What Drives Teachers to Change Their Instruction? A Mixed-Methods Study from Zambia

Andreas de Barros[1], Junita Henry[2], and Jacqueline Mathenge[2]

[1]Department of Economics, Massachusetts Institute of Technology

[2]Teaching at the Right Level (TaRL) Africa

Author Note

Andreas de Barros https://orcid.org/0000-0003-2154-0412

Junita Henry https://orcid.org/0000-0002-5600-7119

This research was supported by the Global Partnership for Education Knowledge and Innovation Exchange, a joint endeavour with the International Development Research Centre, Canada through Grant No. 109295-001 to the Massachusetts Institute of Technology. The study has been registered with the Open Science Framework (OSF). The authors gratefully acknowledge support from Pranav Bharghava, Sharnic Djaker, Maimuna Ginwalla, Isaac Mbiti, Sebastian Muñoz-Najar, Laura Poswell, Nicolás Riveros, Tavneet Suri, and Nico Vromant. The authors have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Andreas de Barros, Abdul Latif Jameel Poverty Action Lab, Department of Economics, Massachusetts Institute of Technology, 400 Main Street, Suite 201, Cambridge, MA 02142. E-mail: debarros@mit.edu

Abstract

There is limited evidence on what drives teachers to change their teaching practices. Using primary qualitative data from 78 Zambian education personnel from the school to provincial level, we combine thematic analysis with unsupervised machine learning to identify drivers of pedagogical shifts. We then combine qualitative analyses with linear probability models to uncover their associations with teacher professional development. Our findings suggest that teaching practices are malleable, with change being predominantly driven by on-site continuous professional development (CPD) opportunities relating to team-based problem-solving, verbal discussions, and skills acquisition. Taken together, this study highlights the potential of school-based CPD opportunities as means to alter teaching practices, in a developing-country setting.

Keywords: continuous professional development, drivers of change, mixed-methods, pedagogical shifts, Zambia

What Drives Teachers to Change Their Instruction?
A Mixed-Methods Study from Zambia

High-quality teaching is a key determinant of student success. A growing body of literature documents how a large proportion of classroom-to-classroom variance in student performance can be attributed to teachers' teaching practices (Araujo et al., 2016; Azam & Kingdon, 2015; Bau & Das, 2017; Buhl-Wiggers et al., 2018).[1] Beyond test scores, teaching quality is also a main driver for the development of socio-emotional skill (Jackson, 2018) and other long-term life outcomes (Chetty et al., 2014; Jackson et al., 2014; Rivkin et al., 2005). At the same time, teachers in many less-developed countries may lack the necessary skills to teach effectively and use teaching methods ill-matched to their students' diverse needs (Bietenbeck et al., 2018; Bold et al., 2017; Bold et al., 2018)—even in countries with comparatively high teacher pay (de Ree et al., 2018; Ramachandran et al., 2018).

In many less-developed countries, governments and international organizations have, therefore, made it a priority to improve teaching quality through professional development for teachers. In a review for anglophone Africa, for instance, all countries had a national in-service training system for teachers (Mulkeen, 2009). Internationally, cooperation for teacher training and in-service development has been recognized as a formal target of the United Nations Sustainable Development Goals (Target 4c). For example, of the World Bank's education projects, approximately two-thirds include professional development for teachers (Popova et al., 2021) and, between 2013 and 2018, the Bank allocated US$12.1 billion towards these projects (The World Bank, 2018b).

Yet, at the same time, this focus on professional development for teachers operates in a context where teacher training has been declared a failure. For example, the 2018 World Development Report concluded that "most teacher training is ineffective" (The World Bank, 2018a, p. 131). Similarly, the 2020 Global Education Evidence Advisory Panel labelled the

[1] In contrast, observable characteristics of teachers (rather than their teaching practices) are often considered a poor predictor of student learning (*ibid.*).

most common forms of in-service teacher training a "bad buy" for policy makers in low- and middle-income countries (The World Bank, 2020, p. 21). It is within this tension that we set out to study how professional development may effectively impact teaching practices in a low-resource setting, at scale.

Conceptualizing In-Service Professional Development for Teachers

Our study conceptualizes in-service professional development for teachers as opportunities for professional learning that may cause improvements in instructional quality and, thus, increased student learning. Our focus on in-service development distinguishes these learning opportunities from others that may be provided as part of teachers' pre-service preparation, apprenticeship period, or induction.

We highlight three key features of this conceptualization (cf. Darling-Hammond et al., 2017). First, we understand in-service professional development for teachers as learning opportunities that may be both externally provided or job-embedded. Thus, learning opportunities may not be "done to" but rather actively co-created by teachers. Secondly, our understanding recognizes that learning may happen during dedicated periods or be continuous. Thus, learning may not be constrained to workshops or training events, but can also occur as prolonged, ongoing processes. Third, we highlight that these learning opportunities can be diverse, both formal or informal. Thus, our conceptualization goes beyond "traditional" in-service teacher training activities such as off-site events, and also encompasses collaborative approaches such as on-site peer learning and coaching, for example.

Our theoretical framework of how in-service professional development for teachers may effectively lead to student impacts is most closely related to Desimone's (2009) model.[2]

[2] It also relates to related models by Fishman et al. (2003), Guskey (2002), Opfer and Pedder (2011), Supovitz and Turner (2000), and Timperly et al. (2007). For an overview and discussion of these models, see McChesney and Aldridge (2019).

Accordingly, we focus on a theory of change whereby (a) professional development activities are expected to lead to (b) changes in teachers' attitudes, knowledge, and skills, which in turn affect (c) changes in teaching practices, that (d) impact student learning. However, we depart from Desimone's (2009) model as it posits that five features define professional teacher development (content focus, active learning, coherence, duration, and collective participation). Rather, as we move away from the context of developed countries, we aim to explore whether these—or other—components feature prominently in those observed development activities that are associated with changes in instructional behavior.

Advances in In-Service Teacher Development in Less-Developed Countries

This study intends to contribute to a nascent body of literature that aims to identify novel in-service teacher development activities that work effectively in less-developed countries. One strand of this literature echos findings from the United States (Kraft et al., 2018), which point to the effectiveness of teacher mentoring and coaching as promising means to improve instruction and raise student achievement. For example, Cilliers, Fleisch, Prinsloo, et al. (2020) document how a South African teacher coaching program led to 0.24 standard-deviation (SD) increases in mother-tongue language and reading proficiency among early-primary grade students.[3] Similarly, Castro et al. (2019) and Majerowicz and Montero (2018) find that a Peruvian teacher coaching system led to 0.25SD improvements in reading comprehension and 0.38SD improvements in mathematics among second graders. Yet another example comes from a coaching program in secondary schools in Brazil, which led to 0.05-0.09 SD improvements in grade-10 mathematics and Portuguese and 0.06SD improvements in grade-12 Portuguese (Bruns et al., 2018).

A second strand of this literature suggests teacher development related to structured pedagogy can be an impactful tool to improve teaching quality and student learning (Conn,

[3] For additional (positive) results from an earlier South African primary school coaching program, see Harvey (1999).

2017; Evans & Popova, 2017; Snilstveit et al., 2015). These learning opportunities are usually centrally designed and include teachers' guides and lesson plans with accompanying teaching and learning materials. Often, these materials are practice-based, (at least partially) scripted, and linked to student materials and textbooks. For example, Piper, Destefano, et al. (2018) find that the Kenyan national literacy program "Tusome" led to large (0.6 to 1 SD) impacts in English and Kiswahili, and Piper, Simmons Zuilkowski, et al. (2018) find structured teachers' guides were a significant driver of this impact. Similarly, structured pedagogy is a key component of teacher development programs that have been found to be effective in Brazil (Leme et al., 2012), the Gambia (Eble et al., 2021), and the Philippines (Tan et al., 1999).

Finally, a third strand of this literature focuses on teaching content that allows instructors to adjust their classes to students' learning level (rather than students' age or grade-level curriculum). These teacher development activities recognize that, in less-developed countries, many students lack foundational skills and lag far behind their respective at-grade content (Azevedo et al., 2021). A series of large-scale randomized controlled trials from India suggests training teachers to target instruction to students' learning levels can effectively improve child learning (Banerjee et al., 2017). Subsequently, revised versions of this approach have been adopted in Botswana, Ghana, Ivory Coast, Madagascar, Mexico, Nigeria, Senegal, and Zambia (Alcott et al., 2018; Duflo et al., 2020), and it is now considered among the most promising strategies to improve student learning in less-developed countries (Angrist et al., 2020).

Gaps in the Literature

Taken together, these advances connect to three of the core dimensions of professional teacher development as identified in the US education literature, concerning its format (coaching), activities (working with structured pedagogy using teaching guides), and focus (learning how to target instruction to students' learning levels) (cf. Hill et al., 2020). At the

same time, however, at least two gaps in the literature limit our understanding of how to successfully deploy in-service professional development for teachers in less-developed countries. Here, we briefly discuss these two limitations.

Identifying Drivers of Change in Teaching Practices. For less-developed countries, there is limited research as to which drivers lead teachers to change their instructional behaviors. We recognize that Desimone's (2009) five program features that are expected to successfully drive change, as identified in the US, may not readily transfer to other contexts (Henrich et al., 2010). For example, a recent review of teacher professional development programs in less-developed countries does not confirm teachers' participation and discussions among teaching staff as predictors of improvements in instruction and student learning (Popova et al., 2021).[4]

While this observation calls for work concerning the model's external validity, we also recognize how additional model elaboration may be required. In particular, the model may be incomplete as it does not identify processes that facilitate linkages between model components (that is, between teachers' participation in professional development activities, changes in teachers' attitudes, knowledge, and skills, and changes in teaching practices) (cf. King, 2014).

In-Service Teacher Development That Remains Effective at Scale. The second limitation revolves around how to identify in-service teacher development that remains effective once it is observed at scale. Consistently, teacher development programs have been found to be less effective (or even detrimental) if implemented without researcher oversight, once substantial external supports are removed, and when responsibilities are transferred from a non-governmental organization to the government (Popova et al., 2021). This observation holds for each of the three areas of professional development we identified above—that is, for coaching programs (Albornoz et al., 2020; Cilliers, Fleisch, Kotzé, et al.,

[4] Yet, the same review also notes how models that stress collaboration among teachers (such as "communities of practice") have not yet been rigorously explored in less-developed countries (cf. Kennedy, 2019).

2020), programs involving scripted lesson plans (Kerwin & Thornton, 2021), and programs promoting that teachers target their instruction to a child's learning level (Banerjee et al., 2017; Duflo et al., 2020). Thus, one can expect large knowledge gains from research that avoids related implementer effects (Vivalt, 2020), site-selection effects (Allcott, 2015), or publication bias (DellaVigna & Linos, 2020) by observing at-scale programs that are implemented under government oversight.

The Current Study

This is an exploratory study with three main aims. The aims of this study consisted of (Aim 1) verifying that the selected context is one in which public school teachers were likely to have changed their instruction, (Aim 2) identifying key drivers that had led to these changes, and (Aim 3) investigating to what extent these drivers were associated with in-service teacher development activities that operate at scale.

To achieve the study's three aims, we used an embedded mixed-methods design, whereby quantitative analytical methods are concurrently nested within a broader qualitative research project.[5] More specifically, we began by purposely selecting a context in which teachers were likely to have changed their instructional behaviors (with variance thereof). Next, we generated qualitative data, through in-depth telephonic interviews. Thereafter, during thematic analysis, we integrated unsupervised machine learning (treating text as quantitative data) with open coding (treating text as qualitative data), observed the extent to which the two methods converged, and thus generated a coding framework. We hand coded all responses following this framework, and identified themes of what reportedly drove teachers to change their instructional behaviors. Finally, we investigated how these

[5] Our use of a single data-set does not fit common convergent designs of mixed-methods research (which usually distinguish qualitative from quantitative data sources). Our exploratory sequencing of analytical steps with a single sample also does not fit common exploratory sequential designs (which usually distinguish qualitative from quantitative samples). See Creswell and Plano Clark (2018), for a discussion of "convergent" and "exploratory sequential" mixed-methods designs.

drivers of change were associated with in-service teacher development activities, both qualitatively and quantitatively, and report the results of the two approaches side-by-side.

Method

We conducted this research with Internal Review Board approvals, both in the United States and in Zambia, in accordance with ethical guidelines for the protection of research participants.

Setting

We conducted this study in a developing-country setting in which (a) teachers had been expected to change their instructional behaviors, and which (b) allowed us to observe a wide array of in-service teacher development opportunities, in public schools, at scale.

Our research took place in two provinces of Zambia (Eastern and Southern Provinces), in public and community primary schools. Teachers had been exposed to two large-scale programs, reaching more than 1,800 schools in these provinces, under government oversight. One program introduced teachers to a "teaching at the right level" approach, which entails grouping learners according to their basic numeracy or literacy level rather than by grade. The program focuses on grades three to five and it is locally known as *Catch Up*. Another program introduced teachers to a simplified five-step literacy program. The program runs from pre-primary through grade three, and it is called *Let's Read*.[6]

Teachers were also (expected to be) exposed to a large array of in-service development opportunities—both through these additional programs and within Zambia's national in-service development scheme. More specifically, the two programs held off-site training workshops and provided regular on-site mentoring and monitoring during school visits. In addition, Zambia's Ministry of General Education established the *School*

[6] Moreover, the government promoted a few small-scale literacy programs such as the *Teaching Handwriting and Spelling Skills (THRASS)* program.

Programme of In-Service Training for the Term (SPRINT) system to carry out both off-site review meetings and biweekly, school-based continuous professional development meetings.

Participants

Our sampling strategy proceeded in two steps. First, we randomly sampled schools. To represent variety in geographical regions, student performance, and school type, we stratified schools based on (1) their province (Eastern; Southern), and (2) above- and below-median student achievement. Student performance was calculated using *Catch Up* assessment data from 2020, and we excluded schools not running the program.[7] We then randomly selected two public schools and one community school in each of the four strata, for a total of 12 schools.

Secondly, we sampled staff that either work in or support the sampled schools. Within each school, we sampled a mathematics and a literacy teacher who taught in grades three to five, as well as three additional roles that are expected to oversee and support these teachers (the headteacher, the "school in-service co-ordinator", and a "senior teacher"). We also sampled the respective zonal, district and provincial coordinators, as well as another 50 percent of staff that cover the remaining zonal and district-level cadres supporting these schools. The total number of sampled roles is 103. Accounting for individuals who take on multiple roles (for example, a headteacher may also be a literacy teacher), a total of 83 individuals covered these 103 roles.

Table 1 provides the sample characteristics. The 12 randomly selected schools cover a geographic region of 10 districts and 12 zones across the two provinces. A little over half (58%) of the schools are located in a rural district. Of the 83 sampled individuals, we were able to interview 78 respondents, for a non-response rate of 6%. A little less than half (45%)

[7] At the time of the study, in the two provinces, the *Catch Up* program ran in 83.4 percent of government-run schools. It ran in 33.3 percent of private, community-run schools. Government schools serve the vast majority of primary-school students (85.6 percent of students who attend government and community-run schools).

of the respondents are female. The average respondent has 4.7 years of experience within their interviewed role. The majority of respondents (71%) are based at the school level (as opposed to higher-level support staff).

Data Collection

Participants were recruited via telephone calls. Interviews were conducted by the second and third author during February and March 2021. Interviews were audio-recorded upon receiving verbal consent from participants. Interview time ranged from twenty minutes to two hours and was 57 minutes in length on average. Upon completing data collection, interviewers manually transcribed interviews verbatim from audio recordings. Respondents were compensated with mobile airtime for participation.

Each week, the research team debriefed to share key information on the data collected. The team used this time to discuss the data and determine whether any adjustments on collection and refinement needed to be made. Interview questions were adjusted where appropriate, to probe deeper into systems and methods the research team did not have a clear understanding of. Additionally, interviewers kept field notes that were written during the data collection process. Reflexivity was an important aspect of the research process and we kept a self-critical account of the process in the form of memos (Nowell et al., 2017). We documented the daily logistics of the research, methodological decisions, and rationales, as well as personal reflections and insights.

The phone interviews followed a semi-structured interview guide. In each interview, we asked respondents how teachers had changed their instruction before the COVID-19 pandemic. Specifically, the leading question was: “In the year before the COVID-19 crisis: Do you think you changed the way you (/ your school’s teachers) went about your (/their) day-to-day teaching in the classroom? If so, how?” In the case of a COVID-19-related response, interviewers were encouraged to ask the question again. We then asked respondents what provoked these changes or what aided teachers in making these changes easier to adapt.

Data-Analytic Strategy

The study's mixed-methods data-analytic strategy embeds quantitative text analysis that exploits unsupervised machine learning within qualitative thematic analysis. To achieve the study's first aim, we identified themes of change in instructional behavior and documented how frequently different changes occurred. To achieve the study's second aim, we identified themes that describe what drove teachers to make these changes. Lastly, to achieve its third aim, we identified themes concerning teachers' engagement in in-service professional development and investigated their association (both quantitatively and qualitatively) with these drivers of change in instruction.

Our thematic analysis expands on the six-step analytical framework proposed by Braun and Clarke (2006). After (1) familiarizing ourselves with the data, we (2) generated initial codes, (3) searched for themes, (4) reviewed these themes, and (5) defined and named them. We then (6) conducted our final analysis and produced the present article. We conducted the first three steps both qualitatively, through "open coding", and quantitatively, through unsupervised machine learning. In the fourth, fifth, and sixth step, we qualitatively analyzed how the qualitative and quantitative methods converged. Our overall approach is largely inductive and "data driven", but it includes deductive elements as we investigated a predetermined set of research questions.

Identifying Themes Through Qualitative Open Coding. To identify themes qualitatively, we treated the interview transcripts as qualitative data to conduct open coding. We open coded all responses following an inductive approach, without a preconceived coding scheme. For this coding, we unitized excerpts at the sentence level unless the subsequent sentence(s) conveyed the same meaning or code application. Next, we searched for themes by condensing and collating these codes into groups, omitting codes that occurred less than five times throughout all interviews. Thereafter, we generated themes from the condensed codes to develop a preliminary coding framework.

Identifying Themes Through Unsupervised Machine Learning. To identify themes quantitatively, we treated the interview transcripts as quantitative data to estimate topic models. Topic modelling is a class of unsupervised machine learning methods. The method infers topics, or distributions over words, that represent semantically interpretable themes. More specifically, we rely on a mixed-membership model, whereby individual documents and their words can belong to multiple topics (instead of just one). We conducted our analyses in the R software, using the "quanteda" and "stm" packages (Benoit et al., 2018; R Core Team, 2020; Roberts et al., 2019).

As we familiarized ourselves with the data, we removed any text spoken by the interviewers and generated an answer-level corpus (containing 5,879 answers). We then used Rapid Automatic Keyword Extraction (RAKE) to add two- and three-word phrases (bigrams, trigrams) to the words respondents had spoken. In turn, we removed a list of common stop words (such as "me", "my", "myself", "we", or "our"). We also set all words to lower case and "stemmed" them (e.g., by removing suffixes such as "ed", "ing", or "ly").

To decide on the number of topics to extract, we trained a range of topic models with varying numbers of topics. We then evaluated their performance using common diagnostic plots, judging models by their residuals, semantic coherence, and exclusivity. Our preferred model identified 45 topics. We present their prevalence and their most distinctive terms in Appendix Figure A1.

Convergence and Qualitative Closed Coding. To develop the final coding scheme, we compared themes generated through qualitative hand coding to topics generated by the computational approach. Table A1 in the Appendix shows the extent to which quantitatively identified themes matched qualitatively identified themes. We do not find strong overlap, with almost half of the machine learning themes remaining uninterpretable. We find stronger convergence for codes and themes that are related to drivers of change in teaching behavior.

During this process, we met bi-weekly to discuss and adapt the framework, which was

refined three times before establishing the final code book. Our final coding framework was a family code-based scheme. "Parent" codes were theme-based and we developed several "child" codes. For example, a "parent" code called "drivers of change" contained several "child" sub-codes pertaining to the various drivers of change mentioned (such as "sharing and discussing challenges").

Final Analysis and Reporting. As with open coding, we used a combination strategy to define units of analysis. We used naturally given units and the meaning of units. We unitized at the sentence level (naturally given unit), but this extended to subsequent sentences if they conveyed the same meaning/code application (meaning of unit). These units, or "meaningful responses" are referred to as excerpts.[8]

In our final analysis, we report on the incidence of excerpts related to changes in teaching behavior. That is, we calculated the percentage of various codes related to change out of the total number of codes related to changes in teaching behavior (Aim 1). We also report on the respective percentage of codes related to drivers of change (Aim 2). We do not report on excerpts that occurred less than five percent of times.

To identify associations between drivers of change and in-service professional development activities (Aim 3), we conducted qualitative and quantitative analyses side-by-side, independently. In qualitative analyses, the study's second author rated the relationship of excerpts related to the three main drivers of change with the data's teacher professional development activities as low, medium, or high. To this end, the author qualitatively analyzed all excerpts related to professional development and their relation to the various inputs within interviews. In quantitative analyses, the study's first author used a series of bivariate linear probability models to quantify these relationships. These analyses estimate the percentage-point increase (decrease) of a driver's occurrence in any excerpt, as a function of whether the same excerpt, or another excerpt immediately before or after, is coded with a given teacher professional development activity. Here, we define "immediately

[8] On average, we identified 71 excerpts per interview.

before or after", as a sliding window of ten excerpts around (and also including) the excerpt in question.[9] To assess the robustness of findings, we also show results for a wider window of twenty excerpts.

Methodological Integrity

We double-coded a random 20% of interviews (16 interviews). We followed Campbell et al. (2013) where one knowledgeable coder identified these meaningful units of analysis, and another coder re-coded the same excerpts. Interviews were coded by the second author and reliability was established by the first and second author.

We determined reliability by dividing the number of times that the initial coder and the second coder used a code (agreements) by the number of times that any coder used it (sum of agreements and disagreements). This is known as percent agreement.[10]

Our codes and re-codes showed 80 percent agreement. Although there is no universally accepted threshold for what indicates acceptable reliability, Miles and Huberman

---

[9] More formally, let Y_{ir} be a binary indicator of whether excerpt i of respondent r was coded with a driver of interest or not (e.g., "sharing challenges"). Let $D_{ir}^{-9,0}$ indicate whether anywhere in a window of nine preceding excerpts up to the same excerpt a professional development activity of interest (e.g., "offsite training") was observed ($D_{ir}^{-9,0} = 1$) or not ($D_{ir}^{-9,0} = 0$). More generally, let integer j denote the first excerpt of a sliding window of ten excerpts around (and including) i, where j reflects the window's starting position relative to excerpt i. We can then calculate the difference in conditional expectations for any sliding window with $[j] \in \{-9, ..., 0\}$ as $E[Y_{ir}|D_1^{j,j+9} = 1] - E[Y_{ir}|D_1^{j,j+9} = 0]$. In the article, we report on the average difference in conditional expectations $\frac{1}{10}\sum_{j=-9}^{0}(E[Y_{ir}|D_1^{j,j+9} = 1] - E[Y_{ir}|D_1^{j,j+9} = 0])$. Finally, for ease of interpretation, we multiply by 100 to report on percentage point differences instead of proportions.

[10] Although this approach does not take into account the possibility of an agreement by chance, other measures that do—such as Krippendorff's coefficient—rely on the assumption that all codes have an equal probability of being used. In our case, this was not appropriate. Not all questions applied to all interviews (depending on the role, time constraints or network connectivity of interviewee); therefore, not all codes may apply to all interviews. Furthermore, we had a large coding framework, which reduced the likelihood that coders agreed by chance (Bernard, 2013; Campbell et al., 2013).

(1994) have suggested a standard of 80 percent agreement.

Results

We report the study's results in three steps. First, we document changes in teaching behaviors. Second, we extract the top three drivers that reportedly provoked these changes. Third, we investigate how at-scale teacher development activities are associated with these primary drivers of change.

Changes in Teaching Behaviors

Table 2 shows the number of respondents who reportedly changed their instructional behavior, along with the incidence of excerpts referring to change (across 236 excerpts). Altogether, 64 respondents reported on changes in their classroom; only 14 of them did not. More specifically, 77.7% of the excerpts refer to technical (or pedagogical) changes, and 16.5% refer to changes in teaching that relate to teachers' attitude or confidence in the classroom. These results lay the foundation that the study operates within a context of changing teaching practices.

Table 2 provides greater details as to what types of pedagogical changes dominate in our sample. Most often, teachers reportedly increased their use of differentiated instruction, where teachers teach to the ability of the learner (18.6% of excerpts). This is followed by an increased use of teaching and learning materials in the classroom (15.3%). Respondents also frequently referred to the use of the Catch Up program's methodology not only in the program's dedicated classes, but also in students' regular, non-Catch Up classes (14.4%). Other changes relate to teachers' increased interaction with learners (7.6%), a greater use of lesson planning and preparation (7.6%), and students' engagement in in-class activities, such as group work or games (5.1%). These changes co-occurred with changes in teachers' attitudes (8.5%) and changes in teachers' intrinsic motivation (8.1%).

We then disaggregate how teachers' instruction reportedly changed across programs (Catch Up, SPRINT, Let's Read, and other programs). By program, we find that changes

related to the "Catch Up" program strictly dominate across teaching behaviors.[11] Catch Up was most frequently related to changes in pedagogy, as about half of the excerpts refer to modifications in classroom instruction due to this program (49.6% overall, or 64.6% of the excerpts related to technical changes). The program is most frequently associated with teachers' increased understanding of learners' needs and differentiated instruction, an increased use of teaching and learning materials, and increased interactions between teachers and students. Respondents also noted how these changes spill over to other classes that are not directly targeted by this program.

Drivers of Change

In Table 3, we then examine what provoked these changes in instructional behaviors, by grouping mechanisms into primary drivers (top three, as per their prevalence across excerpts) and secondary drivers that were mentioned less frequently. The table summarizes 825 excerpts in which respondents explained what had led to (or constitued a "driver" for) the aforementioned changes.

Team-Based Problem-Solving of Challenges through Group Discussions. The most frequently discussed driver of change relates to sharing of challenges and group discussions (11.2% of excerpts referring to drivers of change). This is typically a form of peer mentoring where teachers problem-solve challenges that they are struggling with together. In our sample, 51 out of 78 respondents mentioned this mechanism as a driver of change. A teacher states that it is "because teachers are able to come together to find solutions and support each other. You know when you face a challenge but you do not talk to people, you can do nothing. So, working together during these meetings helps." These discussions were commonly described as being inclusive of all teachers, and they reportedly offer an

[11] This article does not intend to compare the relative effectiveness of various programs. Also recall that the study sampled teachers who taught in grades three to five; other programs with a different grade-level focus may not expect to affect teaching behavior in these grades.

opportunity to identify challenges and use existing knowledgeable personnel who are easily accessible to "sharpen" one another. One teacher mentions that "everyone comes with a certain strength and on the other hand certain weaknesses which can be worked on. Especially if I don't know how to handle certain areas, I find people who can assist there and if I have strength and other teachers don't have that strength, I guide them on how to go about it. So, everyone has a certain weakness and a certain strength. Not everyone can have all strengths or all weaknesses but we share what we can share. Then we see what we need to get from others."

Acquiring New Skills and Learning New Methods. The second most frequently mentioned driver of change relates to teachers' acquisition of new skills and teaching methods (10.3% of excerpts). Within this, a subset of 6.1% of excerpts refer to the learning of new methods at training sessions (not shown in the table). One teacher stated that "knowledge is not like that, it is not static, it changes every now and then". Training equips teachers with a variety of methods and activities that they are able to choose from and use in various situations. Another respondent referred to training by saying that "these programs they have given us knowledge on how we can handle, the teachers have been [given] a basket full of activities or methods that they can pick from, they will not lack anything, they will go into this basket and pick. For example, comprehension, I will go back to my basket and pick one that I think is suitable." Respondents also explained that practical demonstrations were most helpful during training activities (2.8% of excerpts, not shown in the table).

Verbal Encouragement and Discussions. Individual mentoring in the form of verbal encouragement and discussion in a one-to-one setting is also a commonly reported driver of change, although it is mentioned less frequently than the other two drivers (7.0% of excerpts). These discussions mainly consist of identifying areas of weakness particular to a teacher and sharing solutions, ideas or new approaches they may adopt to improve. One teacher put it as "I know my weaknesses, that's why it helped me." Respondents also

mentioned that mere encouragement of program take-up was motivating during these discussions.

Weaker Drivers. Other, secondary elements that reportedly provoked teachers to change their teaching practices are learner outcomes (6.7% of excerpts), access to teaching materials such as flip charts and markers (5.2%), and frequent monitoring (5.0%). Other drivers were mentioned in less than 5% of excerpts and they are omitted from Table 3. It is worth noting that there is little difference in the frequency of excerpts related to verbal encouragement vs. learner outcomes. For simplicity, our following analyses focus on verbal encouragement; however, we note that greater emphasis should be placed on the top two drivers.

Teacher Development Activities Associated With Primary Drivers of Change

Table 4 reports on the prevalence of continuous professional development opportunities and their association with the three most commonly mentioned (or "primary") drivers of change. In terms of prevalence, we note that a large proportion of respondents mentioned teachers' participation in continuous professional development activities. While the range of these activities is broad, on-site capacity building and mentoring through lesson observations dominate the responses. For instance, about twice the percentage of excerpts mention on-site capacity building, as opposed to off-site training (1.7 vs. 0.9 percent of excerpts). On aggregate, across five broader subcategories of professional development, use of technology is mentioned the least (0.5 percent of excerpts).

In terms of associations, the results from quantitative and qualitative analytical approaches closely track each other, with few exceptions. In particular, we find strong agreement on a positive association between on-site capacity building and teachers' sharing of challenges. We also find strong agreement on a positive association between the various mentoring strategies and verbal encouragement. We further observe convergence related to the potential of both off-site trainings and one-on-one mentoring as means to introduce new

teaching methods.

Our results from the qualitative and quantitative analyses moreover agree on the limited potential of technology-based mentoring solutions to trigger the three primary drivers of change. Both approaches also suggest that off-site vs. on-site capacity building may serve different purposes, whereby the latter appears better suited to encourage team-based problem solving. Lastly, as an exception, we note that the two analytical approaches disagree on the extent to which monitoring through lesson observations induces team-based problem solving and sharing of challenges. An additional qualitative review of the interview data revealed the quantitative analysis strictly focused on respondents' description of the actual observation, and its monitoring aspect only, while not capturing short debrief discussions that commonly occur at the end of a lesson observation.

Discussion and Conclusion

In this article, we examined changes in teachers' instructional practices, main drivers that provoked these changes, and the extent to which these drivers are associated with teacher professional development activities that operate at scale, in public schools, in a less-developed country. These analyses and their findings rest on a mixed-method analytical strategy that examines the extent to which results from qualitative and quantitative analytical approaches converge.

We presented three main findings. First, according to our respondents, teachers had indeed altered their classroom instruction. This finding confirms our selection of a study context that allows us to investigate determinants of such changes. Secondly, in analyses of what provoked these changes, we identified three primary drivers of change. In summary, these main drivers relate to teachers' acquisition of new teaching methods and pedagogical skills, to on-site team-based problem solving among teachers, and—to a lesser extent—verbal encouragement. Third, through a mixed-methods approach we presented results from qualitative and quantitative analyses of how these primary drivers relate to teachers'

involvement in professional development activities. Our findings point to different roles of on-site vs. off-site training, whereby off-site training may play a greater role for the acquisition of new skills, yet on-site training appears to be more strongly associated with joint problem solving. Moreover, the mixed-methods results suggest mentoring activities relate most directly to verbal encouragement. Lastly, the quantitative and qualitative findings converge by suggesting a limited role for technology-based solutions in provoking the three main drivers of change.

Results from the current study have several implications for education practitioners and researchers. For practitioners, the findings may suggest continuous teacher development operates best when approached through a multi-pronged approach that combines the individual strengths of each type of professional development activity. In particular, initial off-site teacher training may be best positioned to transfer new skills, yet require subsequent on-site training and mentoring to invoke team-based problem-solving and verbal encouragement. In turn, for researchers, our novel mixed-methods approach highlights how qualitative and quantitative analyses can complement each other. Specifically, we showed how unsupervised machine learning can supplement the discovery of themes in transcripts of qualitative interviews, however to a limited extent. We find that hand coding remains superior to machine learning techniques, but can be validated by machine learning. We also demonstrated how a combination of qualitative and quantitative methods can uncover associations among themes. Results from the two approaches tracked each other very closely, but also identified one interesting discrepancy that led us to additional, qualitative analyses.

We conclude by pointing to several limitations of our study, and how they may be addressed by future research. We begin by noting that our associations between changes in instructional behavior, their drivers, and their relationship to professional development activities are correlational and exploratory. They should not be mistaken as causal relationships. In our ongoing follow-up work, we are preparing a large cluster-randomized controlled trial to measure the causal effect of a continuous professional development model

that aims to tap into the drivers we identified in the present study. Secondly, our analyses rest on self-reported exposure to professional development activities. Future research may compare independent observations of such activities with self-reports. A third limitation relates to the question of whether practitioners should place focus on those elements of professional development that already appear to provoke changes in instruction, as apposed to other elements that are "not yet working." Future research may unearth whether greater benefits result from efforts that seek to leverage what appears to be a working system, as evidenced by our study's snapshot of existing drivers. In contrast, efforts may be more beneficial if they "fix" those components that are yet to provoke positive changes in teaching behavior (e.g., use of data, or educational technology). Lastly, we highlight that our study's external validity may be limited. Challenges for teacher professional development in the Zambian context may indeed reflect similar issues other less-developed countries face as well, including high travel costs for example, and a limited role for roving mentors. In contrast, more densely populated countries with low travel costs may find this study's findings less relevant. To scrutinize the study's external validity, it would be insightful if future research replicated the present study in other contexts.

References

Albornoz, F., Anauati, M. V., Furman, M., Luzuriaga, M., Podestá, M. E., & Taylor, I. (2020). Training to Teach Science: Experimental Evidence from Argentina. *The World Bank Economic Review*, *34*(2), 393–417. https://doi.org/10.1093/wber/lhy010

Alcott, B., Rose, P., Sabates, R., Luisa Alonso, M., & Cherfils, M. (2018). *Experience and lessons of learning intervention programmes across the PAL Network members* (tech. rep.). Zenodo. https://doi.org/10.5281/zenodo.1994919

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, *130*(3), 1117–1165. https://doi.org/10.1093/qje/qjv015

Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, F. H., & Sabarwal, S. (2020). *How to improve education outcomes most efficiently? A Comparison of 150 interventions using the new Learning-Adjusted Years of Schooling metric* (Working Paper No. 9450). The World Bank. Washington, D.C. Retrieved August 11, 2021, from https://openknowledge.worldbank.org/bitstream/handle/10986/34658/How-to-Improve-Education-Outcomes-Most-Efficiently-A-Comparison-of-150-Interventions-Using-the-New-Learning-Adjusted-Years-of-Schooling-Metric.pdf?sequence=1&isAllowed=y

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics*, *131*(3), 1415–1453.

Azam, M., & Kingdon, G. G. (2015). Assessing Teacher Quality in India. *Journal of Development Economics*, *117*, 74–83. https://doi.org/10.1016/j.jdeveco.2015.07.001

Azevedo, J. P., Goldemberg, D., Montoya, S., Nayar, R., Rogers, H., Saavedra, J., & Stacy, B. W. (2021). *Will Every Child Be Able to Read by 2030? Defining Learning Poverty and Mapping the Dimensions of the Challenge* (Working Paper No. 9588). The World Bank. Washington, D.C. Retrieved August 9, 2021, from https://openknowledge.worldbank.org/bitstream/handle/10986/35300/Will-Every-

Child-Be-Able-to-Read-by-2030-Defining-Learning-Poverty-and-Mapping-the-Dimensions-of-the-Challenge.pdf?sequence=1&isAllowed=y

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. *Journal of Economic Perspectives*, *31*(4), 73–102. https://doi.org/10.1257/jep.31.4.73

Bau, N., & Das, J. (2017). *The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers* (Working Paper No. 8050). The World Bank. Washington, D.C. Retrieved April 8, 2018, from https://openknowledge.worldbank.org/handle/10986/26502

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bernard, H. R. (2013). *Social research methods: Qualitative and quantitative approaches* (2nd ed). SAGE Publications.

Bietenbeck, J., Piopiunik, M., & Wiederhold, S. (2018). Africa's Skill Tragedy Does Teachers' Lack of Knowledge Lead to Low Student Performance? *Journal of Human Resources*, *53*(3), 553–578. https://doi.org/10.3368/jhr.53.3.0616-8002R1

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., & Wane, W. (2017). Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa. *Journal of Economic Perspectives*, *31*(4), 185–204. https://doi.org/10.1257/jep.31.4.185

Bold, T., Filmer, D., Molina, E., & Svensson, J. (2018). *The Lost Human Capital: Teacher Knowledge and Student Learning in Africa* (tech. rep. No. 12956). C.E.P.R. Discussion Papers. Retrieved June 9, 2018, from https://ideas.repec.org/p/cpr/ceprdp/12956.html

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bruns, B., Costa, L., & Cunha, N. (2018). Through the looking glass: Can classroom observation and coaching improve teacher performance in Brazil? *Economics of Education Review*, *64*, 214–250. https://doi.org/10.1016/j.econedurev.2018.03.003

Buhl-Wiggers, J., Kerwin, J., Smith, J., & Thornton, R. (2018). *Teacher Effectiveness in Africa: Longitudinal and Causal Estimates* (Working Paper S-89238-UGA-1). International Growth Centre. London, UK. Retrieved August 7, 2021, from https://www.theigc.org/wp-content/uploads/2019/01/Buhl-Wiggers-et-al-2018-Working-paper.pdf

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Castro, J. F., Glewwe, P., & Montero, R. (2019). *Work With What You've Got: Improving Teachers' Pedagogical Skills at Scale in Rural Peru* (Working Paper No. 158). Peruvian Economic Association. Puebla, Mexico. Retrieved December 1, 2019, from http://perueconomics.org/wp-content/uploads/2019/12/WP-158.pdf

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Cilliers, J., Fleisch, B., Kotzé, J., Mohohlwanex, N., Taylor, S., & Thulare, T. (2020). *Can Virtual Replace In-person Coaching? Experimental Evidence on Teacher Professional Development and Student Learning in South Africa* (Working Paper No. 20/050). RISE. Oxford. https://doi.org/10.1257/rct.5148-1.0

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2020). How to Improve Teaching Practice? Experimental Comparison of Centralized Training and In-classroom Coaching.

Journal of Human Resources, *55*(3), 926–962. https://doi.org/10.3368/jhr.55.3.0618-9538R1

Conn, K. M. (2017). Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations. *Review of Educational Research*, *87*(5), 863–898.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third Edition). SAGE.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. Retrieved August 8, 2021, from https://files.eric.ed.gov/fulltext/ED606743.pdf

DellaVigna, S., & Linos, E. (2020). *RCTs to Scale: Comprehensive Evidence from Two Nudge Units* (Working Paper No. 27594). National Bureau of Economic Research. https://doi.org/10.3386/w27594

de Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2018). Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia. *The Quarterly Journal of Economics*, *133*(2), 993–1039. https://doi.org/10.1093/qje/qjx040

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Duflo, A., Kiessel, J., & Lucas, A. (2020). *Experimental Evidence on Alternative Policies to Increase Learning at Scale* (Working Paper No. 27298). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w27298

Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Hsieh, P.-T. J., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., & Elbourne, D. (2021). How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention

from a cluster-randomized trial in the Gambia. *Journal of Development Economics*, *148*, 102539. https://doi.org/10.1016/j.jdeveco.2020.102539

Evans, D. K., & Popova, A. (2017). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*, 242–270.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, *19*(6), 643–658. https://doi.org/10.1016/S0742-051X(03)00059-3

Guskey, T. R. (2002). Professional Development and Teacher Change. *Teachers and Teaching*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Harvey, S. (1999). The impact of coaching in South African primary science InSET. *International Journal of Educational Development*, *19*(3), 191–205. https://doi.org/10.1016/S0738-0593(99)00012-7

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hill, H. C., Lynch, K., Gonzalez, K. E., & Pollard, C. (2020). Professional development that improves STEM outcomes. *Phi Delta Kappan*, *101*(5), 50–56. https://doi.org/10.1177/0031721720903829

Jackson, C. K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy*, *126*(5), 2072–2107. https://doi.org/10.1086/699018

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics*, *6*(1), 801–825. https://doi.org/10.1146/annurev-economics-080213-040845

Kennedy, M. M. (2019). How We Learn About Teacher Learning. *Review of Research in Education*, *43*(1), 138–162. https://doi.org/10.3102/0091732X19838970

Kerwin, J. T., & Thornton, R. L. (2021). Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures. *The Review of Economics and Statistics*, *103*(2), 251–264. https://doi.org/10.1162/rest_a_00911

King, F. (2014). Evaluating the impact of teacher professional development: An evidence-based framework. *Professional Development in Education*, *40*(1), 89–111. https://doi.org/10.1080/19415257.2013.823099

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Leme, M. C., Louzano, P., Ponczek, V., & Souza, A. P. (2012). The impact of structured teaching methods on the quality of education in Brazil. *Economics of Education Review*, *31*(5), 850–860. https://doi.org/10.1016/j.econedurev.2012.05.008

Majerowicz, S., & Montero, R. (2018). *Can Teaching be Taught? Experimental Evidence from a Teacher Coaching Program in Peru* (Job Market Paper). Retrieved April 24, 2019, from https://scholar.harvard.edu/files/smajerowicz/files/coaching.pdf

McChesney, K., & Aldridge, J. M. (2019). What gets in the way? A new conceptual model for the trajectory from teacher professional development to impact. *Professional Development in Education*, *0*(0), 1–19. https://doi.org/10.1080/19415257.2019.1667412

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook, 2nd ed.* Sage Publications, Inc.

Mulkeen, A. (2009). *Teachers in Anglophone Africa: Issues in teacher supply, training, and management*. The World Bank. Retrieved August 8, 2021, from https://doi.org/10.1596/978-0-8213-8053-6

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*(1), 160940691773384. https://doi.org/10.1177/1609406917733847

Opfer, V. D., & Pedder, D. (2011). Conceptualizing Teacher Professional Learning. *Review of Educational Research*, *81*(3), 376–407. https://doi.org/10.3102/0034654311413609

Piper, B., Destefano, J., Kinyanjui, E. M., & Ong'ele, S. (2018). Scaling up successfully: Lessons from Kenya's Tusome national literacy program. *Journal of Educational Change*, *19*(3), 293–321. https://doi.org/10.1007/s10833-018-9325-4

Piper, B., Simmons Zuilkowski, S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, *106*, 324–336. https://doi.org/10.1016/j.worlddev.2018.01.018

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2021). Teacher Professional Development around the World: The Gap between Evidence and Practice. *The World Bank Research Observer*, lkab006. https://doi.org/10.1093/wbro/lkab006

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramachandran, V., Béteille, T., Linden, T., Dey, S., Goyal, S., & Goel Chatterjee, P. (2018). *Getting the Right Teachers into the Right Schools: Managing India's Teacher Workforce*. The World Bank. https://doi.org/10.1596/978-1-4648-0987-3

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417–458.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M. G., & Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review* (tech. rep.). International Initiative for Impact Evaluation. London.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, *37*(9), 963–980. https://doi.org/10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0

Tan, J.-P., Lane, J., & Lassibille, G. (1999). Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments. *The World Bank Economic Review*, *13*(3), 493–508. https://doi.org/10.1093/wber/13.3.493

The World Bank. (2018a). *Learning to Realize Education's Promise. World Development Report 2018*. The World Bank. https://doi.org/10.1596/978-1-4648-1096-1

The World Bank. (2018b). *Selected Drivers of Education Quality: Pre- and In-Service Teacher Training* (tech. rep.). The World Bank. Washington, D.C. Retrieved August 7, 2021, from https://openknowledge.worldbank.org/bitstream/handle/10986/33203/Selected-Drivers-of-Education-Quality-Pre-and-In-Service-Teacher-Training-IEG-Meso-Evaluation.pdf?sequence=1&isAllowed=y

The World Bank. (2020). *Cost-Effective Approaches to Improve Global Learning: What Does Recent Evidence Tell Us Are "Smart Buys" for Improving Learning in Low and Middle Income Countries?* (Tech. rep.). The World Bank. Washington, D.C. Retrieved June 8, 2021, from https://documents1.worldbank.org/curated/en/719211603835247448/pdf/Cost-Effective-Approaches-to-Improve-Global-Learning-What-Does-Recent-Evidence-Tell-Us-Are-Smart-Buys-for-Improving-Learning-in-Low-and-Middle-Income-Countries.pdf

Timperly, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration (BES).* University of Auckland.

Vivalt, E. (2020). How Much Can We Generalize From Impact Evaluations? *Journal of the European Economic Association*, *18*(6), 3045–3089. https://doi.org/10.1093/jeea/jvaa019

Table 1

Sample Characteristics

	n	*%*
Panel A: Schools		
Number of schools	12	
Number of districts	10	
Number of zones	12	
Government school (vs. community school)	8	66.67
Rural school (vs. urban school)	7	58.33
Panel B: Participants		
Number of sampled individuals	83	
Non-response	5	6.02
Average interview time (hours)	0.95	
Female	35	44.87
Years of experience in role ($n = 72$)	4.7	
Works at school level (vs. above school level)	55	70.51

Notes. This table displays sample characteristics for the study's schools (Panel A) and participants (Panel B). Schools' classifications (community vs. government schools), and schools' geographic location (rural vs. urban) as per Zambia's 2018 educational management information system (EMIS). For years of experience in interviewed role, data is missing for six respondents. "Works at school level" captures whether a respondent holds a school-based (e.g., a teacher) or non-school-based position (e.g., a district official).

Table 2

Prevalence of Changes in Instruction

	n	% of excerpts				
		Overall	**By program**			
			Catch Up	SPRINT	Let's Read	Other
Technical changes	64	76.69	49.58	13.56	5.93	7.63
Increase in differentiated instruction based on learners' needs	19	18.60	11.40	5.10	1.70	0.40
Application of Catch Up to non-Catch Up classes	16	14.41	14.41	0.00	0.00	0.00
Increased use of materials	22	15.25	10.59	2.12	1.27	1.27
Increased interaction with learners	12	7.63	5.51	1.27	0.85	0.00
Increased lesson preparation	14	7.63	4.24	1.69	0.85	0.85
Increased student participation in activities	12	5.08	2.97	1.27	0.42	0.42
Teacher attitude / confidence	22	16.53	9.32	3.39	1.69	2.12
Attitude change	13	8.47	5.08	1.69	0.85	0.85
Change in intrinsic motivation / commitment to learners	15	8.05	4.24	1.69	1.27	0.85
No change	14	9.75	2.12	6.36	1.27	0.00
Total			63.98	31.36	8.90	11.86

Notes. This table presents the reported prevalence of changes in teachers' instructional behaviors. "Excerpts" refers to 236 excerpts in which respondents discussed changes in instructional behaviors (including absence thereof). The first column shows the number of respondents that mentioned a given change. The remaining columns reflect the percentage of excerpts that pertain to various types of change (overall, and by programs). "Technical changes" include 14.8% of excerpts that were related to pedagogy with no further specification. Changes due to COVID-19 (13%) and codes with incidence rates lower than 5 percent are omitted. Totals may exceed 100% due to excerpts that refer to more than one program or denote more than one change.

Table 3

Drivers of Change in Instruction

	n	**% of excerpts**		
		Overall	**By cadre**	
			Teachers	Other respondents
Primary drivers				
Sharing and discussing challenges	51	11.15	6.18	4.97
Acquisition of new skills and teaching methods	34	10.30	7.39	2.91
Verbal encouragement and/or discussions	36	7.03	2.30	4.73
Secondary drivers				
Learner outcomes	37	6.67	2.42	4.24
Materials (such as flipcharts and markers)	26	5.21	2.42	2.79
Frequent monitoring	29	4.97	1.09	3.88

Notes. This table presents the reported prevalence of drivers that provoked teachers to change their instructional behaviors. "Excerpts" refers to 825 excerpts in which respondents discussed such drivers of change in instructional behaviors. The first column shows the number of respondents that mentioned each driver. The remaining columns reflect the percentage of excerpts that pertain to various drivers (overall, and by respondent cadre). Codes with incidence rates lower than 5 percent are omitted. Totals may exceed 100% due to excerpts that refer to more than one cadre or denote more than one driver of change.

Table 4

Prevalence of Professional Development Opportunities, Their Association With Primary Drivers of Change

	Prevalence		Association with primary drivers of change								
			Sharing challenges			Verbal encouragement			Learning new methods		
			Qual.	Quant.		Qual.	Quant.		Qual.	Quant.	
	% Respondents	% Excerpts		Within 10	Within 20		Within 10	Within 20		Within 10	Within 20
Training											
Off-site training	38.46	10.97		-11.65	-14.98		10.80	15.95		7.81	13.53
On-site training	67.95	20.25		26.75	33.58		1.98	6.18		-2.79	3.38
Monitoring											
Monitoring through lesson observations	71.79	31.43		-9.29	-7.04		3.67	6.96		-2.17	3.38
Monitoring through file-checking	24.36	4.43		-7.32	0.84		-3.29	-3.86		-10.10	-8.76
Monitoring by walking-around	6.41	1.27		0.54	-4.38		-3.98	-0.07		-4.54	1.71
Mentoring											
Capacity building during 1:1 mentoring	55.13	14.98		5.49	15.26		9.80	9.05		5.82	11.59
Encouragement during 1:1 mentoring	24.36	6.96		-2.54	-0.16		5.11	17.58		-9.20	-11.60
Lesson planning support during 1:1 mentoring	16.67	3.59		-5.13	-10.88		8.57	12.85		-5.55	-6.19
Technology											
Mentoring over technology	29.49	5.27		-6.61	-10.74		n.a.	-11.77		-10.45	-11.21
Data use	28.21	5.49		-2.41	5.18		4.39	6.60		5.10	9.43

Notes. This table shows the prevalence of professional development opportunities and their association with primary drivers of change in teaching behaviors. Professional development opportunities are shown in rows. "Prevalence" reports on the percentage of respondents and excerpts that relate a given professional development opportunity to changes in instruction. Under "Prevalence", "Excerpts" refers to 474 excerpts in which respondents discussed teacher professional development opportunities, given that a driver of change was mentioned in the interview. The quantitative (Quant.) component, reports the percentage point increase (decrease) in the probability of a driver being mentioned in an excerpt, if a professional development opportunity is mentioned within the neighborhood of ten or twenty excerpts, respectively. These "associations" are color coded, whereby green highlights positive, yellow highlights weak, and red highlights negative associations.Results from a quantitative analysis are shown with a continuous color scale. The qualitative (Qual.) component is shown in three discrete colors, only. Green indicates strong qualitative evidence of associations between each respective driver and its professional development opportunity. Orange indicates weak evidence and red indicates no evidence.

Appendix

Additional Tables and Figures

Table A1

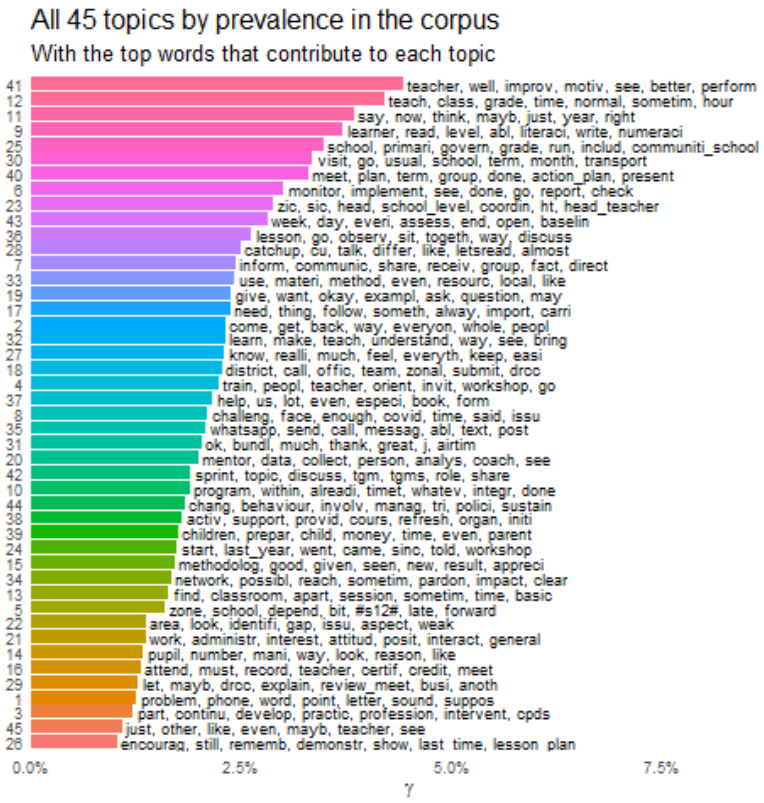
Comparison of Qualitatively Identified Themes With Quantitatively Identified Themes

	Number of codes in qualitative coding framework	**Number of topics in text analysis**
Changes in teaching practices	27	2
Drivers of change	52	4
Professional development inputs	243	6
Training	73	1
Monitoring	78	1
Mentoring	55	3
Technology	16	1
Data use	21	0
Other categories	86	14
Nonsensical categories	-	19
Total	**408**	**45**

Notes. This table presents the number of codes as per the qualitative coding framework and as per the quantitative text analysis, respectively. Each row represents a category. We focus on categories of codes that were included in the current study. "Other" categories represent unrelated codes, such as background information on the setting of a school. "Nonsensical" categories represent topics as per the quantitative text analysis that, after review, did not prove meaningful (e.g., topic 11 in Appendix Figure A1: say, now, think, mayb, just, year, right).

Figure A1

Identification of Themes Through Unsupervised Machine Learning



Note. This figure presents the 45 themes identified through topic modelling. Themes are ordered by prevalence and presented along with their seven most distinctive terms. The corpus of text relies on all sentences spoken by respondents (5,879 answers); it excludes any text spoken by interviewers. Bi- and trigrams were added via Rapid Automatic Keyword Extraction (RAKE). The analysis excludes common stopwords (such as “me”, “my”, “myself”, “we”, or “our”), sets all text to lower case, and “stems” words (e.g., by removing suffixes such as “ed”, “ing”, or “ly”).