. (2019). *Read Liberia Activity: Cost-effective, evidence-based coaching model.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00W5XQ.pdf

RTI International. (2018). *Early Grade Reading (EGR) Time-on-task study report.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00T59S.pdf

RTI International. (2015). *Nigeria Reading and Access Research Activity (RARA): Results of an approach to improve early grade reading in Hausa in Bauchi and Sokoto States.* Retrieved from https://pdf.usaid.gov/pdf_docs/PA00KVM1.pdf

RTI International. (2014). *Task Order 15: Data for Education Programming in Asia and the Middle East (DEP/AME): Nepal education management efficiency study and teacher observation study: Final report.* Retrieved from https://pdf.usaid.gov/pdf_docs/PBAAF536.pdf

Schoenfeld, A.H. (2008). Chapter 2: On modeling teachers' in-the-moment decision making. *Journal for Research in Mathematics Education. Monograph,* 14, 45–96.

Serbessa, D.D. (2006). Tension between traditional and modern teaching-learning approaches in Ethiopian primary schools. *Journal of International Cooperation in Education, 9*(1), 123–140.

Smart, A. (2021). Teachers' guides: isn't that what they should be? *IARTEM e-journal, 13*(1).

Stipek, D.J., Givvin, K.B., Salmon, J.M., & MacGyvers, V.L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*(2), 213–226.

United Nations. (2015). *Transforming our world: the 2030 agenda for sustainable development.* Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

Wilkins, J.L. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139–164.

World Bank Group. (2017). *Conducting classroom observations: Stallings 'classroom snapshot' observation system for an electronic tablet.* Retrieved from https://openknowledge.worldbank.org/handle/10986/28339