Selected Drivers of Education Quality

Pre- and In-Service Teacher Training

IEG MESO EVALUATION

IEG | WORLD BANK GROUP
INDEPENDENT EVALUATION GROUP | World Bank • IFC • MIGA

WHAT WORKS

Selected Drivers of Education Quality

Pre- and In-Service Teacher Training

IEG Meso Evaluation

November 15, 2019

Contents

Contents

Boxes

Figures

Table

Appendixes

Abbreviations

FY   fiscal year
IEG   Independent Evaluation Group
Lao PDR   Lao People's Democratic Republic
PAD   project appraisal document
TEDS-M   Teacher Education and Development Study in Mathematics
TTL   task team leader
VNEN   Vietnam Escuela Nueva Project

All dollar amounts are U.S. dollars unless otherwise indicated.

Acknowledgments

Overview

Highlights

- World Bank engagement in training teachers before entry into the profession (preservice training) has been limited and has prioritized coursework, with less emphasis on other drivers of quality, such as screening, practicum, and quality assurance. The World Bank has instead relied heavily on engagement with continued training during employment (in-service training) to address shortcomings in preservice training through support to programs for both underqualified and qualified teachers.

- Most countries where the World Bank provides support require discipline-specific in-service training that is adapted to teachers' needs and capacity, models adult learning style, and includes follow-up support. Some of these features are evident in the operations examined, though often not in combination.

- Well-designed and well-implemented training programs alone cannot improve teachers' pedagogical practices, particularly without strong instructional leadership. This area received support in just 40 percent of the operations, which is low considering the challenges country clients face. Education systems need to create an enabling environment to sustain teacher professional development within a broader career framework supported by instructional leaders and incentives.

- Scaling up of training programs needs to increase both the breadth of coverage and the depth and sustainability of the training for it to achieve long-term changes in teaching practices. World Bank–supported scaling efforts have achieved success largely by increasing the number of teachers trained. Some desirable conditions to ensure sustainability of the training programs—longer-term strategic focus with financing sufficient to support sustainability, ongoing communication with key stakeholders, and political support—are missing in some cases.

- The World Bank could give more attention to evaluating the training programs it supports to improve program effectiveness, resource use, and learning, and to provide data to support the achievement of progress and outcomes. Limited monitoring and evaluation can also undermine sustainability. The data collected on scaling-up efforts typically served an accountability function rather than a feedback process to refine efforts and show the value of in-service training to build sustained system change and stakeholder support.

Evaluation Motivation and Rationale

The World Bank has increased its emphasis on teachers and their training in its operations and with initiatives such as the Human Capital Project and the Global Platform for Successful Teachers. Better-trained teachers perform better in the classroom, and better-performing teachers improve student learning. Therefore, strengthening the preparation (preservice) and training of teachers throughout their careers is key to addressing low student learning attainment in many low- and middle-income countries. This evaluation recognizes the renewed emphasis on quality teaching and supports that effort with evidence on what works in training programs.

The evaluation examines how the World Bank supports preservice and in-service training and how these programs can be better designed, implemented, and scaled up. The World Bank can use the learning provided here to aid decision-making related to scaling up its Human Capital Project by reviewing how its interventions might better support teacher training. Ultimately, the evaluation seeks to inform both the World Bank's support for teacher training and that of its clients, partners, and donors active in education development.

Approach and Methods

The evaluation approach used a mix of methods to identify quality drivers and key features for preservice and in-service training and conditions that aid in scaling up in-service training. The two training systems are conceptually different but linked, with preservice training following a logical progression through a series of steps, each with a singular purpose and specific quality characteristics. In-service training, by nature, is not sequential, can serve a variety of purposes, and encompasses a cluster of quality drivers, all of which need to be addressed to some degree to ensure success.

The key features and quality drivers were derived from the literature and secondary data analysis of evidence related to preservice training and applied to the lending portfolio. The identified drivers are screening (mechanisms to select teaching candidates), coursework, practicum (practice teaching), quality assurance (of training programs), and in-service training (itself a cluster of features). The evaluation team applied these drivers, along with conditions that support scaling up, to the project appraisal documents for an identified lending portfolio of 110 World Bank operations and examined in field-based case studies in Ghana, Uruguay, and Vietnam. Findings were triangulated with interviews with task team leaders (TTLs), which provided valuable information about the current direction of the World Bank and supplemented information contained in the examined documents.

Scope of Training Support

Although the amount of World Bank education lending with training components is sizable, the amount

allocated to training is tracked in just half of the 110 projects and, where it is tracked, typically comprises only 10 to 15 percent of project financing. This figure likely underestimates the amount devoted to training. World Bank lending for education between fiscal year (FY)13 and FY18 was $18.4 billion for 207 operations; the total financing of the 110 projects with a training component approved (and tracked) since FY13 accounts for $12.1 billion. The World Bank could enhance the specificity of government data to understand how much governments devote to training as a share of total educational expenditures.

Projects approved since FY13 emphasize in-service training over preservice training. Of the 110 projects examined, 68 exclusively support in-service training, 2 support only preservice training, and the rest support both. Recent operations usually support both types of training. Key stakeholders, including TTLs, indicated in interviews that both systems should be addressed simultaneously because they saw a need to build greater alignment between them. Preservice training was addressed only when the government supported the need to reform preservice institutions, so the World Bank tends to focus more on in-service training, which is often easier to reform. For this reason, the World Bank has elected to focus on select preservice quality drivers and has rarely addressed all drivers in combination.

Preservice Training Findings

Filtering and screening. The World Bank has tended to focus on motivating potential teachers, with less attention to recruiting the strongest candidates. Nearly one-third of the operations examined tried to motivate trainees through support for scholarships and stipends. However, TTLs say that attracting better candidates is often inhibited by the unattractiveness of teaching as a profession and the low capacity of students who enter tertiary education. Operations that address screening of those exiting preservice training generally focus on strengthening existing examination functions. When operations adopt alternative approaches, they more often seek to address scarcity than to create a new mechanism to bring in candidates with stronger content knowledge. In part, this is because governments do not want to deter candidates when education systems are growing.

Coursework. The World Bank has focused on enhancing the curriculum, building the capacity of teacher trainers, and improving infrastructure related to teacher preparation. The literature and TTLs both emphasize the importance of aligning the preservice curriculum and methods preparation with the curriculum to be taught in schools. This contrasts with many preservice training classrooms, which rely on conveying theoretical information and ineffective teaching methods such as rote memorization and

copying. This helps explain why half of the operations examined provide some form of capacity development for teacher trainers.

Practicum. The World Bank has given some attention to teacher practicum, but there is room to do more. TTLs noted that practicums are critical, suggesting a growing awareness of the need to include support mechanisms related to practicums beyond what was noted in one-third of appraisal documents. In some instances, the support is extensive, linking up with multiple aspects of preservice training and incorporating key principles. A common description of practicums in many low-income countries is "sink or swim," an approach in which trainees are given too much autonomy too soon and without a qualified mentor.

Quality assurance. The World Bank may be under-engaged in areas important to quality assurance, such as establishing education standards, teacher training curriculums, teacher educator requirements, practicum requirements, and other system-level matters. Data show that countries with weak accreditation systems have no effective control over training institutions or rely on voluntary participation mechanisms. The accreditation process is not just about holding preservice training institutes accountable to a standard of quality but also about supporting institutions to develop to the standard—both areas where the World Bank might expand its support. However, the existing quality

assurance systems (or lack of such systems) are constraints.

In-Service Training Findings

The key features of effective in-service training—adequate duration, discipline specificity, active and applied learning based on teachers' needs and capacity, and follow-up support—are evident in nearly half of the World Bank operations examined, but they are not present in combination. In the literature, the presence of all the key features is associated with improvements in student learning. The main findings are as follows:

- All in-service training supported by the World Bank meets at least the minimum requirement for duration lasting between 5 and 20 days (or more).

- The design and implementation of training is weakest in the provision of follow-up support. TTLs reported encouragement that includes coaching, but there have been issues with the level of quality and variation in quality of coaching, suggesting a need to address these in the future.

- Although effective training programs focus on the content to be taught—such as mathematics, science, or literacy because discipline-specific training programs produce greater learning gains—pedagogy was the focus in about half of the World Bank

operations, even those supporting secondary education, where subject specificity is critical.

- Effective training also supports the way in which adults learn through application, modeling, and demonstration during training, as was found in at least one World Bank–supported operation.

Achieving the aims of training programs depends partly on the coherence of those aims with the enabling conditions within the education system. Instructional leadership, shown to be critical to training programs, was simultaneously supported in nearly 40 percent of the operations. Leadership skills were a shortcoming in principals in some low- and middle-income contexts. Training programs are also not implemented consistently within a broader framework for teacher development. TTLs believe an important way to improve training is to anchor it within a broader career framework. This would require greater clarity of the outcomes of training than World Bank operations currently put into planning efforts.

Scaling up that increases the breadth of training coverage without ensuring depth and sustainability of the training is less likely to achieve long-term changes in teaching practices because some of the key conditions to ensure sustainability are missing in the World Bank's current approach. Each of the six case studies included well-sequenced plans for scaling up training programs with good examples

of support for administrative capacity, and logistics, procurement, and human resources management. These cases tended to have more success with gains in the reach of in-service training than with ensuring that the programs became embedded in the education system. Project evaluations in all six cases document success in meeting targets for teachers trained, but the government has sustained only one of the programs. Some of the conditions that need to be addressed more systematically to better support scaling up include planning for longer-term funding support, consultation and communication with beneficiaries and other key stakeholders, and ongoing political support, especially to ensure lasting improvements.

Weak monitoring and evaluation undermine the potential for sustainable approaches to in-service training and learning. Systematic evidence (including robust monitoring in addition to evaluation that is used to adapt to implementation challenges) is critical to sustainability, but the quality of the evidence to support sustained scaling was weak in the countries examined. Less than half of the operations examined evaluate in-service training. Monitoring is infrequently used to ensure the fidelity and implementation of the training and follow-up support. Without such data, it is not possible to identify lack of consistency or bring greater oversight and generate learning.

Lessons

The World Bank's limited focus on preservice training institutions may not ensure effective teaching. Instead, more active policy dialogue may be needed to convince clients to reform. A more contextualized assessment of preservice institutions that highlights individual institutional strengths and weaknesses, for example, may help overcome political economy constraints. Including such assessments in policy dialogue would provide government clients with the evidence they require to understand the issues affecting the quality of preservice training and move toward reform. Dialogue may also facilitate development of a long-term plan to sequence improvements to the quality drivers. The shortcomings of graduating candidates have repercussions for in-service training programs, which alone cannot improve these candidates. Thus, the rationale for attending to both types of training systems is clear.

The effectiveness of in-service training programs depends on consistent attention to all key features. The degree to which key features are integrated with the education system matters. Ensuring integration may require the World Bank to consider comprehensive in-service training reforms and embed them in the education system. Effective in-service training programs alone are not enough to give teachers a broad repertoire of skills or make them more reflective practitioners. For this reason, sustained follow-up is critical and requires mechanisms that provide opportunities for peer learning and coaching, as well as participatory approaches that elicit teachers' views about their training and needs. These mechanisms require effective instructional leaders who can provide follow-up support that differentiates teachers' capacity and helps teachers address varying learning levels among students. Additionally, training programs need to be anchored in teaching standards, career ladder progression, and screening throughout the career. Embedding key in-service features in the education system and in the design and implementation of operations (including monitoring and evaluation) is also needed.

Sustainable scaling up of in-service training requires attention to key conditions for the planning, implementation, and monitoring of the scaling. In-service scaling was well planned and well implemented in the cases examined. In the training programs visited, scaling-up plans covered logistics, costs, and modalities; targets were set and met for numbers trained, and materials were distributed as relevant. Yet sustainability remains an issue. In this regard, World Bank operations may benefit from greater focus on quality assurance for in-service training, arrangements to evaluate training outcomes, and planning to embed system-related aspects of in-service training into the scaling. Some of the gaps in planning and design were associated with short

implementation periods for operations.
Thus, TTLs may need to plan for scale-up
from the start and address constraints
such as political support, long-term
financing, and monitoring and evaluation.
Efforts to assess quality and outcome can
help build a case for more sustained in-
service provision.

1. Introduction

Highlights

- Student learning depends on quality teaching, which requires effective teacher preservice and in-service training.

- The World Bank primarily supports in-service training, mostly in low- and lower-middle-income countries, which is consistent with need.

- The World Bank and clients do not consistently collect data related to the cost of training programs, so it is unclear how much is spent.

- Five drivers are essential to effective teacher professional development—screening and filtering, coursework, practicum, quality assurance, and in-service training—but the school and education system context are at least equally important for quality results.

Teacher effort and capacity are critical to student learning and educational outcomes. The *World Development Report 2018: Learning to Realize the Promise of Education*, highlights the centrality of learning for education systems. Furthermore, the World Bank's Human Capital Project emphasizes the importance of education for human development. Hence, an Independent Evaluation Group (IEG) evaluation focused on selected drivers of education quality and learning is timely and relevant. Student learning depends on skilled, adequately prepared teachers. This evaluation focuses on understanding how the World Bank has supported two types of professional development to improve teacher capacity—preservice and in-service training (box 1.1)—and how these drivers of education quality can be better designed, implemented, and scaled up to make World Bank operations in this area more effective.

Box 1.1. What Are Preservice and In-Service Training?

The term "professional development" in this evaluation includes all learning opportunities for teachers, from the initial stages of their careers to their entry into the profession (known as preservice training) and throughout their employment as teachers, known as in-service training.

The purpose of preservice and in-service training is to "enhance the quality of students' learning by improving the quality of teaching through constant review and assessment of teachers' instructional approaches, identifying effective teaching approaches, and capitalizing on them for the benefit of the learners." Effective training equips teachers with knowledge in subject content, pedagogical strategies, and classroom management.

Source: Luneta 2012.

Evaluation Context: Why Teacher Development Matters

Teachers and effective teaching matter to improve learning. The literature has shown that individual teachers can have a sizable and direct effect on student performance (Hanushek and Rivkin 2010). Having an effective teacher also makes a considerable difference in students' learning trajectory (Rockoff 2004; Chetty, Friedman, and Rockoff 2014).

Professional development is essential to effective teaching. Observations in many countries have shown that teachers are often inadequately prepared to teach well (Reimers and Chung 2018, among others). For example, in some systems, teachers' basic knowledge cannot be assumed, as the World Bank's Service Delivery Indicators found. In this context, the literature shows that professional development is important to improving teacher capacity, but training needs to be high quality (Darling-Hammond and Richardson 2009; Darling-Hammond, Hyler, and Gardner 2017; Garet and others 2001).

The World Bank is now emphasizing teachers and their training. In 2018, the World Bank launched the Global Platform for Successful Teachers, aiming to have all children taught by effective teachers whose education systems support them. This renewed emphasis is important because there is limited guidance on what constitutes good practices for preservice training is limited. Although some reviews have identified key characteristics of in-service training, much less is known about how to scale up quality in-service training. Providing evidence on how to situate teacher training better during design, implementation, and scaling within the education system is critical because teachers do not operate in a vacuum. As the *World Development Report 2018* argued, all actors in the education system need to be aligned. Better designed, implemented, and scaled-up training programs could potentially offer better value for money.

World Bank Assistance for Preservice and In-Service Training

The World Bank has approved 110 projects in 67 countries since fiscal year (FY)13 (of 207 projects approved by the Education Global Practice between FY13 and FY18) that support the professional development of teachers, but the relative share of resources devoted to training is unclear. The World Bank's global lending for education between FY13 and FY18 was $18.4 billion among the 207 operations, and the total financing of the 110 training projects approved since FY13 accounted for $12.1 billion. It is not clear how much money the World Bank devotes to training because less than half of the project appraisal documents (PADs; 51of 110) present detailed cost data. The amount allocated to in-service training ranged from $1.5 million to $58 million—typically accounting for 10 to 30 percent of total project costs (the median was 15 percent), but four cases

accounted for more than half of the total project funding. The figure likely underestimates total resources devoted to training because other costs may be part of other components. Preservice training costs were usually included with in-service training, and only three PADs differentiated them. In these cases, the amount allocated to preservice training was $2 million, which is consistent with task team leaders' (TTL) reports that the World Bank placed more emphasis on in-service training. In the future, it will be important for the World Bank to address the lack of available data in PADs and support better production of cost data to ensure that the resources governments devote to teacher training as a share of total public educational expenditures is transparent.

More emphasis has been placed on in-service training within the operations reviewed. Sixty-eight projects exclusively support in-service training, two support only preservice, and the rest support both. Recent operations include both. TTLs believed both systems should be tackled simultaneously because they saw a need to build greater alignment. However, operations generally did not address preservice training for many reasons, the most common of which were the amount of project financing for teacher professional development and the political economy surrounding preservice training institutions. TTLs found it easier to support in-service training and more difficult to address preservice. Other reasons that drove the selection were context-specific, such as government financial constraints in hiring new teachers, which decreased the rationale to focus on preservice. It was also easier to engage in countries where the preservice training system was small because support to individual institutions requires context-specific understanding of their strengths and weaknesses.

Teacher training projects,[1] like education projects in general, are implemented more often in low- and lower-middle-income countries,[2] where the need is greatest. To illustrate the relevance of the World Bank's support, Figure 1.1 shows that these programs are weighted toward countries in the lower quartiles of the Human Capital Index—nearly three-quarters support countries in quartiles one and two.[3] Teaching practices in low-income countries in Africa need substantial improvement, based on Service Delivery Indicators. In Mozambique, for instance, only 15 percent of teachers answered the pedagogical questions correctly.[4] Thus, it is important for the World Bank to focus on these countries.

Figure 1.1. Pre- and In-Service Countries by Human Capital Index Quartile
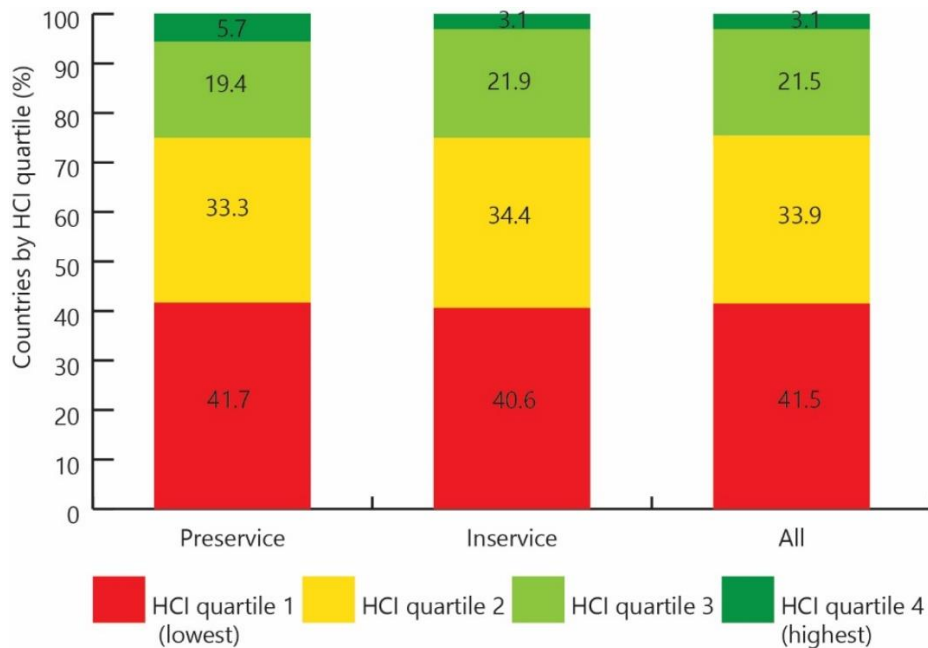
Source: Independent Evaluation Group coding and World Bank Human Capital Index.
Note: HCI = Human Capital Index; n = 64.

Quality Drivers for Teacher Professional Development: Preservice and In-Service Training that Are Well-Designed, Scaled, and Sustained

The conceptual framework for the evaluation covers professional development throughout a teaching career (Figure 1.2). It includes initial training (preservice) and training received while employed (in-service) and captures best practices for each. The key features of the framework were derived from the literature and secondary data analysis for evidence related to preservice training. The two systems are conceptually different but linked, with preservice training following a logical progression through a series of steps, each of which has a singular purpose and specific quality characteristics. In-service training, by nature, is not sequential, can serve a variety of purposes, and encompasses a cluster of quality drivers, all of which need to be addressed to some degree to ensure success. The two training systems are linked, which is represented by the framework's sequential arrangement. These links are necessary because the preservice curriculum should equip teachers with the pedagogical and content knowledge (math, language, and other subjects) they will need for the curriculum and classes they will teach. Moreover, if preservice training is inadequate, in-service training will be required to address the shortcomings in teachers' pedagogical and content knowledge, making it even more important to address all the quality drivers.

Many features of the education sector context—including financing, governance, curriculum, and incentives—affect teacher professional development. The attractiveness of some aspects of the education system, including well-established incentive mechanisms related to career advancement and pay, facilitate lifelong teacher learning and development. By contrast, the in-service training program is subjected to extra pressure when teachers are not well prepared initially and not effectively motivated during their careers. Education systems that do not initially screen candidates require effective filtering mechanisms at later stages through teacher performance management, typically after a probationary period or at various points throughout their careers. Although coherence with all aspects of the education system is needed to "transform the processes and policies to support teachers, their education, and their work," this report gives particular attention to instructional leadership and monitoring and evaluation (Villegas-Reimers 2003).

The first stage of preservice training, screening and filtering, is about getting quality candidates into the education field. Screening relies on high demand among high-performing students to enter the teaching profession, which relates closely to the attractiveness of teaching relative to other professions. Many factors contribute to the demand for teaching positions, including initial pay, career opportunities, incentive and support structures, classroom and school working conditions, and even cultural aspects related to how society views teachers. Additionally, it is not just about the supply of candidates. In low-income countries, it is common to have more teacher applicants than jobs, but this does not guarantee a high-quality cadre of teachers or teacher candidates (Bold and others 2017). Screening depends on transparent and meaningful requirements to enter and exit preservice institutions, such as examinations, grades, or graduation requirements.

Figure 1.2. Evaluation Conceptual Framework

OVERARCHING EDUCATION SECTOR FEATURES

REPORT FOCUS: school instructional leadership; **monitoring and evaluation of in-service training.**
Finance, governance, curriculum, student assessment, career opportunities/incentives and performance management.

PRE-SERVICE TRAINING   IN-SERVICE TRAINING

QUALITY DRIVER 1
Screening/filtering

Entry/Exit requirements
to attract quality
candidates

QUALITY DRIVER 2
Coursework

Content and
pedagogical knowledge
for actual curriculum
Capable teacher
educators

QUALITY DRIVER 3
Practicum

Supported experience
Coursework and
practicum aligned
Monitoring of
practicum

QUALITY DRIVER 5
In-Service Training

Adequate duration
Discipline-focused
Reflective of adult learning
Sustained follow-up support
Adapted to teachers
capacity

Sustained scale-up

**Intermediate
outcomes**

Improved pedagogical
knowledge and teaching
practices

Outcomes

Enhanced student
achievement,
attendance,
attainment, completion

QUALITY DRIVER 4
Quality Assurance Features
institutional accreditation; program oversight and support; alternative preparation mechanisms;
transparent screening/filtering mechanisms; teacher certification

Source: Independent Evaluation Group.

The second driver, coursework, aims at preparing teacher trainees for the classroom. Coursework needs to provide candidates with both content and pedagogical knowledge grounded in the curriculum of the schools where they will eventually teach. This requires teacher educators who can impart these skills and have the necessary learning materials. For this reason, the effectiveness of teacher educators matters. The length of training programs largely depends on the trainees' skill level, so there is little guidance on the optimum duration. Longer programs are not necessarily effective if the quality in the preservice training is low.

The third driver, practicum, is a critical component of a well-rounded professional development experience. Concerns about quality and effectiveness of coursework have led some researchers to focus instead on practicums (Béteille and Evans 2019; Béteille and others 2018; Lewin 2004). The practicum needs to be a supported experience for the trainee and be monitored. Effective practicums help teachers gradually assume more tasks and more responsibilities through developmentally appropriate clinical experiences (AMTE 2017, 38). This requires effective monitoring and mentoring, which begins by forming productive partnerships with schools. Experienced mentors who are familiar with the needs of beginning teachers are crucial for creating a trainee-centered experience. Formative assessments are required with feedback, accompanied by reflection and dialogue.

The fourth driver, quality assurance, regulates many aspects of the pre- and in-service training systems to ensure quality and transparency. This driver regulates teacher education program providers, which can help ensure adherence to training standards, removal of political influences, and effective control over the number of candidates entering the system. Quality assurance aspects also regulate screening mechanisms, such as entrance and exit examinations, to ensure that they are implemented well, provide clear signals about quality, and are free from manipulation. They also accredit training institutions and provide a support component to meet standards.[1] Certification and alternative preparation for teaching are other quality assurance mechanisms to ensure transparent filtering. Thus, this driver relates not just to preservice training but to the entire education system.

The fifth driver, in-service training, is about providing additional improvements in teachers' instructional practices and knowledge conducive to student learning. This training needs to be consistent with adult learning principles, that is, socially situated in teachers' context and adapted to teachers' capacity (Reimers and Chung 2018, among others). Some studies show that discipline-focused training is more effective than general content (Popova and others 2018; Darling-Hammond, Hyler, and Gardner 2017). An adequate duration of training with sustained follow-up support through coaching or

feedback to promote reflection is also more effective (Darling-Hammond, Hyler, and Gardner 2017; Popova and others 2018). The need to sustain this effort and to achieve sufficient scale in teachers' coverage to improve teachers' practices in certain contexts make the quality and sustained scaling up of these programs another important element. Conditions within the education system to support this aspect will also be discussed.

The framework includes outcomes to show the connection between professional development, teacher practices, and ultimately, student outcomes, such as achievement and attainment (Carrillo, van den Brink, and Groot 2016, for example). These outcomes are assumed (based on extensive literature showing associations between quality training and teaching practices and student learning) and are not the focus of this evaluation. Education sector features also affect outcomes directly. For example, career advancement and pay affect teachers' practices, independent of teachers' professional development, through incentives to perform.

Evaluation Objectives, Analytical Approach, and Methods

The objectives of this learning evaluation are to understand how the World Bank supports preservice and in-service training and how these interventions can be better designed, implemented, and scaled up. Better-trained teachers perform better in the classroom. Therefore, strengthening teacher training is key to addressing the crisis of low student learning attainments in many low- and middle-income countries. The goal of the evaluation is to provide the World Bank with information that will aid in decision-making related to scaling up the Human Capital Project by providing knowledge from the literature on how World Bank interventions can better support teacher training. Ultimately, this evaluation intends to inform both the World Bank's support for teacher training and that of its clients, partners, and donors active in education development.

Box 1.2. Evaluation Questions

- What are key features of effective preservice training programs, and to what extent do World Bank operations reflect these characteristics?

- What are key features of effective in-service training programs, and to what extent do World Bank operations reflect these characteristics?

- What factors determine the effective scaling up of effective teacher in-service training financed by the World Bank?

The evaluation used a mixed methods approach. For preservice training, the team used a structured literature review and analysis of secondary data (such as Tatto 2013) to identify key quality drivers, and then used these elements to examine the portfolio of

World Bank–supported training operations. For in-service training, existing evidence from systematic reviews provided effective features of in-service training programs that were applied to the portfolio review. For scaling up, a structured literature review supported the development of a theory of change and fed into a background paper that informed case studies. The three literature reviews provided the theoretical foundation for the evaluation, and the basis for developing theories of change and the coding template applied to the portfolio of operations for which PADs and other sources of evidence were reviewed.[2] The evaluation considered only lending operations and did not examine Advisory Services and Analytics or other forms of World Bank engagement.

The evaluation team examined additional design elements in field-based case studies in Ghana, Uruguay, and Vietnam. These cases were selected, based on a review of PADs and consultation with TTLs, to ensure the presence of key in-service features. Focus groups with TTLs supplemented the portfolio review and provided contextual details. Concerning scaling up in-service training, the case studies in Ghana, Uruguay, and Vietnam also provided relatively homogenous effective training programs to examine the scaling-up process in varying contexts. A theory-driven, cross-case analysis was used to detect conditions that facilitate scaling up.[3] Details of the evaluation's design and sources of evidence are in appendix A.

The evaluation faced some limitations. PADs and operational manuals provided relatively limited details about preservice and in-service training, and many had no description of important characteristics or associated costs.[4] Interviews with TTLs (individual and group) were conducted to supplement missing information and gain context-specific understanding. The additional verification comprised one-third of the portfolio and of cases with the largest data gaps, and they were consistent with patterns in the overall portfolio. Additionally, for preservice training, analysis of secondary data, combined with limited research studies, provided an indicative list of features that are often based on experiences in developed countries and therefore may not reflect the context in many low-income countries. Lessons can still be learned from more advanced countries with care to contextualize them. The team also examined data from developing countries, but these data sets contained fewer variables about preservice preparation.

2. Where Teachers Are Made: Preservice Training

Highlights

- The World Bank's experience with preservice training is limited.

- Support has focused predominantly on coursework with less attention to the other drivers—screening, practicum, and quality assurance—needed to improve preservice training quality.

- The World Bank has found it easier to use in-service training to address underprepared or unqualified teachers, which tended to constrain the ability of in-service training to have a long-term effect on quality.

For several reasons, the World Bank's experience with preservice training is limited to 40 operations out of 207 education operations approved between FY13 and FY18. The number of operations supporting both preservice and in-service training slightly increased in recent years. One reason for the limited focus is a complex political economy in this area, characterized by differing goals among multiple stakeholders. For example, one TTL reported, "The education sector is used to address youth unemployment without concern for the quality of the candidates in one country." Governments are reluctant to address preservice training, which limits the World Bank's ability to engage in this area. Limited subsector governance is also a concern. Thus, the World Bank has provided support when the government has recognized the need for reform. Aspects that TTLs reported were important for engagement are fully understanding the institutional landscape and the individual institutional strengths and weaknesses, and leveraging the World Bank's tertiary education support for preservice training.

This chapter considers the quality drivers presented in figure 1.2 in broad strokes. (See appendixes C and D for more detail about literature and secondary data analysis.) The drivers covered are screening of teacher candidates using entry and exit examinations, teacher preparation coursework, teaching practicum, and quality assurance. Figure 2.1 shows the number of operations, generally limited, that discuss each of these drivers. It also indicates how rare it is for an operation to support all of them (only one project does).

Screening of Teacher Candidates

The level of selectivity applied by preservice institutions is associated with student achievement. The literature on the screening of teacher candidates indicates that

Figure 2.1. Operations Covering Each Quality Driver
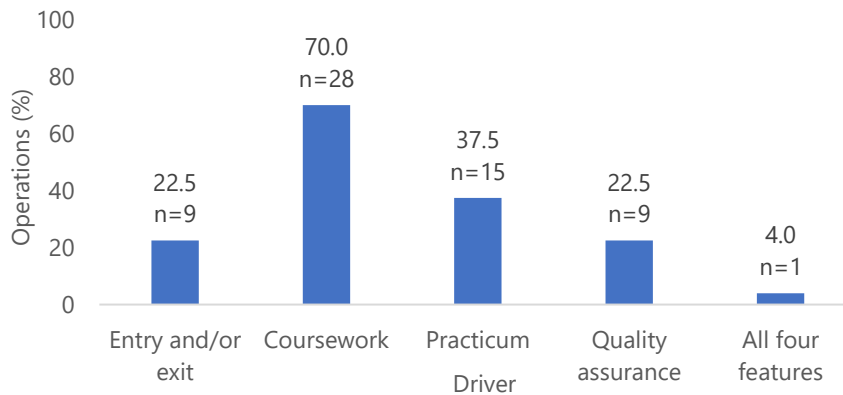
Note: The total number of preservice operations is 40.

selectivity throughout teacher education programs is associated with teacher effectiveness and therefore with student achievement. Wang and others (2003) finds that developed countries with the highest levels of student achievement are selective—and not just at the point of initial selection for teacher education programs but throughout the professional lifecycle of teachers.[1] For example, developed countries that use high-stakes screening mechanisms at multiple points along this sequence, such as Japan and Korea, tend to outperform countries with less stringent screening on international assessments.[2] Additionally, relatively high-performing countries like the Russian Federation and Switzerland, along with Taiwan, China, have more than five requirements for graduating, which is consistent with the desirability of setting substantive screening requirements. In the multivariate analysis, the number of graduation requirements reported by individual institutions was positively associated with student knowledge in five countries at the secondary level, although the results were more mixed at the primary level (appendix D). In most operations reviewed, the preservice institutions often rely on a single requirement for entry, such as level of completed education. This would be expected in education systems experiencing growth in enrollment because governments might not want to deter entry of candidates.

Support to exit or entry exams was limited in the World Bank operations examined; strengthening an existing examination function was the most common program feature related to screening. Nine of the 40 operations refer to exit or entrance examinations. Operations emphasized exit screening more than entry screening, considering the importance of ensuring that appropriate candidates enter the profession. An assessment of capabilities was used at entry into preservice to tailor training programs when students did not have competencies. For example, the Dominican Republic Support to the National Education Pact implements entry exams to ensure that new students with

low levels of preparation undergo intensive remediation before formally starting their course of study. TTLs in some of the focus groups were skeptical that these programs improved skill deficiencies. However, the Mauritania Basic Education Sector Support Project assesses competencies in French, Arabic, and mathematics to identify potential areas of improvement and tailor training to specific needs.

World Bank efforts to ensure adequate teacher training through the verification of content and pedagogical learning in operations focus largely on exit exams. Exit or proficiency exams might help guarantee a basic level of preparation in the necessary content, pedagogy, and methods,[3] and nine PADs indicated support for the development of guidelines for exit exams. The Democratic Republic of Congo's Education Quality Improvement Project, for example, administers exit exams to ensure that graduates effectively master the content of the basic education curriculum, such as the use of learning and teaching materials. Strengthening this exam was necessary because of concerns with the exam's transparency and adequacy.

In some cases, the World Bank tried to address the factors that contribute to ensuring that a steady supply of motivated students enters the teaching profession (Box 2.1). Multiple factors contribute to the demand for teaching positions, including benefits during training (such as subsidies), initial pay, career opportunities, incentive and support structures, classroom and school working conditions, and even cultural aspects related to how teachers are viewed in the society.[4] Nearly one-third of the PADs, 13 of the 40, refer to interventions related to improving the attractiveness of the teaching profession, which is promising given that recruitment of quality candidates will require a longer-term approach to move beyond filling immediate needs to strategically planning steps to improve the quality of candidates and attractiveness of teaching.

Box 2.1. Encouraging Good Students to Become Teachers

The World Bank has supported several ways to encourage motivated students to take up the teaching profession. The Dominican Republic's Support to the National Education Pact Project introduces a scholarship program that targets students from the top quartile of the university entrance exam. Moldova's Education Reform Project finances a new remuneration program designed to attract, develop, and retain teachers and school directors while enhancing performance. China's Guangdong Compulsory Education Project improves school facilities by creating dormitories to serve as temporary housing as an incentive to attract teachers to rural areas. An oversupply of candidates creates opportunities for effective filtering, but when the overall quality of the system is low, the potential pool of future teachers may be of low quality. The World Bank is trying to prevent an oversupply of candidates in the Democratic Republic of Congo through a study of the government's pilot preservice reform. The study examines ways to professionalize teacher trainees and the government's attempt to attract new graduates through career advancement while also providing adequate management.

The World Bank has also supported alternative routes to attract teacher candidates, in some cases more rapidly than through existing routes. Trying to attract teachers into the profession through alternative routes has sometimes addressed concern about teacher education system curriculums and effectiveness (Glazerman, Mayer, and Decker 2006). Alternative routes have potential to introduce some dynamism into the teacher training system, as the Teach for All Program has shown in numerous developing countries (Bruns and Luque 2014).[5] Twelve of the 40 World Bank PADs propose new short-term or accelerated training, delivered within an existing training modality. Only a couple of projects support a new training modality operating outside the current one (Box 2.2). TTLs in the Africa Region believed the candidates who enter preservice are the best available considering the low competencies of students overall in higher education. This could explain why only two operations supported alternate models.

Box 2.2. Using Existing or New Training Modes

Nicaragua's Second Support to the Education Sector Project coped with the scarcity of teachers and incomplete schools in rural areas through an accelerated preservice teacher training program for lower secondary education graduates. The young graduates were trained in teacher training institutions over six months to become fifth and sixth multigrade teachers.

Meanwhile, the Kazakhstan Education Modernization Project will undertake a pilot that will try to experiment with a new model of teacher preparation through a technical degree program and enhanced pedagogical coursework, practicum, and induction.

Teacher Preservice Coursework

World Bank studies are consistent with the frequent criticism of teacher training curriculums as outdated and emphasizing theory over practical knowledge applicable to classroom teaching. New teachers often lack preparation in how to manage the classroom and stimulate effective discussions, make content knowledge accessible to learners, perform informal (formative) assessments, and teach in heterogeneous environments. Even the basics of teacher preparation cannot be assumed in poor countries. Recent World Bank Service Delivery Indicator studies found that few primary teachers demonstrated a mastery of primary-level content: 3 percent in the Lao People's Democratic Republic (Lao PDR), 19 percent in Uganda, and 39 percent in Kenya.[6]

World Bank project documents frequently propose to develop teacher coursework, in most cases balancing academic content and pedagogical skills. The review of World Bank operations found that 70 percent of the projects (28 of 40) proposed to finance activities that would develop curriculums for preservice training, including new development and redesigning of preservice training programs with the focus on subject matter and pedagogical methods. The results from the Teacher Education and Development Study in Mathematics (TEDS-M; see appendixes A and D) provide some guidance on coursework that is consistent with the limited empirical literature that models student achievement on the basis of the teachers' preservice training experience (Boyd and others 2008; Goldhaber 2019; mathematics course work study). This analysis identified some classroom features and coursework topics that were positively associated with trainee pedagogical content knowledge, including asking questions during class time, participating in class discussions, exploring how to apply mathematics to real-life problems, and exploring how to use manipulative materials to solve math problems.

The complex matter of training duration requires careful analysis and clear decision-making criteria that are not apparent in the World Bank's approaches. The optimal duration of training programs depends on multiple factors. The TEDS-M data offer evidence that longer programs produce higher levels of trainee content knowledge and pedagogical content knowledge. However, the same is not true for secondary-level trainees.[7] One potential explanation is that education systems use stronger filtering for secondary-level teacher candidates than for primary-level teachers. This finding from TEDS-M could suggest the consequential role that training could play for teachers with limited background knowledge, but more research is needed to confirm this hypothesis. Longer programs are not necessarily effective when quality is low. Furthermore, the optimal length of training relates to the minimum skill level of the average teacher trainee, which again highlights the importance of effective screening measures. Eleven

World Bank operations appear to be lengthening the preservice training period, but five programs propose to shorten that period and supplement it with on-the-job training. The factors driving these differing choices are not clear, and it is unclear whether the reduced training duration is temporary or permanent (Box 2.3).

Box 2.3. Reduced or Increased Training

Nicaragua's Second Support to the Education Sector Project will reform its teacher training module to include all grades modalities, which will increase the length of studies from six months to one year. Instead of recruiting candidates with 11 years of education, the government will select students with only 9 years of education and provide an additional year of training to become teachers.

The World Bank frequently supports capacity development for teacher educators, perhaps because ineffective teaching methods observed frequently in less-developed country classrooms—such as rote memorization and copying—are also common in preservice training classrooms (Béteille and others 2018; UNESCO 2012). To ensure training efficacy, most teacher training institutes in TEDS-M countries require mathematics content course instructors to have an International Standard Classification of Education 6 degree. However, based on descriptive summaries and multivariate analysis, there is no clear pattern between teacher educator credentials and trainee outcomes like content knowledge. The guidance from high-performing systems is that teacher educators are hired based on a portfolio of skills with transparent selection procedures because the effectiveness of teacher educators is critical. For this reason, approximately half the World Bank operations provide some form of capacity development to teacher educators focused on pedagogical methods versus content or subject matter, which is important (figure 2.2).

Figure 2.2. Capacity Building Activities for Teacher Educators
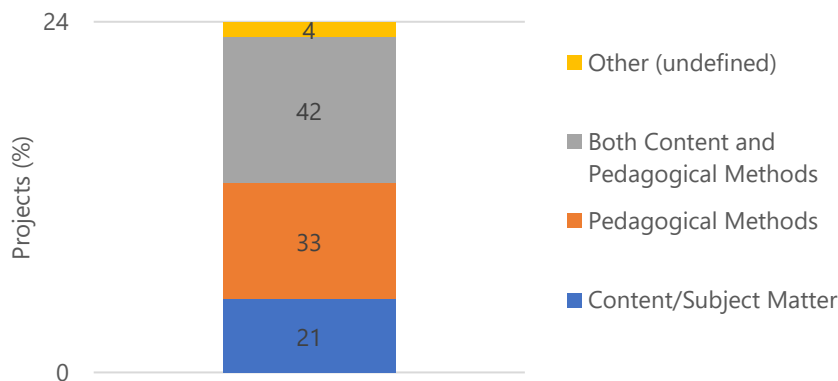
Source: Independent Evaluation Group coding of project appraisal documents.
Note: n = 24.

Box 2.4. Building Teacher Training Capacity

Capacity building includes providing workshops, conferences, and exchanges with local and international institutions to enhance trainers at teacher training institutes. For example, a Kazakhstan project proposes to finance technical assistance to pilot a preservice training program for university staff to enable them to teach one subject using English as the language of instruction. The technical assistance will cover English language instruction and ongoing support to the teacher trainers. A project in Vietnam proposes to build the capacity of lecturers and managerial staff of teacher training institutions by providing relevant training programs and courses, holding national and international conferences and workshops, engaging in scholar and academic exchanges, and reviewing and renovating regulations on recruitment, work position, and other procedures related to human resources. Likely government preferences constrain the World Bank from complementing capacity development with other intervention (as in Vietnam) to monitor the performance of teacher educators, consistent with practices in high-performing countries.[a]

a. Guidance from high-performing systems is that teacher educators are hired based on a portfolio of skills with transparent selection procedures. The core challenge is matching the teacher educator's skills with the course demands. It is not enough to set minimum education or experience levels to determine teacher educator credentials, although these can help avoid hiring inappropriate staff. High-performing systems also provide clear guidelines about performance, and the environment is positive and rewarding.

More than half of the reviewed World Bank projects consider the conditions in which teacher preservice training takes place, including infrastructure and materials—especially locally relevant materials. Deficiencies in materials are common in less-developed country descriptions of teacher training experiences (Lewin 2004; Sorto and Luschei 2010). Training centers in poor countries also often lack basic infrastructure and amenities, which can affect both the quality of the learning experience and the attractiveness of the profession. About 48 percent of the World Bank projects finance infrastructure or renovations for preservice training institutions. Seventy-two percent of operations supported soft infrastructure, including teaching materials such as textbooks, videos, and information and communication technology.

Teaching Practicum

World Bank clients often give too little attention to practice teaching, a critical phase of teacher preparation (box 2.5). Support for this activity can begin with a national policy that defines practicum features, such as minimum duration for classroom experiences and rules for supervision, support, and responsibilities for both the teacher training center and the school where the practice takes place. This type of systemic guidance is typically missing in less-developed countries, which leaves individual institutions in charge of defining the specifics. The practicum is nearly universal among TEDS-M participants, but there is a lot of variation in practicum length (Figure 2.3).[8] There is no

agreed-on minimum number of days for practicums—the critical aspect is meaningful experiences with time for practice, reflection, and feedback.

Box 2.5. The Importance of the Practicum

In the Teacher Education and Development Study in Mathematics data, only 1.5 percent of the secondary teaching training institutes did not offer teaching practice. However, in the Southern and Eastern Africa Consortium for Monitoring Educational Quality and the Latin American Third Regional Comparative and Explanatory Study (TERCE) data, significant percentages of teachers did not participate in a practicum (see appendix A for an explanation of these data sets and appendix D for the analysis of that data).

This has important implications for learning outcomes. The association between practicum experience and student achievement was analyzed with TERCE data. The results show some evidence that student achievement is higher in classrooms where the teacher reports experience with a practicum, and the association is more significant (and larger) among younger teachers, for whom the practice experience is more recent.

Figure 2.3. Total Hours in Practice Teaching by Training Level and Country, TEDS-M
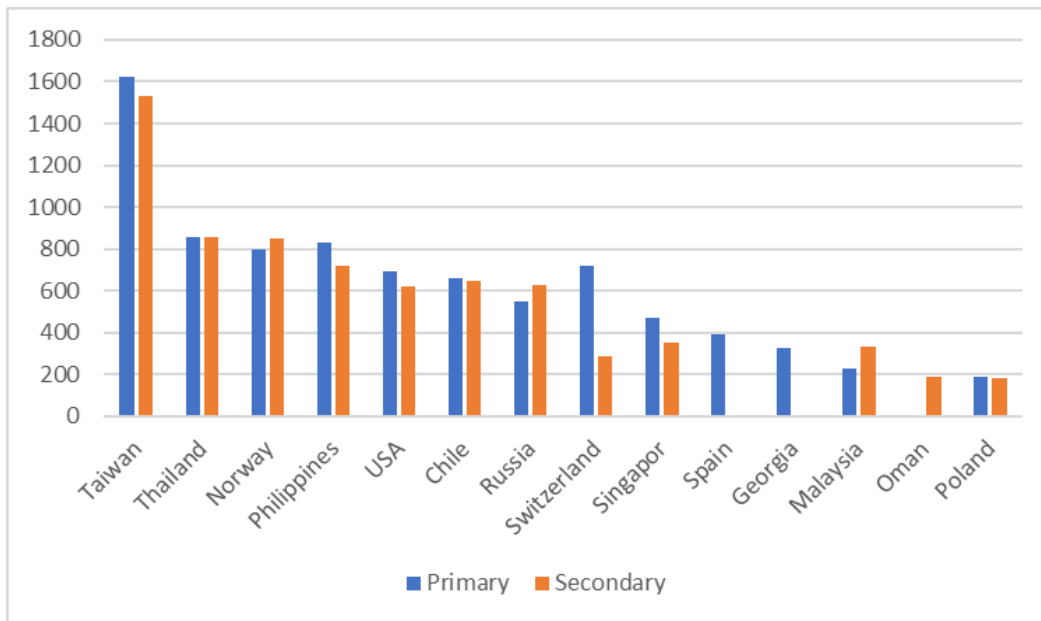
Source: Independent Evaluation Group analysis of the Teacher Education and Development Study in Mathematics.
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

The World Bank attention to practicums in its projects is low relative to the need. Fifteen of the World Bank projects support mechanisms related to practicums, which is low considering it is often deficient in less-developed countries (Akyeampong 2017). TTLs consistently affirmed the importance of the World Bank's support for enhancing the nature of the practicum to focus on practice by teacher trainees with quality mentors.

For this reason, the limited support for this aspect in the operations examined is surprising.

When it does support practicums, the World Bank has taken a variety of approaches, overall aligned with some key desirable features. A common deficiency in the implementation of practicums is "sink or swim," an approach in which trainees are given too much autonomy too soon, highlighting the importance of a supported experience. A critical issue is the lack of qualified teacher mentors in the schools where trainees spend their practicum time. This is a difficult situation to correct in countries with widespread quality deficiencies. The TEDS-M data suggest that the frequency at which trainees are observed in the classroom during the practicum is another important element. One negative factor is when trainees use methods in their practicum that are different from what they learned in class. Finally, the most consistently positive predictor of pedagogical content knowledge (significant in four countries) is the frequency that trainees reported having to demonstrate their ability to apply teaching methods learned in coursework in their actual classes, which suggests the necessary alignment between coursework and practicum. The World Bank has supported practicums in multiple countries (**Error! Reference source not found.**).

Box 2.6. World Bank Support for Practicums

In some instances, the World Bank's support for practicums has been extensive, linking with multiple aspects of preservice training and incorporating key principals (Mauritania). Project documents also refer to selecting schools for the best practicum experiences (Haiti), training supervisors to reinforce practicums (Guinea), strengthening the links with practice school partners (the Democratic Republic of Congo and Ethiopia), exposing trainees to urban and rural school settings during practice teaching (the Democratic Republic of Congo), or the development of a national policy on practicums. All these aspects are encouraging because teacher trainees need to practice with support in a context like they will face in their work.

Quality Assurance for Preservice Training

For various reasons, the World Bank has given limited attention to quality assurance or accreditation. The TEDS-M study on quality assurance provides an excellent summary of accreditation's role in affecting the quality of teacher preparation (Ingvarson and others 2013). The data show that countries with weak accreditation systems have no effective control over training institutions or rely on voluntary participation mechanisms.[9] TTLs highlighted this reality, noting that their country clients either have no accreditation mechanism or have one that is subject to political interference. Governments often oppose efforts that would "shake the comfort zone." However, some cautioned that accrediting a flawed system would not solve the problem. Thus, nine of

40 World Bank projects support processes or institutions that conduct accreditation. Operations in Ethiopia and Bihar, India, finance quality assurance mechanisms for preservice training institutions, which might provide better accountability for what the institutions are doing. Effective monitoring and accreditation mechanisms regulate teacher education program providers to ensure adherence to training standards and remove political influences. These mechanisms also regulate entrance and exit examinations to ensure quality.

Very few of the World Bank PADs examined refer to specific monitoring and inspection functions for teacher training institutes, with support related to policies, materials, and procedural improvements (and training). Accreditation is only one means of quality assurance; other actions, such as monitoring and other support, are also important (box 2.7). Projects that support the preparation of early childhood educators typically contain quality assurance activities to monitor services or certification because the subsector is often unregulated.

Box 2.7. Monitoring Quality

A few World Bank projects address monitoring activities, and among them is the assistance to monitor practice teaching by the Departmental Directorate Offices of Pedagogic Support in Haiti's Education for All (Phase II) Project. Similarly, in the Dominican Republic, support is provided to several agencies with responsibility for implementing and evaluating policies for preservice training institutes. Early childhood education projects promote quality assurance of the services, but few of them monitor the standards in teacher training institutes, as the project in Yuan, China does.

One consequence of the low quality of preservice training systems is the need to use in-service training to address teachers who are underprepared or unqualified. There is an argument that the task of addressing teacher content knowledge deficiencies is so challenging in poor countries that it should be taken out of preservice training institutions and based instead in an intermediate (or parallel) system in which trainees upgrade their basic knowledge (Lewin 2004). Consistent with this approach, one-quarter of in-service training projects supported teachers with limited preservice training and used on-the-job training to improve their skills and qualifications. For example, the Ghana Global Partnership for Education helped teachers in isolated areas complete a formal certification if they lacked these credentials. This training program also addressed weaknesses in these teachers' competencies in literacy, numeracy, and science. The literature offers no guidance on the optimal duration for such use of in-service training.[10] The World Bank's targeting of low-skilled teachers is important because this group needs more guidance to reach minimally acceptable levels of instruction (Ganimian and Murnane 2016). The World Bank has found it easier to work

with ministries of education to improve in-service training rather than reform individual preservice training institutions.

Although reform of preservice training institutions is more complex, the rationale for attending to both types of training systems simultaneously is clear. Without improvements in the candidates exiting from preservice institutions, stronger screening by the ministry throughout the teacher's career becomes necessary. In-service training then becomes the sole quality driver for teacher professional development and the way to address the needs of unqualified or undertrained teachers. Reliance on in-service training carries costs for ministries of education because in-service training needs to address many competing goals. Improving teacher quality will require improvements in the candidates graduating from preservice training institutions—an aspect that in-service training alone cannot address. Results from the World Bank's SABER (Systems Approach for Better Education Results) Teachers highlight the low minimum levels of education and preservice classroom practice (especially in Africa), which have far-reaching consequences for these systems (box 2.8).

Box 2.8. SABER Teachers: The Reality of Preservice Institutions across the Globe

Systems Approach for Better Education Results (SABER) Teacher data have been collected in more than 40 countries and regions, focusing on Sub-Saharan Africa, Middle East and North Africa, and Latin America and the Caribbean. The tool has two domains relevant to preservice training: attracting the best candidates into teaching and preparing teachers with useful training and experience. These domains are consistent with other reviews that focus on filtering and structural features of the profession that affect teachers throughout their careers.

Overall, the SABER Teacher data show that between 40 and 60 percent of the SABER countries were classified as latent in three subindicators: teacher pay, working conditions, and minimum standards for becoming a teacher. These countries are relatively more effective in teacher policy areas related to entry requirements, career opportunities, and classroom practice. Teacher pay levels are perceived as low, and working conditions can be quite difficult. The teaching career is often perceived favorably for long-term benefits, which helps to counteract the effects of low pay. The fundamental challenge for preservice training is attracting teachers with high levels of education and the best performing students in high schools. The results show that the Sub-Saharan African countries generally have the lowest results, with a number of latent or emerging classifications, and very few established. By contrast, the Middle East and North Africa and South Asia Regions have fairly high scores.

Similarly, better alignment between preservice and in-service training is also needed. Strategic links were lacking between preservice and in-service training in all cases that IEG examined except for one. In the positive example, preservice institutions were part of the delivery of in-service training. Thus, in the remaining cases, in-service training compensated rather than complemented what was learned in preservice, suggesting a need for explicit links in operations.

3. Where Teachers Grow: In-Service Training

Highlights

- Discipline-specific training that models adult learning style and adapted to teachers' needs and capacity with follow-up support have not received the attention needed under prevailing circumstances.

- Well-designed and well-implemented training programs can improve teachers' pedagogical practices, but they cannot do it alone. The education system needs to create an enabling environment for teacher professional development.

- Training programs should provide robust monitoring and evaluation data to give a clear indication of progress and outcomes.

- Scaling up of training programs needs to be considered at all stages, from planning through monitoring and evaluation. Although some well-planned and well-implemented scaling-up processes have resulted in success, some of the desirable conditions for each stage to ensure sustainability and depth of the training programs are missing in the World Bank's current approach.

The 108 training operations focused mainly on primary education in low- and middle-income countries and used a variety of modalities. The most commonly used were teacher training centers, school-based training, and distance training (17 percent each). The cascade approach was by far the least common modality (9 percent). Details on the characteristics of the in-service training operations are in appendix B.

Key Features of In-Service Training in World Bank Operations

In-service training can improve teachers' performance, knowledge, skills, and motivation. Effective training, implemented with the key features shown in the conceptual framework, is also linked with improvements in student learning (Darling-Hammond, Hyler, and Gardner 2017; Popova and others 2018). The essential features are adequate duration, discipline specificity, active and applied learning based on teachers' needs and capacity, and follow-up support to provide opportunities for feedback and reflection. Some of these features were evident in nearly half of the World Bank operations examined (table 3.1) but are not present together.[1] However, well-designed and well-implemented training programs alone cannot improve teachers' pedagogical practices, suggesting the importance of the broader education system to create an enabling environment. Finally, sustained scale-up is another key feature of effective in-service training.

Table 3.1. Key In-Service Features

Key Feature	Projects (no.)
Adequate duration	49
Focused on discipline	57
Follow-up support	54
Adapted to teachers' capacity	42
Every key feature	10

Source: Independent Evaluation Group coding of project appraisal documents.
Note: n = 108.

Adequate Duration

Training supported by the World Bank meets at least the minimum requirement for duration. Adults need time to learn and apply a new skill, so training needs to be spread over time.[2] Of the 51 projects that provide information about the duration, all but two of them provided the minimum amount of time thought to be required (50 to 60 hours over multiple days), typically at a time when students were on recess. Most of the training sessions last between 5 and 20 days. Four projects promote ongoing training, suggesting that more than the minimum amount of time is provided, considering the reference (from research studies) is adequately prepared teachers in the United States.

Discipline Focus

The World Bank focuses mainly on pedagogical training, though training in the relevant subject matter is equally important for teacher quality. Effective training programs focus on the content to be taught—mathematics, science, or literacy—and give teachers the opportunity to study their students' work or a particular element of pedagogy in the relevant content area (Darling-Hammond, Hyler, and Gardner 2017; Garet and others 2001; Popova and others 2018). Some studies show that discipline-specific training programs produce greater learning gains (Darling-Hammond, Hyler, and Gardner 2017; Popova and others 2018). However, pedagogy was the focus in about half of the World Bank PADs (figure 3.1), regardless of the school level. Given the relatively greater importance of course content at the secondary level, the expectation would be to see more discipline-focused training for secondary teachers, but this was not the case. This observed focus on pedagogy might have resulted from the need to address shortcomings in preservice preparation. TTLs now recognize the need to focus much more on discipline-specific content, given teachers' limited capacity to teach numeracy, literacy, and science.

Figure 3.1. Content Focus of Training Programs
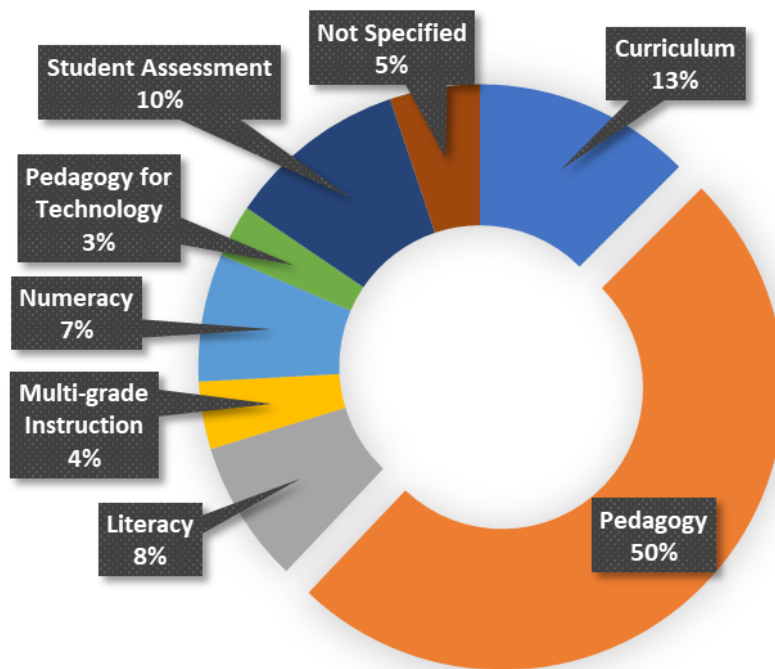
Source: Independent Evaluation Group coding of project appraisal documents.

Training supported by the World Bank also addresses the teaching of thinking and communication skills. The fieldwork for this evaluation confirmed a broader focus beyond pedagogy in the training programs. Ideally, training programs should equip teachers with knowledge and skills to promote a range of skills among students, such as critical thinking, communication, and collaboration, among others (Reimers and Chung 2018). TTLs asserted that training helps teachers focus on students' skills development, including higher-level thinking skills and socioemotional skills that are needed to succeed in the workplace. The impact evaluation for the Vietnam Escuela Nueva Project (VNEN) found that the program, which was supported through extensive training, had a positive result on the socioemotional skills of children enrolled in supported schools (Parandekar and others 2017). A notable omission overall is technology learning, which was mentioned in only 3 percent of PADs. However, TTLs considered the limited focus on getting teachers to use technology was appropriate, given the context in those classrooms.

Learning Environment

World Bank operations have unevenly applied active learning that addresses adult learning styles. Adults learn through application, modeling, and demonstration by engaging teachers directly in designing and trying out teaching strategies rather than

lecture and discussion.[3] Fieldwork revealed that some training programs embodied adult learning principles, which was consistent with TTL reports. One explanation for this was that trainers were unfamiliar with the concept and the techniques that facilitate adult learning. Some TTLs noted that modeling was applied during coaching sessions; however, they were unsure whether training sessions were reflective of how adults learn. When teachers acquire new knowledge and apply it, there is a greater chance of influencing teaching practices, suggesting an aspect to improve in future training programs.

Box 3.1. Active Learning in World Bank Projects

A good example of active learning from a World Bank project is the training that the Vietnam Escuela Nueva Project uses. The initial workshop actively engaged teachers by putting them in the role of students to experience a project classroom. This allowed the teachers to learn how to become facilitators and advisers of children's learning—a dramatic difference from the teacher-led instructional practice they used and were taught during preservice training. The training program in Uruguay practiced teaching by using modeling, examining student work, and watching videos. By contrast, training programs in Ghana focused on discussion and listening rather than modeling, according to teachers who participated in the training, though it is possible that some sessions used learning techniques that are more active.

Meeting Teachers' Needs, Capacity, and Context

Effective professional development is designed according to teachers' needs and takes account of the teaching context and capacity; operations typically rely on data sources other than feedback from teachers. Training needs to be connected to teachers' actual practices and classroom context (Darling-Hammond, Hyler, and Gardner 2017), aligned to their needs (Westbrook and others 2013), and appropriate to their level of experience (Popova and others 2018). This can be derived from an analysis of needs obtained from teachers or feedback from teachers (Luneta 2012; Villegas-Reimers 2003). Box 3.2 illustrates the stages of development in the education system and the implication for teachers and the design of training programs to facilitate movement to reflective professionals. Having a solid data foundation to design the training program is an important prerequisite (Westbrook and others 2013). Teacher focus groups were rare, however, with only three operations mentioning use of this method.[4] TTLs agreed that design often omitted asking teachers what they needed, which they viewed as a shortcoming. Without such consultation, it is difficult to gear training specifically to their needs and capacity. Soliciting teacher feedback can also help avoid common pitfalls with training, such as a one-size-fits-all approach, or derived by central-level planners (Luneta 2012), suggesting that teacher consultation may need more emphasis. However, other sources of evidence were also used to design the training program, such as

international literature, SABER, classroom observational data, needs assessments, learning data, or the curriculums.

Box 3.2. Stages of Development and the Implications for Teacher Training

---

An education system has several stages of teacher development:

- Unskilled teaching: those who are unprepared. Training programs should focus on improving their content and pedagogical knowledge and skills in a structured manner, such as scripted lessons, and education systems should provide supervision and support.

- Mechanical teaching: teachers instruct in a mechanical way. Training programs need to provide teachers with a variety of guides and textbooks, information, and modeling of new teaching techniques. They also need continuous support.

- Routine teaching: teachers have a limited repertoire of methods. Training needs to expand their methods and knowledge to give them a broader set of experiences.

- Reflective teaching: teachers are able to change their repertoires of methods to their own circumstances. These teachers are reflective and skilled; thus, training needs to draw on their reflective ability to further support their development.

Consistent with these stages, Aslam and others (2014) stresses that it is not about a particular method but the ability of teachers to select among methods.

Source: Villegas-Reimers 2003.

Some systems solicit teacher input when designing training. This might be related to the context in which educational systems operate, with more centrally managed systems taking a narrower view of consultation. This could explain why teacher feedback in both training programs in Vietnam (VNEN and the School Readiness Promotion Project) did not gather teachers' input in the design stage. By contrast, training programs in Uruguay broadly consulted across a variety of stakeholder groups, including trainers, inspectors, school directors or principals, teachers, and members of education councils, including representatives of the teachers union. During the design process, developers also consulted educational experts and those in other fields such as psychology, sociology, and the arts. The Ghana Secondary Education Project used school plans, needs assessments, and current student results to develop the focus of the training. The primary teacher training program in Ghana also consulted widely with teachers and teacher colleges responsible for delivering aspects of the program. Representatives of teachers unions were not included in the process, however.

Although less than 40 percent of World Bank operations (42 of 108) were geared to teachers' capacity, awareness of the need to do so could be increasing. This feature was distributed across country income levels, suggesting an emerging recognition of the need for tailored programming. TTLs reported that some of the cases addressed a

specific need of untrained or underprepared teachers, but the majority of them built specific competencies, such as English skills. Fewer projects addressed other types of capacity, such as the Lebanon RACE (Reaching All Children with Education) Support Project that targets support to teachers based on classroom observations and criteria established in the teacher standards. The training program in Uruguay focuses flexibly on teaching practice through school-based training. Trainers customize their support according to the school environment or teachers' specific challenges. Such an approach develops teachers' capacity to be more skilled and reflective practitioners (Box 3.2).

The World Bank's targeting of low-skilled teachers is important because this group needs more guidance to broaden their repertoire of skills. Unskilled and mechanical teachers need more highly structured materials and scripted lessons (Ganimian and Murnane 2016). They need support tailored to their (low) skill level, such as flashcards for teacher-led or student-directed drills or other individually paced learning materials (Evans and Popova 2016). Appraisal documents rarely specified complimentary materials, so the extent to which training programs provide materials beyond textbooks is unclear. In the Democratic Republic of Congo, lessons with a script of subject information were to be provided to primary teachers in math and French so that teachers deliver accurate content, especially those with weak content knowledge or pedagogic practice. Low-capacity teachers are a concern across country income levels, which is another reason to ensure that training addresses teachers' capacity.

Fieldwork found that training programs mostly are not differentiated by student grade, age, or ability level. Matching instruction to student learning or level is important to student achievement (Evans and Popova 2016, Ganimian and Murnane 2016, among others). There are many strategies to accomplish this aim. For example, teachers who received coaching used small group instruction to provide more individualized instruction and more time for students to practice reading.[5] Fieldwork found limited application of such techniques in the reviewed cases. Primary teacher training in Ghana, for example, included training in literacy, numeracy, and science but was not geared to student grades or levels. Similarly, the early childhood education training in Vietnam provided the same training for teachers who teach children ages three, four, and five, with examples to illustrate the significant development differences in learning among these children. The VNEN training, although not differentiated, was likely justified because all teachers were new to the model and likely at a similar starting point. The secondary education training courses in Ghana were responsive to the context of deprived districts and students whose exposure to science, technology, engineering, and math courses was more limited. Uruguay's training program focused on underperforming students. Considering that effective teachers place their students at the

center of the teaching-learning process, training programs should attend more to this aspect (Westbrook and others 2013).

Follow-Up

World Bank operations that offer opportunities for feedback and reflection tend to be school-based or distance training. Adults require time to apply, practice, and test new learning.[6] Follow-up support like this is not yet common in World Bank–supported training programs. TTLs noted that coaching is important to training, which is consistent with evidence showing that this approach can shift teaching practice and improve student learning in developing contexts.[7] TTLs also reported that they are being encouraged to include coaching in their operations, and 54 of the 108 operations specified it, particularly in recent operations. Mentoring or coaching (44 PADs) was the most common format. Programs also used follow-up training (9 PADs) and text messaging (1 PAD). Projects using training workshops and a cascade approach seldom offered formal follow-up support. Thus, the training modality might matter.

Early World Bank experience with coaching has revealed some issues, including level of quality and variation in quality. Coaches help teachers apply material covered in the training program, and sustained coaching has positive effects on teachers' practices (Conn 2014; Popova and others 2018). Experienced or expert coaches have also been shown to be effective in changing the instructional practices of untrained or undertrained teachers, but this type of follow-up support is expensive and labor intensive (Orr and others 2013). Coaching has been implemented in the Democratic Republic of Congo, Lao PDR, and Pakistan, where observational and instructional leadership skills—prerequisites for coaches—might not be thoroughly developed in these contexts. These coaches will likely need training, follow-up support, and mentoring in their new role. World Bank operations most frequently draw coaches or mentors from among trainers, head teachers, or principals. Inspectors or other ministry staff are mentors less frequently used (Box 3.3).

Box 3.3. Finding Good Coaches

The training program in Punjab illustrates the need to select the right people to be mentors or coaches because the inspectors from the district or province continued to function in their usual manner. As a result, "No teacher could share any concrete or substantive teaching tool or technique as an example learned from the [coach]" (World Bank 2016, 86). According to teachers in Ghana, the mentoring program had substantial variation in its effectiveness—from regular school engagement and planning to virtual nonexistence. A significant part of the coaching was dependent on the district office's capacity and commitment, which the TTLs acknowledged in focus groups. This highlights the need for greater accountability and monitoring for follow-up support.

Fieldwork showed that some systems bring together the participation of teachers from the same school or grade level to encourage peer learning or collaborative work but not systematically (Box 3.4). Collaborative work or observations make teaching practices more public and open teachers to feedback (Darling-Hammond, Hyler, and Gardner 2017). This process helps build teacher reflection.

Box 3.4. Building Peer Learning

In Vietnam's education system, school visits, principal feedback, and peer-to-peer learning at school cluster meetings, among other mechanisms, provide teachers with follow-up support. Hence, the Vietnam Escuela Nueva Project (VNEN) used follow-up support quite extensively. During the school year, two to four technical support team meetings were conducted for VNEN schools. These involved classroom observations, interviews with teachers and school management, review of logbooks, and exchange of experiences with principals and teachers. VNEN also developed a website to post and share lessons, videos, and examples. However, despite the availability of these mechanisms, there was a high level of heterogeneity in its implementation (Parandekar and others 2017). Follow-up support provided by trainers in Uruguay was not systematic and was based on teachers' initiative. Contact with trainers, who teachers said were responsive and knowledgeable, occurred through email and by phone.

Sustained follow-up support is needed because of the significant pedagogical changes required of teachers. The types of changes envisioned in the training programs visited in the field were not simple or easy, and teachers often started at a disadvantage given their skill levels and gaps in their preservice training. Thus, teachers needed more on-the-job support to understand complex topics introduced in their training more fully. This highlights the need for the World Bank and its clients to commit to support teachers continually for the long term.[8]

Enabling Education System Environment

The coherence of training programs' aims with the education system they serve can influence training impact (Villegas-Reimers 2003). In-service training needs to reflect key features of the enabling education environment and context, such as management, governance, and finance (Villegas-Reimers 2003). Hence, training can be more effective when it is part of a larger reform effort (Garet and others 2010, Darling-Hammond and Richardson 2009), aligned with standards and assessment (Darling-Hammond and Richardson 2009), and embedded in the curriculum (Popova and others 2018). Nearly all the appraisal documents examined lacked detail or analysis of these structural features, but these analyses may have been contained in other documents.

The fieldwork and focus groups with TTLs provided examples of where the education system was not consistent with the aims of the training. When rolling out a new curriculum does not accompany training (76 percent of the cases do not), coherence may

be lacking between the aim of the training and the existing curriculum. Such timing issues can undermine effectiveness (Box 3.5).

Box 3.5. The Timing of Training Matters

In the Vietnam Escuela Nueva Project (VNEN), timing of the introduction of the model and respective teacher training was problematic considering the existing curriculum and the aspirations for the new pedagogy. Key stakeholders reported the pedagogical changes envisioned under VNEN would ideally have been sequenced subsequent to reframing the overall objectives of the education system and the consequent revision of the curriculum, which would then open the way to the production of appropriate learning material and other supports. However, the VNEN had to use the existing curriculum and single text, even though it made dramatic changes in how teachers would instruct. The project produced textbooks based on the existing but essentially outdated curriculum.

Some training operations support activities to improve sector governance and student assessment to create an enabling environment to complement the effectiveness of training programs (Figure 3.2). Sixty-nine percent of World Bank operations supported governance—often governance linked with teacher management and assessment. For example, Bihar, in India, had inadequate accountability and incentive mechanisms, resulting in issues with teacher motivation, deficiencies in planning, and monitoring and managing of teachers. In the Democratic Republic of Congo, the roles and responsibilities of state and religious organizations were unclear, resulting in layers of administrative offices that can impede efficiency and accountability. Operations in both places strengthened institutional capacity for better teacher management. Forty-four percent of the training operations examined established a new assessment or strengthened an existing one, including getting teachers to understand and use the data. The key issues identified in PADs were then addressed, except for educational finance (38 percent), taking at face value that the issues presented were the most critical ones. It is not clear why the PADs rarely identified curriculum (15 percent) and school management (10 percent) as a major issue.

Figure 3.2. Key Issues in Project Appraisal Documents and Areas Supported

Share of projects by key issues and support (n=108)
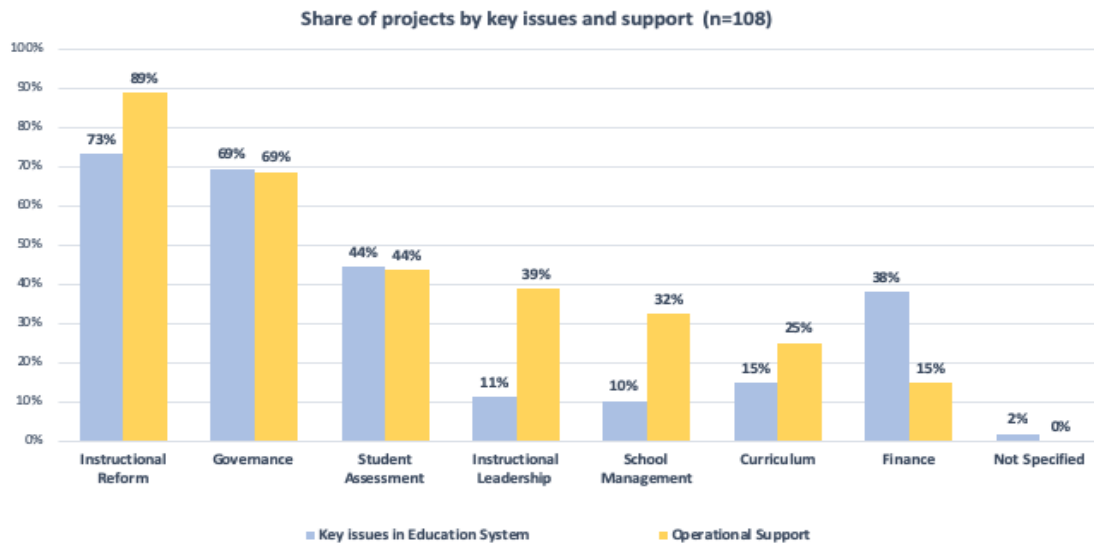
Source: Independent Evaluation Group coding of project appraisal documents.

Although 73 percent of PADs name instructional reform as a key challenge, World Bank–supported training programs are not consistently implemented within a broader framework for teacher development or teaching standards. Training was the main activity supported to improve instruction, which would likely be inadequate because operations do not embed the training program within a broader framework consistently and simultaneously. For example, in São Tomé and Príncipe, the World Bank developed a competence-based training framework that defines critical competencies, training plans to respond to these needs, and a certification process. TTLs considered anchoring training in such a framework important to improving training. This would require greater clarity of the training outcomes than currently put into planning efforts, according to some TTLs. It would also help advance another point shared by TTLs: the need to link training to teachers' career opportunities and ladder.

Feedback from TTLs and the literature suggest the importance of a career framework to provide teachers with incentives to facilitate lifelong learning and development, but the World Bank's attention in this area tends to focus on certification and qualification rather than career development. Incentives for training are important (Popova and others 2018), and opportunities for professional development are needed throughout the teaching career, likely accompanied by filtering or screening mechanisms, as suggested by the analysis of high-performing countries (Appendix DD). In these countries, the screening and support mechanisms diagnose weaknesses and provide sustained support. These systems also recognize that teachers' capacity grows over the course of a career (Reimers and Chung 2018).

Teacher training cannot be effective without adequate instructional leadership from head teachers, school principals or managers, and school inspectors. Instructional leadership skills were lacking in principals in West African countries—skills typically not developed before assuming this new role (Bush and Glover 2016). However, principals have to deal with inadequately prepared teachers and build a school culture that focuses on learning and collegiality, which are important enabling conditions for training and peer learning. Thus, even if teachers intend to implement training, instructional leaders must also support the practices (Villegas-Reimers 2003).

Despite its importance to nurturing a culture of support and collaboration, relatively few World Bank operations recognize instructional leadership as an issue and invest in developing the capacity of head teachers and school leaders. Few operations (11 percent) examined recognized instructional leadership as a key issue (Box 3.1), and 39 percent of them would improve the capacity of head teachers, inspectors, or school leaders. This is a small number given the challenges that low-income countries face. These leaders must have the instructional skills to model the practices and give feedback, and to foster positive and nonbiased attitudes among staff (Aslam and others 2014), which is important for learning among poor or ethnic minority students. Some operations seek to replace unqualified principals with more qualified ones (Sri Lanka, for example). Capacity development such as in Ghana is more typical, where the project supports school leadership training to improve teaching, coaching, school management, and teacher assessment. Such operations are developing the necessary instructional leadership skills, but likely too few projects are doing so.

Training programs should provide robust monitoring and evaluation data to give a clear indication of progress and outcomes. Such data serve multiple purposes: providing information that can be used to adapt and improve the program and providing accountability and transparency. Three types of data could be collected: monitoring the implementation of the training program (including feedback from trainees), monitoring the outputs of the training program, and evaluating the training program and its outcomes. These data are important for the training programs' sustainability and ensuring that the training program meets teachers' needs.

The World Bank could give more attention to evaluating the training programs it supports. Evaluations can improve program effectiveness, resource use, and learning in the World Bank and among country clients. Less than half of the operations plan an evaluation of in-service training, and many fewer plan a rigorous impact evaluation. Classroom observations and teacher testing were the most common methods used to measure the performance changes of teachers resulting from training. Fifteen projects plan to implement an impact evaluation,[9] including projects in some challenging

country contexts. Outcomes predominantly relate to changes in teachers' knowledge and practice rather than student learning, which is the ultimate measure of whether training influences what teachers actually do in the classroom.

Box 3.6. Use of Evidence and Learning in Ethiopia

Analysis identified the quality of teachers—both current teachers and new graduates—as the priority issue for the education system in Ethiopia because teaching is the career of last resort. When a graduate cannot find a job, he or she takes an additional one-year course to become a teacher. The Ministry of Education has taken steps to reform preservice requirements and eventually improve the stock of candidates. First, candidates will need to identify their interest before graduation and take pedagogical and content coursework during their studies. Second, lessons learned from what was not working in previous training was applied to school-based follow-up in the General Education Quality Improvement Project to improve observations and supervision by school leaders.

Source: Ethiopia Ministry of Education.

Monitoring may need more attention to assess fidelity and implementation of the training and follow-up support to ensure its effectiveness. Data of this type can help identify noncompliance or lack of consistency and bring greater oversight to training. Among the countries visited in the field, the fidelity or implementation of the training was monitored in Uruguay, but this observation might not reflect actual practice because most TTLs reported examples. A review conducted in a sample of full-time schools in Uruguay found weaknesses in teachers' pedagogical approach. The review discovered that assignments do not stimulate students' interest in reading and writing, and there were limited links between classroom activities and didactic sequences (Bentancur and Gabbiani 2016). Ethiopia provides another example of how to use evidence to improve the current training program (box 3.7), an aspect present in operational designs and, to some extent, in training program design. One project is trying to find out if existing mechanisms at the school level can support effective coaching and mentoring. The Early Childhood Education Project in Lao PDR is testing two approaches, and the evaluation will help answer implementation questions related to follow-up support. The monitoring team in the Yuan Early Childhood Innovation Project will assess the follow-up support (aligned with key features) to assist the provincial department of education's efforts. Despite these notable examples, TTLs agreed that more effort should be put into monitoring and believed more should be expected from training.

Simple tools can address capacity barriers, suggesting an important role for the World Bank. Capacity to manage and analyze monitoring and evaluation data was a large barrier, according to TTLs. Simple apps or tablets have helped to immediately digitize the data and avoid piles of paper. These tools help in two ways: They reduce the

significant amount of staff time needed to input data, and they make it easier to give immediate feedback to central and decentralized ministry staff about what is practiced in schools and classrooms. This same process could be applied to monitor the implementation of training and follow-up support. TTLs were enthusiastic about a role for the World Bank's observational tools, *Teach* (see Box 3.7) and *Coach* in future data collection.

Box 3.7. How Has *Teach* Supported Learning?

The World Bank recently developed an open source teacher observation tool appropriate for low- and middle-income countries. The *Teach* tool assesses the quality of the instruction to support students' cognitive and socioemotional skills and provides diagnostic information for the education system. For example, the data can be used in planning a training program to address teachers' strengths and weaknesses. From the early pilots, *Teach* was used in Guyana to inform teachers' professional development because results showed instructional weaknesses, including their attention to children's socioemotional development.

Source: World Bank 2019 and interviews.

Scaling Up

To get the most from in-service training, the Education Global Practice emphasizes the need for effective scaling of teacher training. Effective scaling involves a sustained expansion of coverage while ensuring the depth of change necessary to support and sustain lasting educational improvement (Fullan and Quinn 2016, for example). To sustain certain features of the training, such as follow-up support, and achieve sufficient teacher coverage to improve teachers' practices on an appropriate scale, effective scaling considers the processes needed to scale up the program, policy, or innovation (Christina and Vinogradova 2017, for example). Therefore, it is important to understand what works and what pitfalls could be involved in taking a successful training program to scale.

To assess the World Bank's experience, this evaluation developed a theory of change (Figure 3.3), derived from the literature and case studies, that is applicable to the education sector. The theory of change considers three main stages in scaling: planning, implementation, and monitoring and evaluation, with related conditions for each stage. The evidence for the findings in each of these areas comes from six purposively chosen case studies conducted in Ghana, Uruguay, and Vietnam (see appendix A for more on the methodology and appendixes E and F for the literature and detailed findings from the cases).

Figure 3.3. Theory of Change for Scaling Up

Planning for scale up	Implementation of scale up	M&E	Outcomes
Relevant stakeholders consulted on scale up	Regular, ongoing communication with key stakeholders	M&E systematically deployed to support its viability	Vertical scale up
Scale up plan developed:	Political support	M&E used to provide a feedback loop for adaptation	Horizontal scale up
Clear sequencing	Influential champions	Training quality monitored	
Financial planning beyond initial (project) period	Sufficient financing to support the intended scale up	Anticipated number/level of beneficiary related outputs achieved	
Plan includes M&E	Financing likely to support sustainability	Training program evaluated	
	Adequate logistical support and human resources		

ASSUMPTIONS
Identified need, clarity regarding nature of scale up (how much, for how long, and in what direction), continuity of support (political, financial), long-term commitment

Source: Independent Evaluation Group.
Note: M&E = monitoring and evaluation.

Effective scaling needs to support depth of change. The literature suggests that it is important to plan for scaling and know from the start what type of scaling is intended so that the features noted in the theory of change can be integrated into each stage. Scaling can be horizontal, vertical, or functional (box 3.8); scaling has been horizontal in most World Bank operations. The conditions for scaling up might be required to differing degrees depending on the type of scaling desired. Less complex forms of scaling up that focus on enlargement or increased numbers without seeking to impact systems (such as horizontal scaling) would typically require fewer conditions. Some of the necessary conditions for scaling up relate to the sustainability of the scaling, a critical feature of this process. As described below, typically, a combination of horizontal and vertical scaling could be more effective and sustainable than horizontal scaling. This combination ensures the depth of change necessary to support and sustain lasting educational improvement (Fullan and Quinn 2016, for example).

Box 3.8. Types of Scaling

The literature identifies three types of scaling, often pursued in parallel: horizontal scaling that focuses on an intervention's breadth of coverage, vertical scaling that involves a deeper embedding of the scaling-up process within the policy making and implementation system, and functional scaling that pertains to the expansion of the type of activities or areas of engagement (for example, expanding the range or level of subject matters offered in existing training or including functional aspects of the education system).

Source: Robinson, Winthrop, and McGivney 2016.

Horizontal scaling that increases the breadth of training coverage without ensuring depth and sustainability of the training engagement is less likely to achieve long-term changes in teaching practices. Scaling up in the six World Bank cases that attempted it tended to have greater success with targeted gains in the reach of in-service training than it did with ensuring that the programs became embedded in the education system.[10] Project evaluations (Implementation Completion and Results Reports) in all six cases document project success in meeting targets for teachers trained (Box 3.9), but the government has sustained only one of the programs.

Box 3.9. Target Achievements for Teacher Training

Under the Vietnam School Readiness Promotion Project, 98 percent of school managers and 93 percent of early childhood teachers completed core modules. The in-service training under the Vietnam Escuela Nueva Project trained 52,792 teachers against a target of 30,000. Targets were exceeded for in-service training delivered in Ghana and Uruguay. The investment in in-service training in Uruguay resulted in the creation of new institutions to support ongoing training.

Regarding conditions for planning scale-up, the in-service training operations that the World Bank supported included plans to scale up in-service training that were well sequenced in all cases. The experience of a pilot project supported the VNEN, and that experience informed the sequencing of the scaling up using an adapted cascade model that ultimately led to a greater number of teachers training than originally anticipated. In Uruguay, each training cycle has a carefully managed two-year duration—the targeted schools receive training in math and sciences during the first year and training in language and social sciences in the second year.

In most cases, however, a plan extending beyond the life of the project was lacking, particularly regarding funding. Positive drivers associated with planning for scale—such as a scale-up plan, capacity building, and systemwide policies—support ongoing depth and scale. At least part of the reason for the lack of long-term planning in the case studies could be associated with the short-term, project-based funding model. However,

the in-service training in Uruguay has broad support, institutional capacity development, and government financing, which has sustained the training program. In Ghana, in-service training is also project dependent, an approach that continues to support ad hoc planning. The project cycle for the Global Partnership for Education–funded VNEN was just three years, after which no further funding was available (Box 3.10). To achieve sustainable scaling, a more programmatic approach to in-service training is indicated.

Box 3.10. Success without Sustainability

The Vietnam Escuela Nueva Project met its targets within its short, three-year time frame and acted as a catalyst for promoting the Escuela Nueva model to a sizable number of voluntary schools adopting the new approach but not supported by the project. This significant success came from careful planning that encompassed, for example, targeting, the operation of a modified cascade approach to training, and production of textbooks and other supporting materials. Additionally, a carefully designed plan for scaling up was developed after the initial pilot that ring-fenced support for 1,447 schools supported by a clear plan regarding identification, rollout, and associated supports. The approach also anticipated voluntary take up by nonproject schools, and this materialized. However, the promoters of the new approach did not plan for vertical scaling and the model's longer-term sustainability, perhaps because of the understandable concentration on completing the project within its three-year project window.

The literature and the case studies emphasize the importance of consultation and communication for the realization of effective scaling up, but consultation is not an ingrained practice in some countries and contexts. Sustaining an innovation in training beyond a pilot phase requires capacity among key stakeholders and trust in the system, which consultation can build. All the case studies show a significant investment of time and effort in close consultation and communication, but teachers were not consulted in all cases. Differing approaches to the inclusion of teachers in the consultation process highlight the influence of context and culture (Box 3.11). Therefore, the country context can present World Bank teams with a design and implementation challenge, especially where consultation of direct beneficiaries is less evident or developed, which suggests the need for sensitivity to context.

Box 3.11. Consultation During Scaling

Consultation is embedded in how things are done in Uruguay, a country characterized by secularism, liberal social laws, and a well-developed education system. School principals have significant autonomy over school management, teachers have significant autonomy over classroom management, and in-service training is demand driven. In Ghana, the approach is similarly inclusive, using feedback to make minor changes to the Untrained Teachers Diploma in Basic Education program. By contrast, Vietnam operates with a more centrally managed model of public administration in which there is significant consultation among actors within the state and Peoples Committee but limited consultation beyond.

Scaling up was successfully executed in all case study countries, within the boundaries of the objectives set for horizontal scaling of in-service training. In alignment with the theory of change, this reflects elements of capacity (such as administrative capacity, logistics management, procurement, and human resources) that might not be as evident in less-developed contexts. The Untrained Teachers Diploma in Basic Education training in Ghana was supported by an explicit plan for scaling premised on implementation by teacher colleges, improved certification requirements, and data on the numbers of unqualified teachers. The scale-up was effectively costed, adequately resourced, and complemented by human resource support from teacher colleges. Although the in-service scaling efforts did not encounter significant implementation challenges in the cases studies, challenges can arise in horizontal scaling, particularly in countries with less implementation capacity. The scaling up can progress relatively smoothly where the enabling environment and political support is robust, but where the baseline is low, it could be necessary to allow for a longer period and more technical support.

Although the nominal conditions required for implementing both horizontal and vertical scaling might be similar, the level of intensity required for vertical scaling is typically more onerous. The VNEN, for example, aspired to more complex and longer-term scale-up but lost its key champions, who reached retirement age when the project closed. Together with other factors, this had a profound effect on the model's sustainability. The pursuit of horizontal scaling was simply defined in most instances—objectives and indicators were largely output-oriented—and in-service training was generally executed within a single project cycle. Most of the scale-up efforts explored in the case studies did not face challenges as acutely as those associated with longer-term, system-focused scaling efforts (vertical scaling) did as identified in the literature.[11] For example, challenges arose from the time required for scaling to come to fruition and the associated misalignment with project cycles, which are complicated by fluctuating donor priorities. To address those challenges, elements like political support and influential champions become important. Weak monitoring and evaluation undermines the potential for sustainable approaches to in-service training. Systematic evidence—

including robust monitoring in addition to evaluation that is used to adapt to implementation challenges—is critical to sustainability, but the quality of the evidence in the case studies to support sustained scale-up was weak. The Untrained Teachers Diploma in Basic Education and VNEN benefited from impact evaluations. For the latter, where no further scaling was required, the impact evaluation reflected positive appreciation of good practice. In VNEN, the impact evaluation was an important part of making the case for sustained support. However, the late start and associated late production of the evaluation lessened its value and utility.

4. Conclusion

This assessment of World Bank support for the professional development of teachers through preservice and in-service training considered the main drivers of education quality and their application in World Bank operations. It turns now to ways in which future operations could be designed, implemented, and scaled to improve effectiveness and help ensure results.

The World Bank has taken steps in recent years to enhance its attention to teachers and their training with a variety of operations and with initiatives such as the Human Capital Project and Global Platform for Successful Teachers. Between FY13 and FY18, the World Bank supported teacher training in 110 projects in 67 countries. Those projects, like education projects in general, have been implemented most often in low- and lower-middle-income countries, where the need is greatest, and the challenges are numerous. The scope of the issues involved has contributed to a selective approach to teacher training. For example, partly because of the political economy challenges of working on preservice training, most of the projects reviewed focused on in-service training: 68 projects exclusively supported in-service training, two supported only preservice, and the rest supported both.

Attention to Preservice Training

For several reasons, World Bank experience with preservice training has been limited and has supported only selective elements. The overall quality of initial teacher development in many low- and lower-middle-income country clients is low; however, because of the political economy challenges of working with diverse stakeholders who have varying goals, the World Bank has tended to intervene only where the government understands the need to reform preservice training and uses in-service training to address deficiencies. Improving teacher quality will require improvements in the candidates graduating from preservice training institutions, which in-service training alone cannot address. This provides a strong rationale for the World Bank to use its support for tertiary education to improve teacher formation in preservice institutions and strategically addressing specific weaknesses in preservice training. Context-specific research may be needed to assess the best way to sequence the quality drivers, given the existing level of quality in the system. TTLs believed it would be most helpful to understand the institutional landscape and the individual institutional strengths and weaknesses fully, and to leverage the World Bank's tertiary education support for preservice training.

When the World Bank addressed screening and filtering, it focused on ensuring a steady supply of motivated students to enter the teaching profession. TTLs say that attracting

better candidates is often inhibited by the unattractiveness of teaching and overall low capacity of students who enter higher education, areas that could get more attention in World Bank operations. A starting point might involve increasing dialogue with clients to plan for longer-term measures to improve the quality of candidates. Some operations focused on creating incentives to motivate students through support for scholarships and stipends during training, and others strengthened an existing examination function to enable better screening of those exiting preservice. Some operations adopted alternative approaches, more often using existing mechanisms to address scarcity than creating a new mechanism to bring in candidates with stronger content knowledge.

World Bank operations have addressed teacher coursework with attention to enhancing the curriculum, building the capacity of teacher trainers, and improving infrastructure related to teacher preparation. The literature and TTLs emphasize the importance of aligning the preservice curriculum and methods preparation with the actual curriculum in schools. This contrasts with many preservice training classrooms, which rely on conveying overly theoretical information and ineffective teaching methods such as rote memorization and copying. This is one reason half of the PADs examined provide some form of capacity development for teacher trainers. Project documents frequently propose to develop teacher coursework that is balanced between academic content and pedagogical skills. This work supporting locally relevant curriculums and material needs to continue.

The teacher practicum, although critical to teacher preparation, received little attention in World Bank operations. Only 15 of the operations examined included support mechanisms related to practicums. In some instances, the support was extensive, linking with multiple aspects of preservice training and incorporating key principles. TTLs are broadly aware of the need to prepare candidates for the actual environments in which they will eventually teach through better-supported practicums and alignment with the coursework. It is not only a matter of increasing the duration of the practicum but of ensuring the practicum is designed with supported opportunities for practice, reflection, and feedback.

Quality assurance related to preservice education begins with establishing education standards, teacher training curriculums, teacher educator requirements, practicum requirements, and other system-level aspects such as accreditation. The accreditation, monitoring, and support of teacher training institutions are important to ensuring the quality of training. TTLs viewed accreditation as important, but its effectiveness in ensuring quality among preservice training institutions was not always evident. However, the accreditation process is not just about holding preservice training

institutes accountable to a standard of quality but also about supporting institutions to develop to the standard—both opportunities for World Bank support.

Broaden Interventions in In-Service Training

The World Bank operations examined had features necessary for effective in-service training, though often not in combination. Supported training programs meet the minimum duration and impart a broad range of skills, and some systems consult widely with stakeholders, including teachers. The provision of follow-up support and adaptation to teacher capacity to move them toward being more reflective professionals are not adequately emphasized considering the prevailing context in most World Bank clients. Adults learn through application, modeling, and demonstration during training—that is, engaging teachers directly in designing and trying out teaching strategies rather than focusing solely on lecture and discussion.

Effective programs make the training relevant to adult learning by having teachers analyze students' work or watch videotaped lessons, critiquing, and then trying the strategy. Fieldwork and TTL reports suggest that active learning is a feature of some, but not all, training programs, suggesting an area in which the World Bank might expand its activities. TTLs agree that trainers might not fully understand this concept or use it adequately, suggesting an area for improvement. Except for one project, training programs are of adequate duration—at least 60 hours spread over time. These training programs broadly focus on pedagogy rather than being discipline-specific, which is striking for teachers at the secondary level.

The programs examined respond to teachers' general needs and are geared to teacher capacity in nearly 40 percent of the operations. Programs geared to teacher capacity sometimes focus on untrained and undertrained teachers, which is an important group to target. However, training programs need to build competencies and move teachers from mechanical to reflective professionals. The training program in Uruguay is an example of better practice: It flexibly addresses the practice of teaching through school-based training. Trainers customize their support according to the school environment or teachers' challenges.

Fieldwork also found that training programs typically are not differentiated by student grade, age, or ability level, except for Uruguay's training, which helps teachers identify struggling students and provide them with more tailored support. This kind of tailoring of training, if done more consistently in World Bank operations, would help teachers foster a better understanding of students' levels, which relates to another important success factor: matching teaching to students' learning levels.

Effective training is typically followed up with opportunities for reflection and time to apply, practice, and test new learning, which needs more attention in World Bank projects. Just 54 of 108 operations specify such follow-up support, particularly in more recent operations. School-based training and distance training are more likely to include a design for follow-up support, whereas those that use workshops and a cascade approach seldom offer formal follow-up. Variability has been observed in the quality of coaches, and it is a challenge the World Bank will need to address more fully. This highlights the need for greater accountability, monitoring of follow-up support, and capacity building of coaches.

Effective in-service training requires instructional leadership from head teachers, school principals or managers, and school inspectors. Less than half of the operations examined support instructional leadership simultaneously. Principals need to build the school culture to focus on learning and collegiality, aspects that are important enabling conditions for training and peer learning. Thus, even if teachers have the intention of implementing training, instructional leaders must also support the practices.

Training programs need to be consistently implemented within a broader framework for teacher development. Opportunities for professional development are needed throughout the teaching career, likely accompanied by filtering or screening mechanisms, as suggested by the analysis of high-performing countries. In these countries, the screening and support mechanisms diagnose weaknesses and provide sustained support. TTLs believe that an important way to improve training is to anchor it within teacher development frameworks that link training with teachers' career opportunities.

Monitoring and evaluation are limited in training programs. Less than half of the operations examined evaluate in-service training. Classroom observations and teacher testing are the most common methods used to measure performance changes resulting from in-service training. Few impact evaluations are planned, and monitoring could have a role to ensure the fidelity and implementation of the training and follow-up support. Without such data, it is not possible to identify noncompliance or lack of consistency and bring greater oversight and learning. TTLs agree that monitoring needs more effort, including greater clarity on the outcome the training is expected to achieve. Monitoring and evaluation are also essential to ensuring the sustainability of scaling up in-service training.

Scaling up has achieved some success in the six case study countries, but the gains are largely measured in outputs rather than lasting improvements in teachers' professional development or program sustainability. The scaling has been well implemented, though this was in countries with good implementation capacity. However, the efforts have

been largely in increasing the number of teachers trained and were limited to delivery within a single project cycle. Notwithstanding the implementation success based on targets met efficiently and on time, support for scaling up in-service training lacks a longer-term strategic focus, particularly regarding ensured funding to sustain any improvements made. Some of the conditions that need to be addressed more systematically include planning for longer-term funding support, consultation and communication with beneficiaries and other key stakeholders, and ongoing political support.

Lessons

The World Bank's limited focus on preservice training institutions may not ensure effective teaching. Instead, more active policy dialogue may be needed to convince clients to reform. A more contextualized assessment of preservice institutions that highlights individual institutional strengths and weaknesses, for example, may help overcome political economy constraints. Including such assessments in policy dialogue would provide government clients with the evidence they require to understand the issues affecting the quality of preservice training and move toward reform. Dialogue may also facilitate development of a long-term plan to sequence improvements to the quality drivers. The shortcomings of graduating candidates have repercussions for in-service training programs, which alone cannot improve these candidates. Thus, the rationale for attending to both types of training systems is clear.

The effectiveness of in-service training programs depends on consistent attention to all key features. The degree to which key features are integrated with the education system matters. Ensuring integration may require the World Bank to consider comprehensive in-service training reforms and embed them in the education system. Effective in-service training programs alone are not enough to give teachers a broad repertoire of skills or make them more reflective practitioners. For this reason, sustained follow-up is critical and requires mechanisms that provide opportunities for peer learning and coaching, as well as participatory approaches that elicit teachers' views about their training and needs. These mechanisms require effective instructional leaders who can provide follow-up support that differentiates teachers' capacity and helps teachers address varying learning levels among students. Additionally, training programs need to be anchored in teaching standards, career ladder progression, and screening throughout the career. Embedding key in-service features in the education system and in the design and implementation of operations (including monitoring and evaluation) is also needed.

Sustainable scaling up of in-service training requires attention to key conditions for the planning, implementation, and monitoring of the scaling. In-service scaling was well planned and well implemented in the cases examined. In the training programs visited, scaling-up plans covered logistics, costs, and modalities; targets were set and met for numbers trained, and materials were distributed as relevant. Yet sustainability remains an issue. In this regard, World Bank operations may benefit from greater focus on quality assurance for in-service training, arrangements to evaluate training outcomes, and planning to embed system-related aspects of in-service training into the scaling. Some of the gaps in planning and design were associated with short implementation periods for operations. Thus, TTLs may need to plan for scale-up from the start and address constraints such as political support, long-term financing, and monitoring and evaluation. Efforts to assess quality and outcome can help build a case for more sustained in-service provision.

Chapter 1

[1] The regional distribution of training projects is similar to the regional distribution of all education project approved in this period.

[2] These are countries with gross national income per capita below $4,000

[3] Figure 1.1 is based on a single country per group (in-service, preservice, or both) and no double counting of multiple projects. The population of preservice countries fits into the in-service population with just one exception, Cameroon, a country that has a preservice intervention and no in-service interventions. Thus, there is no difference in the Human Capital Index makeup of countries between pre- and in-service country groups. The data correspond to summaries by income level (see appendix B).

[4] The Service Delivery Indicators initiative benchmarks service delivery performance in education and health in Africa.

[1] The heterogeneity of the institutions charged with quality assurance and accreditation makes it hard to test hypotheses about effective approaches (Ingvarson and others 2013).

[2] For all three literature reviews, search terms were entered into databases such as World Bank Library databases, Web of Science, Google Scholar, and Google (see appendix A for more on search terms). For the first two reviews (education quality and preservice), the gray literature from the following organizations was also searched: 3iE, Campbell Collaboration, Inter-American Development Bank, the National Bureau of Economic Research, Organisation for Economic Co-operation and Development, the U.K. Department for International Development, the United Nations Children's Fund, the United Nations Educational, Scientific, and Cultural Organization, the U.S. Agency for International Development, and the World Bank. The gray literature

searched for the third review, scaling up, was from the following organizations: World Bank, the United Nations Children's Fund, Organisation for Economic Co-operation and Development, the U.K. Department for International Development, Inter-American Development Bank, the Brookings Institution, the Center for Global Development, the National Bureau of Economic Research, the U.S. Agency for International Development, and the United Nations Educational, Scientific, and Cultural Organization. The literature review for scaling up was further supplemented by a Google search for scaling up of public policy, programs, and projects to identify generally applicable lessons or issues, models, and frameworks.

[3] Case studies are based on multiple sources of evidence, including stakeholder interviews, documents and reports, observation, and task team leader interviews.

[4] Operational manuals were also reviewed in a subset of in-service and preservice training operations, and these documents did not provide any more details than the project appraisal documents did on cost and features of training.

Chapter 2

[1] Wang and others (2003) considers teacher preparation only in Australia, England, Hong Kong, Japan, Korea, the Netherlands, Singapore, and the United States.

[2] However, there are several caveats regarding the connection between teacher education and student outcomes. Teacher education is not a consistently significant predictor of outcomes like student achievement. It tends to be moderate in size, even when significant, and explains little of the effect individual teachers have on students.

[3] Methods-related teacher preparation should be embedded in the curriculum that teachers will be using in their actual work and the contextual reality of their future classrooms and communities (Grossman, Hammerness, and McDonald 2009).

[4] In the United States, some research shows a positive effect of the Teach for America program that recruits high-performing college graduates to work on a short-term contract basis (Béteille and Evans 2019).

[5] The World Bank supported an impact evaluation of the program in Chile.

[6] A complicating factor in any review of preservice teacher preparation is the differing requirements for each teaching level. Primary school teachers are usually generalists who need mastery of basic content across multiple subjects. In theory, this means more time should be available for training in methods; however, the evidence cited suggests that in many poor countries, even basic levels of content knowledge cannot be assumed. The sizable content demands for secondary education specialist teachers often require additional course work that can come at the expense of methods-related courses. Pedagogical and content knowledge was geared for primary and preprimary teachers, but few operations specifically discussed preparation for secondary-level teachers (World Bank 2016; Wane and Martin 2013; Martin and Pimhidzai 2013).

[7] The descriptive comparisons show that not all high-scoring countries or economies have longer programs. For example, Taiwan, China's training programs are all 50–60 months long (4 or more years), but a sizable portion in Singapore is trained in 18 to 24-month programs.

[8] Duration alone is not a strong predictor of systemic performance. Trainees in Taiwan, China spend more than 1,400 hours (175 days) in extended teaching practice, but Thailand and the Philippines also have some of the longest practice periods among Teacher Education and Development Study in Mathematics countries.

[9] Taiwan, China is an example of a strong system with several important features: a clear legislatively defined accreditation function, a single agency with national power, program evaluation conducted by professional experts (from universities, among others), a collection of firsthand evidence from a variety of sources, and termination (or disaccreditation power for the agency.

[10] Teachers in Angola, for example, have typically completed only eight years of schooling. In this case and others like it in Brazil, Ghana, Liberia, and Tanzania, the program was conducted over a year (or more), indicating that the World Bank recognizes the need for longer duration in such cases.

Chapter 3

[1] Project appraisal documents (PADs) did not discuss some key features, such as active and applied learning, so other sources of evidence were also examined.

[2] Although the amount of time considered adequate is not defined, programs that operated over consecutive days had a positive impact (Popova and others 2018). Programs in the United States with a minimum of 50 to 60 hours spread over time produced changes in teachers' pedagogical practices (Garet and others 2010). Other authors similarly report this figure (Carrillo, van den Brink, and Groot 2016). For example, the Middle School Math Professional Development Program trained teachers during a three-day summer institute and then held a series of one-day follow-up seminars over the course of the school year (Garet and others 2010). An in-school coaching session was provided after each seminar.

[3] Effective programs make the training relevant to adult learning style by having teachers analyze students' work or watching videotaped lessons, critiquing, and then trying out the strategy (Garet and others 2001; Darling-Hammond, Hyler, and Gardner 2017; Popova and others 2018).

[4] The basis for training was clear in less than half of World Bank PADs (approximately 40 percent), but about half (56 percent) did not specify a basis of assessment because training related to a general need "to improve instruction." Needs assessment was the most common type of information (17 PADs). Others were classroom observation (13 PADs), student achievement data (12 PADs), or study (12 PADs).

[5] Cilliers and others (2018) finds that pupils exposed to two years of the program improved their reading proficiency by 0.12 standard deviations.

[6] For example, one researched training program allocated time during the seminar for teachers to plan lessons to apply the new material in their classroom instruction. Coaches visited immediately after one of the seminars and provided small group or individual support (Garet and others 2010).

[7] There is promising evidence that this approach can succeed at shifting teaching practice and improving student learning. For example, see Cilliers and others (2018) and Bruns, Costa, and Cunha (2017).

[8] The need for sustained training is consistent with practice (Reimers and Chung 2018; Darling-Hammond, Hyler, and Gardner 2017).

[9] Impact evaluations are planned in various countries: The Arab Republic of Egypt; Cambodia, Chad, the Dominican Republic, Ethiopia, Ghana, the Lao People's Democratic Republic, the Republic of Congo, the Republic of Yemen, São Tomé and Príncipe, St. Vincent and the Grenadines, Tanzania, Uruguay, and Vietnam.

[10] Functional scale-up (expanding specific activities or areas of engagement), also mentioned in the literature, was not found in the cases examined.

[11] The literature includes, for example, Hartmann and Linn (2007), Banerji and Madhav (2016), Fleisch (2016), Rincon-Gallardo (2016), Colbert and Arboleda (2016), Gilson and Schneider (2010), Yew and others (2014), and, Spicer and others (2014).

Bibliography

Akyeampong, K. 2017. "Teacher Educators' Practice and Vision of Good Teaching in Teacher Education Reform Context in Ghana." *Educational Researcher* 46 (4): 194–203.

AMTE (Association of Mathematics Teacher Educators). 2017. *Standards for Mathematics Teacher Preparation*. Raleigh, NC: AMTE.

Aslam, Monazza, Shenila Rawal, Geeta Kingdon, Bob Moon, Rukmini Banerji, Sushmita Das, Manjistha Banerji, and Shailendra K. Sharma. 2014. *Reforms to Increase Teacher Effectiveness in Developing Countries: Systematic Review.* London: Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, University College London Institute of Education, University College London.

Bentancur, L., and B. Gabbiani. 2016. *Enseñanza de lectura y escritura de maestros en escuelas de Tiempo Completo participantes en el TERCE*: Segundo Informe de Resultados, Instituto Nacional de Evaluación Educativa, Quito, Ecuador.

Béteille, Tara, and David K. Evans. 2019. *Successful Teachers, Successful Students: Recruiting and Supporting Society's Most Crucial Profession.* Washington, DC: World Bank.

Béteille, Tara, Michelle Riboud, Shin Nomura, Namrata Tognatta, and Yashodhan Ghorpade. 2018. *Ready to Learn. Ready to Thrive: Before School, In School, and Beyond School in South Asia*. Washington, DC: World Bank.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane. 2017. "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa." *Journal of Economic Perspectives* 31 (4): 185–204.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Teacher Preparation and Student Achievement." National Bureau of Economic Research (NBER) Working Paper 14314, NBER, Cambridge, MA.

Bruns, Barbara, and Javier Luque. 2014 *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean.* Washington, DC: World Bank.

Bruns, Barbara, Leandro Costa, and Nina Cunha. 2017. "Through the Looking Glass: Can Classroom Observation and Coaching Improve Teacher Performance in Brazil?" Policy Research Working Paper 8156, World Bank, Washington, DC.

Bush, Tony, and Derek Glover. 2016. "School Leadership in West Africa: Findings from a Systematic Literature Review." *Africa Education Review* 13 (3-4) 80–103.

Carrillo, Camilo, Henriëtte Maassen van den Brink, and Wim Groot 2016. "Professional Development Programs and Their Effects on Student Achievement: A Systematic Review of the Evidence." Top Institute for Evidence Based Education Research Working Paper 16/03, Maastricht University, Netherlands.

Chetty, R., J. N. Friedman, and J. E. Rockoff. 2014. "Measuring the Impact as of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 259–632.

Christina, Rachel, and Elena Vinogradova. 2017. "Differentiation of Effect Across Systemic Literacy Programs in Rwanda, the Philippines, and Senegal." *New Direction of Child Adolescent Development* March (155): 51–65.

Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor. 2018. "How to Improve Teaching Practice? Experimental Comparison of Centralized Training and In-Classroom Coaching." Research on Improving Systems of Education Working Paper 15/2018, Stellenbosch University, Stellenbosch, South Africa.

Colbert, Vicky, and Jairo Arboleda. 2016. "Bringing a Student-Centered Participatory Pedagogy to Scale in Colombia," *Journal of Educational Change* 17 (4): 385–410.

Conn, Katharine M. 2014. "Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Rigorous Impact Evaluations. Colombia: Academics Commons." PhD thesis, Columbia University.

Darling-Hammond, Linda, and Nikole Richardson. 2009. "Research Review/Teacher Learning: What Matters?" *How Teachers Learn* 66 (5): 46–53.

Darling-Hammond, Wei, Ruth Chung; Aleathea Andree, C. W. Ruth, and N. Richardson. 2009. *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad.* Washington, DC: National Staff Development Council.

Darling-Hammond, Linda, Maria E. Hyler, and Madelyn Gardner. 2017. *Effective Teacher Professional Development.* Palo Alto, California: Learning Policy Institute.

Evans, David K., and Anna Popova. 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." The World Bank Research Observer 31 (2): 242–270.

Fullan, Michael, and Joanne Quinn. 2016. *Coherence: The Right Drivers in Action for Schools, Districts, and Systems.* Ontario, Canada: Corwin.

Ganimian, Alejandro J., and Richard J. Murnane. 2016. "Improving Education in Developing Countries: Lessons from Rigorous Impact Evaluations." *Review of Educational Research* 86 (3): 719–755.

Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Kirk Walters, Mengli Song, Seth Brown, Steven Hurlburt, Pei Zhu, Susan Sepanik, and Fred Doolittle. 2010. *Middle School Mathematics Professional Development Impact Study: Findings after the First Year of Implementation* (NCEE 2010-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bibliography

Garet, Michael S., Andrew C. Porter, Laura Desimone, Beatrice F. Birman, and Kwang Suk Yoon. 2001. "What Makes Professional Development Effective? Results from a National Sample of Teachers." *American Educational Research Journal* 38 (4): 915–945.

Glazerman, Steven, Daniel Mayer, and Paul Decker. 2006. "Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes." *Journal of Policy Analysis and Management* 25 (1): 75-96.

Goldhaber, Dan 2019. "Evidence-Based Teacher Preparation: Policy Context and What We Know." *Journal of Teacher Education* 70 (2): 90–101.

Grossman, Pam, Karen Hammerness, and Morva McDonald. 2009. "Redefining Teaching, Re-imagining Teacher Education." *Teachers and Teaching: Theory and Practice* 15 (2): 273–289.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267–71.

Hartmann, Arntraud, and Johannes F. Linn. 2007. "Scaling Up: A Path to Effective Development." 2020 Focus Brief on the World's Poor and Hungry People, International Food Policy Institute, Washington, DC.

Ingvarson, Lawrence, John Schwille, Maria Teresa Tatto, Glen Rowley, Ray Peck, and Sharon L. Senk. 2013. *An Analysis of Teacher Education Context, Structure, and Quality Assurance in TEDS-M Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M).* Amsterdam: International Association for the Evaluation of Educational Achievement.

Lewin, Keith M. 2004. "The Pre-Service Training of Teachers: Does It Meet Its Objectives and How Can it Be Improved?" Background Paper commissioned for the Education for All Global Monitoring Report 2004, The Quality Imperative.

Luneta, K. 2012. "Designing Continuous Professional Development Programmes for Teachers: A Literature Review." *Africa Education Review* 9 (2): 360–379.

Martin, Gayle H., and Obert Pimhidzai. 2013. *Service Delivery Indicators: Kenya*. Washington, DC: World Bank.

Orr, David, Jo Westbrook, John Pryor, Naureen Durrani, Judy Sebba, and Christine Adu-Yeboah. 2013. *What Are the Impacts and Cost-Effectiveness of Strategies to Improve Performance of Untrained and Under-Trained Teachers in the Classroom in Developing Countries?* London: EPPI Centre, Social Science Research Centre, Institute of Education, University of London.

Parandekar, Suhas D., Futoshi Yamauchi, Andrew B. Ragatz, Elisabeth K. Sedmik, and Akiko Sawamoto. 2017. *Enhancing School Quality in Vietnam through Participative and Collaborative Learning: Vietnam Escuela Nueva Impact Evaluation Study*. Washington, DC: World Bank.

Popova, Anna, David K. Evans, Mary E. Breeding, and Violeta Arancibia. 2018. "Teacher Professional Development around the World: The Gap between Evidence and Practice." Policy Research Working Paper 8572, World Bank, Washington, DC.

Reimers, Fernando M., and Connie K. Chung, eds. 2018. *Preparing Teachers to Educate Whole Students: An International Comparative Study.* Cambridge, MA: Harvard Education Press.

Robinson, Jenny Perlman, Rebecca Winthrop, and Eileen McGivney. 2016. *Millions Learning: Scaling Up Quality Education in Developing Countries*. Washington, DC: Brookings Institution.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52.

Sorto, M. A., and T. F. Luschei. 2010. "Teachers' Education, Supervision, and Evaluation in Costa Rica." In *International Handbook of Teacher Education World-Wide: Issues and Challenges, Volume II*, edited by K. G. Karras and C. C. Wolhuter, 727–746. Athens: Atrapos Editions.

Tatto, Maria Teresa. 2013. *The Teacher Education And Development Study In Mathematics (Teds-M): Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries.* Amsterdam: International Association for the Evaluation of Educational Achievement.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2012. *Antecedentes y criterios para la elaboración de políticas docentes en América Latina y el Caribe*. Paris: UNESCO.

Villegas-Reimers, E. 2003. *Teacher Professional Development: An International Review of Literature.* Paris: International Institute for Educational Planning.

Wane, Waly, and Gayle H. Martin. 2013. *Education and Health Services in Uganda: Data for Results and Accountability*. Washington, DC: World Bank.

Wang, Aubrey H., Ashaki B. Coleman, Richard J. Coley, and Richard P. Phelps. 2003. *Preparing Teachers around the World*. Princeton, NJ: Educational Testing Service.

Westbrook, Jo, Naureen Durrani, Rhona Brown, David Orr, John Pryor, Janet Boddy, and Francesca Salvi. 2013. *Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries. Education Rigorous Literature Review.* London: U.K. Department for International Development.

World Bank. 2016. "Pakistan—Third Panjab Education Sector Project." Project Appraisal Document PAD1641, World Bank, Washington, DC.

———. 2018. *World Development Report 2018: Learning to Realize Education's Promise.* Washington, DC: World Bank.

———. 2019. *Teach.* Washington, DC: World Bank.

Appendix A. Methodological Approach

Evaluation Questions

The evaluation's objective is to understand how the World Bank supports the design and implementation of preservice and in-service teacher training and the scaling up of in-service training.

The objective and associated concerns prompted development of three evaluation questions that guided the collection and analysis of data and the framing of findings and lessons:

- What are the key features of effective preservice training programs, and to what extent do World Bank operations reflect these characteristics?

- What are the key features of effective in-service training programs, and to what extent do World Bank operations reflect these characteristics?

- What factors determine the effective scaling up of effective teacher in-service training financed by the World Bank?

Evaluation Approach and Design

The evaluation uses a mixed methods approach—structured literature review, background papers, secondary data analysis, portfolio analysis, interviews, and case studies—to support data collection and analysis. Data were systematically triangulated to ensure the findings' robustness.

For the first two evaluation questions, the design sought to triangulate key findings from literature reviews, portfolio analysis, and interviews with task team leaders (TTLs) and key stakeholders. These data were triangulated with secondary data analysis for preservice training and case studies for in-service training. Links between preservice and in-service training in World Bank operations and other interventions were examined. In addition, views elicited through interviews with TTLs were used to identify how World Bank operations can be designed better to address key constraints to maximizing effectiveness. The third evaluation question, on scaling up in-service training, was supported by a literature review and case studies. A theory of change for scaling up in-service training was developed from the literature and findings from case studies.

Evaluation Components

Table A.1 lists the evaluation components applied for each question. The text that follows elaborates on the content and function of each component and the selection and analysis process.

Table A.1. Application of Evaluation Components

Data Collection and Analysis Methods	Preservice Training	In-Service Training	Scaling Up of Training
Structured literature review	Preservice literature	Education quality and training	Scaling up
Secondary data analysis	Yes	No	No
Portfolio review analysis	Yes	Yes	No
Interviews with task team leaders and key stakeholders	Yes	Yes	Yes
Case studies	No	Yes	Yes

Literature Review

Selection and process. Three structured literature reviews were conducted on education quality (systematic reviews of interventions that improve education outcomes supplemented with studies examining the impact of training), preservice training, and scaling up. In addition, supplemental searching of key authors was performed to identify relevant studies. The reviews provided the theoretical foundation for the evaluation and the basis for developing theories of change and the coding template for the portfolio of operations reviewed.

For the first search (education quality), search terms such as "education quality," "teacher professional development," and "teacher in-service training" were keyed into the sources referenced below and combined with "review" or "meta-analysis" under a joint search: ("education quality" OR "teacher professional development" OR "teacher in-service training") AND (review); ("education quality" OR "teacher professional development" OR "teacher in-service training") AND ("meta-analysis").

For the second search (preservice teacher training), terms such as "teacher preservice," "teacher preservice training," and "preservice teacher training" were keyed into the sources referenced below and combined with "review" and "analysis" and "meta-analysis": ("teacher preservice" OR "teacher preservice training" OR "preservice teacher training") AND (review); ("education quality" OR "teacher professional development" OR "teacher in-service training") AND ("analysis").

The third literature review (scaling up) included search terms such as "scaling-up," "teacher professional development," and "teacher preservice training," combined with "review" or "analysis" for a joint search: ("scaling-up" AND "teacher professional development" OR "teacher preservice training" OR "preservice teacher training") AND (review); ("scaling-up" AND "education quality" OR "teacher professional development" OR "teacher in-service training") AND ("analysis"); ("scaling-up" AND "public policy" OR "public program" OR "project"); ("scaling-up" AND "public policy" OR "public program" OR "project") AND ("education").

The databases used for the literature reviews consisted of World Bank Library databases (including education and economics) and Web of Science, and Google Scholar and Google were used to capture additional papers. The gray literature searched for the first and second reviews was from the following organizations: Campbell Collaboration, 3iE, the World Bank, the U. S. Agency for International Development, Inter-American Development Bank, Organisation for Economic Co-operation and Development,[1] the United Nations Children's Fund, the National Bureau of Economic Research, the U.K. Department for International Development, and the United Nations Educational, Scientific, and Cultural Organization. The gray literature searched for the third review was from the following organizations: World Bank, the United Nations Children's Fund, Organisation for Economic Co-operation and Development, the U.K. Department for International Development, Inter-American Development Bank, the Brookings Institution, the Center for Global Development, the National Bureau of Economic Research, the U. S. Agency for International Development, and the United Nations Educational, Scientific, and Cultural Organization. The literature review for scaling up was further supplemented by a Google search for scaling up of public policy, programs, and projects to identify generally applicable lessons or issues, models, and frameworks.[2] This general search covered the first 100 items identified (ordered by relevance) within the period October 1, 2008, to October 1, 2018.

Review. The inclusion criteria applied to all searches were academic peer-reviewed and gray literature (quantitative and qualitative studies); survey papers and studies of specific projects; and studies of and analysis of teacher preparation for preprimary, primary, basic, and secondary levels.[3] Studies in low- and middle-income countries were searched, supplemented with studies from developing countries. Exclusion criteria applied for all three searches (except for the general search on scaling up) were publication date before 2000, and studies and analyses of technical and vocational education and training. The preservice review process also drew on teacher professional development literature.

The number of research results, articles reviewed, and articles found relevant after application of the protocol for literature reviews are presented in table A.2.

Table A.2. Results: Application of Protocol for Literature Reviews (number)

Literature Review	Search Results	Reviewed Articles	Selected Articles for Relevance
Education quality	1,559	66	21
Preservice	360	41	25
Scaling up	1,151	52	21

The literatures for preservice and in-service training were synthesized in three steps. The first step was a thorough literature review. The synthesis consisted of a summary and detailed description of each training characteristic acknowledged in each article. The second step was an in-depth review of the details of each characteristic—captured information included the description, type of activities, and contextual features to ensure effective implementation. Third, all the information was compiled into a table that provided a description of all elements acknowledged by the authors to achieve effective preservice or in-service training programs.

A background paper on preservice training was prepared as described in appendixes C and D. The background paper supported the development of key drivers and helped identify the elements that were reviewed in the portfolio.

The literature review undertaken for scaling up of in-service training became the basis for a background paper. The breadth of the literature reviewed allowed for an appreciation of common issues related to scaling up, independent of sector. The review identified, as available, aspects and characteristics of scaling up more particular to the education sector and to in-service teacher training, noting that the research undertaken found certain core characteristics of scaling up that were independent of sector. The process involved a thorough review of the literatures, identification of major concerns (such as definition, obstacles and challenges, and characteristics of successful scaling), and grouping of literatures for use in the background paper.

Both background papers supported the evaluation through the identification of insights into evaluation questions and concepts, and both were shared among team members to ensure that a common understanding informed the approach to other evaluation components. The authors of the papers also took part in team meetings and contributed to discussion that further informed case studies and interviews.

Quantitative Analysis of Secondary Data

Selection and process. The quantitative analysis focused on three data sets:

- 2008 Teacher Education and Development Study in Mathematics—data from 17 countries used to assess the relationship between features of teacher training programs and teacher capacity.

- 2006–11 Southern and Eastern Africa Consortium for Monitoring Educational Quality—data from 15 countries; in each, representative samples of grade six students completed tests in reading, mathematics, and health and HIV knowledge, as did grade six teachers.

- 2013 Latin American Laboratory for the Assessment of the Quality of Education—a regionwide assessment of student achievement (the Third Regional Comparative and Explanatory Study) in 15 countries, including representative samples of grade three and grade six students who completed standardized tests in reading and mathematics; they were linked with their teachers, who completed detailed background questionnaires (see appendix D).

Analysis. The analysis examined associations between teacher preservice training characteristics and teacher quality and/or student achievement outcomes (see appendix D).

Portfolio Review of World Bank Projects

Selection criteria and process. The Independent Evaluation Group's (IEG) identification methodology used the World Bank's sector and theme codes and relevant World Bank databases, together with a manual review, to systematically capture and categorize the relevant portfolio subsets. The portfolio identification consisted of three steps: selection of projects for review based on the application at least one relevant World Bank theme or sector code;[4] selection of projects from the World Bank Education Projects Database using relevant activity and subsector filters;[5] and review of the project appraisal documents (PADs) for the 208 projects identified through step two to confirm whether they had preservice or in-service activities. This process yielded 110 relevant projects.

Coding and analysis. IEG coded and extracted data from the selected projects using coding protocols developed through insights from literature reviews of preservice and in-service interventions. IEG reviewed every appraisal document and a sample of operational manuals (n = 18) to identify as much detail as possible on key training features, although the operational documents did not provide any more detail in that regard. The literature reviews were critical for identifying the existing body of evidence on what kinds of in-service and preservice teacher training features are most effective.

The coding protocol for preservice training captured the type of interventions used to support preservice training (infrastructure, learning materials, and the like); whether the

project supported the design and/or implementation of features to increase the attractiveness of the teaching profession; whether the project supported alternative routes to traditional preservice training; whether the project supported extensions to preservice training duration; whether the project supported the design and/or implementation of entry or exit examinations for preservice training; whether the project supported capacity development for teacher training through content knowledge or pedagogical methods; whether the project supported any type of infrastructure for preservice training; whether the project supported curriculum development in preservice institutions; whether the project supported the design and/or implementation of a practicum and the development of a national policy on practicums; whether the project supported processes or institutions for conducting accreditation of monitoring centers; and whether the project supported the establishment of a national center for teacher training.

The coding protocol for in-service training captured whether the design of the training included an ex ante assessment of needs; whether the training had a national scale; the school level toward which the training was targeted; the content focus of the training; whether the training was complemented with materials; whether there was some sort of follow-up provided as part of the training; training duration and modality; and whether the training had built-in incentives for teachers to participate.

Interviews and Consultation with Stakeholders

There was broad stakeholder engagement over the course of the evaluation. First, the evaluation's approach and scope were discussed with seven staff, managers, and directors from the Education Global Practice.

Second, interviews (individual and group) were held with TTLs and other stakeholders—including staff from ministries of education at the central and other levels and teachers—during IEG's missions to Ghana, Uruguay, and Vietnam. IEG interviewed (or held focus groups) with nearly 200 people.

Selection criteria and process. TTLs were identified for participation in focus groups and interviews from the universe of 110 projects included in the portfolio review. TTLs were selected either because the operation was the subject of the case study or the appraisal documents lacked information relating to several key training features. In addition, TTLs were invited, by email, to one of three focus groups scheduled between April 1 and 3, 2019. An additional step was taken to supplement information obtained from appraisal documents. TTLs were interviewed in May and June 2019 to identify key features not discussed in the appraisal documentation. Open-ended questions were also asked to better identify constraints and understand why the World Bank may not

support particular key features of in-service training. A sample of 40 operations (and corresponding TTLs) were selected based on cases that had gaps in data from appraisal documents. Additional information was obtained for 38 operations. Thus, the interviews with the TTLs filled in missing data not specified from appraisal documents. When TTLs shared additional training documents, these documents were also reviewed. Thirty-six TTLs participated in focus groups or interviews.

Collection and analysis. Notes from the interviews with the TTLs and stakeholders were transcribed and triangulated with the other sources of evidence (literature review and portfolio analysis).

Case Studies

Selection criteria and process. Case studies were undertaken in support of the exploration of scaling in-service training. Six cases for analysis (six efforts supported across five projects) were identified in consultation with the Education Global Practice. These cases were selected from the initial review of the portfolio and additional screening questions answered by TTLs to ensure that programs were scaled up and possessed key features. The purposively selected cases included Ghana, Uruguay, and Vietnam, reflecting various implementation contexts. The case studies were supported by a protocol that was informed by the literature review on in-service training and the literature review and background paper on scaling up.

Collection and analysis. The responses to each question from the protocol for each of the cases were summarized and collated for ease of analysis. Cross-case analysis was used to examine patterns and divergence. The analysis examined the extent to which particular characteristics (of in-service training or scaling up) were evident and which contributed to the development of the theory of change.

Quality Assurance

The evaluation was subject to IEG's standard quality review. The external peer reviewers were Barbara Bruns (Center for Global Development) and Eleanor Villegas-Reimers (Boston University). The team also followed IEG's quality assurance process and worked closely with IEG's Methods Advisory Team during all phases of the evaluation.

In line with the practice of meso evaluations, a highly consultative process was used with staff and management in the Education Global Practice to develop the evaluation's scope and focus. The consultation was to ensure the relevance of evaluation methods and questions.

Ensuring the Validity of Findings

The mixed methods approach triangulated the findings from multiple sources of data collected using the evaluation questions set out in the concept note. The extensive collection of data from multiple sources was necessary to cover the various subject matters addressed by the evaluation and ensure robust interaction of layered perspective to arrive at secure findings.

In collecting and analyzing data, the evaluation team consistently and cohesively used a common protocol for the literature review and a common protocol (built from the literature) for the case studies, backed by common definitions grounded in background papers. The evaluation design balanced the trade-off between breadth of coverage (as a basis for generalizability) and depth of analysis (as a basis for understanding contextual factors). Hence, to support generalizability of the findings, the evaluation assessed the extent of convergence across multiple sources of evidence, and the methods used built on each other to form a secure, layered base that held the multiple sources in tension to validate or negate emerging themes. Collated material was subject to further analysis through the construction of word tables and team discussion in an iterative manner that sought to verify and support emerging findings.

Limitation

Data. There are limited quantitative data captured and available regarding World Bank engagement preservice and in-service teacher training. As such, it was not possible to identify the precise cost of training supported by the World Bank and relative efficiencies in that regard.

Available evidence. The evaluation engaged in extensive portfolio review work to unearth key features associated with preservice and in-service training supported by the World Bank; however, the PADs provide limited information in that regard. To address this gap, the evaluation team conducted interviews with TTLs. This mitigating exercise was successful in generating additional evidence about 38 of the 40 operations for which basic data were unavailable through the PAD. The exercise also confirmed the overall patterns identified in the review of the PADs for which information was available.

Selection of country cases. It was not possible within the scope of this evaluation to select countries for case work on a representative basis. Instead, the cases were purposively selected in consultation with the Education Global Practice so that the evaluation could engage with broadly successful training efforts that involved some element of scale-up. This allowed the team to focus on the evaluation's core concerns and, in particular, identify factors and characteristics associated with scale-up.

---

1 For the Organisation for Economic Co-operation and Development Library, the literature search dates to 2003, not 2000 like the others.

2 This general search covered the first 100 items identified (ordered by relevance) within the period October 1, 2008, to October 1, 2018. All 100 items were reviewed, and 30 were considered relevant.

3 Inclusion and exclusion criteria were also applied to a small number of reports that the task team leader preidentified.

4 Search fiscal year (FY)13–18, OPCS Sector and Theme Codes: http://www.worldbank.org/projects/sector?lang=en&page= http://www.worldbank.org/projects/theme?lang=en&page= File from World Bank Business Intelligence Portal and Analysis for Office Application. Applying filters: **Theme codes:** 654—Teachers. **Sector Codes**: EC—Early Childhood Education; EP—Primary Education; ES—Secondary Education; ET—Tertiary Education; EW—Workforce Development and Vocational Education; EL—Adult, Basic and Continuing Education; EF—Public Administration—Education; and EZ—Other Education**.**

5 Search FY13–18, World Bank Education Projects Database" http://datatopics.worldbank.org/education/wQueries/qprojects. Applying filters: **Activities:** (i) Curriculum and Textbooks: Teacher Training for learning materials use; (ii) Teachers: In-service teacher training; (iii) Teachers: Preservice teacher training; (iv) Teachers: teachers' certification; (v) Teachers: teacher training system restructuring; and (vi) Teachers: training of teacher educators.

Appendix B. Basic Characteristics of World Bank Teacher Training Programs

Nearly half of the appraisal documents (51 of 108 project appraisal documents [PADs]) contained an identifiable component or subcomponent for training with cost data. Overall, the pedagogical training median cost per project was 15 percent of the total; the average per project training cost was 22 percent (figure B.1). These ratios varied from as little as 2 percent to as much as 80 percent of operation cost. However, in some instances, such as the case of the data point with 80 percent training cost to total funding, the operation was an additional financing supplement to the original operation (thus probably an outlier). Therefore, the median ratio of 15 percent training to total cost is more representative of the reviewed sample of 51 operations. This ratio suggests that training is largely implemented as part of larger development objectives, such as a general education reform or improving quality of education.

Figure B.1. Distribution of Training Program Cost as Share of Total Project Cost
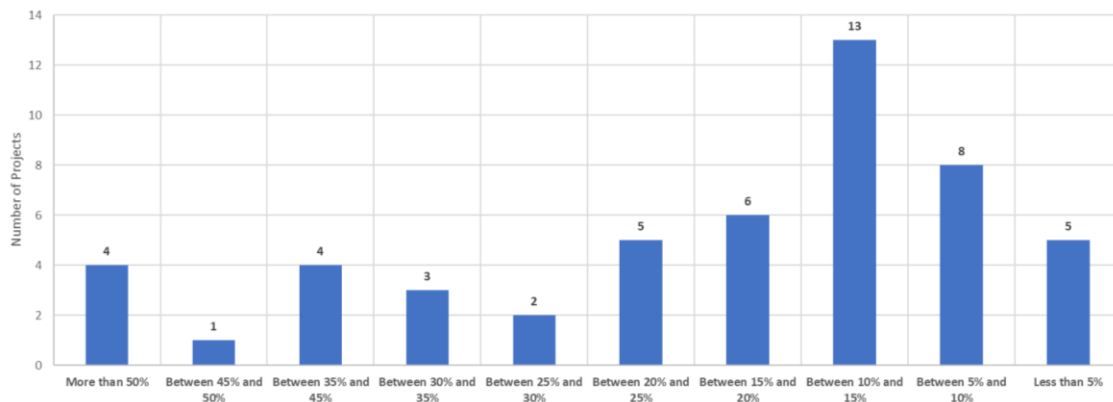
Appraisal documents did not provide the cost of training as a share of public education expenditures. It is important for governments to monitor overall spending on both pre- and in-service training activities. This makes it possible, at a minimum, to monitor the overall share of spending devoted to these activities, which can then be used to set targets based on goals and outcomes, and to compare spending shares on training with other countries in the same way that overall spending on education is often used to assess adequacy. Ideally, spending on training activities would be disaggregated, at least to some degree, to allow for even more detailed monitoring and tracking.

The spending tracking is relatively straightforward for in-service training because the costs are related mainly to trainers, materials, and travel for both trainers and trainees. Preservice training is another matter: Following the key drivers in this report, preservice

training investments include infrastructure, materials, training staff and personnel, scholarships, practicums (which are similar to in-service training), and examinations.

There are several challenges to improve data collection related to the cost of training programs. First, the costs of training are potentially shared across ministries for things like paying civil servants (teacher trainers) and infrastructure. Second, personnel involved in training might have other functions, such as district officers who are in charge of in-service training activities. Third, in many countries, there is a sizable university presence in preservice training that greatly complicates spending summaries compared with countries where preservice training is conducted in government-run institutions.

Eighty-six projects in low- and lower-middle-income countries supported training compared with 24 projects in upper-middle- and high-income countries. Only three projects were in high-income countries (Argentina and Uruguay). As figure B.2 shows, the Africa Region was most frequently supported, followed by Latin American and the Caribbean, South Asia, and Europe and Central Asia. The Middle East and North Africa and East Asia and Pacific regions had fewer training programs, which is consistent with education operations overall.

Figure B.2. Projects Approved since FY13 with Training Programs
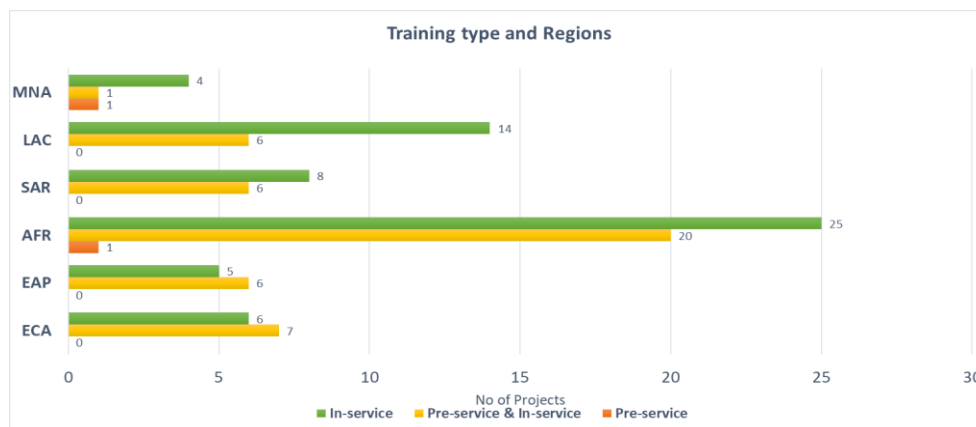
Source: Independent Evaluation Group coding of project appraisal documents.
Note: AFR = Africa; EAP = East Asia and Pacific; ECA = Europe and Central Asia; FY = fiscal year; LAC = Latin America and the Caribbean; MNA = Middle East and North Africa; SAR = South Asia.

A variety of training modalities were used in the 108 programs identified, but the most common were teacher training centers (17 percent), school-based training (17 percent), or distance learning (16 percent). Workshops were specified in 13 percent of PADs. The cascade approach (figure B.3) was by far the least common modality (9 percent). However, more than one-quarter of all operations did not specify the modality, leaving some room for interpretation.

Figure B.3. In-Service Training Programs Approved FY13–18, by Modality and Percent
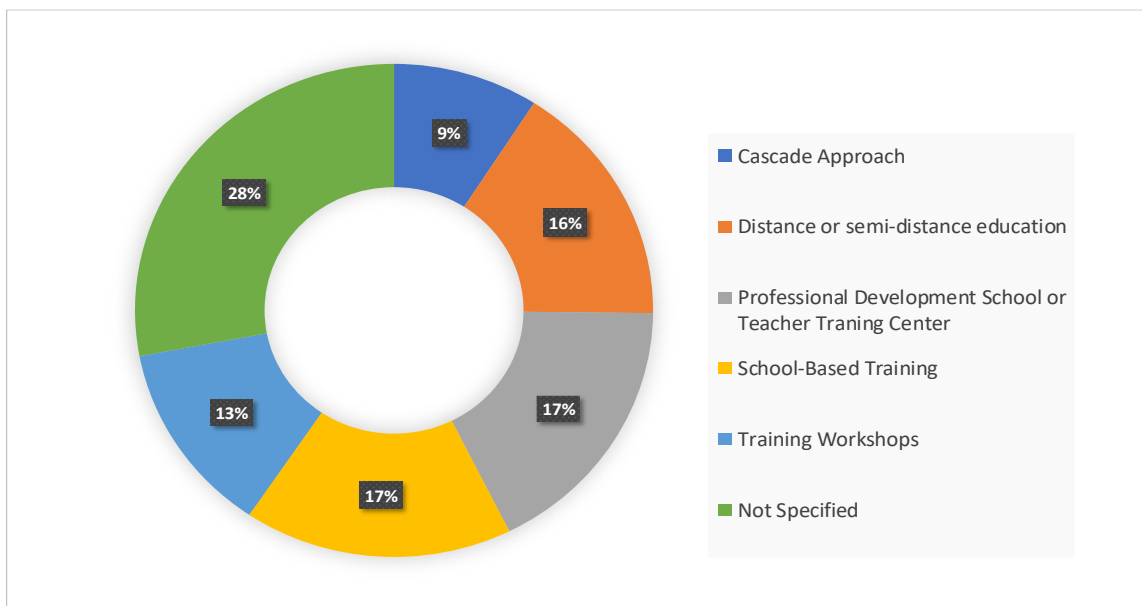
Source: Independent Evaluation Group coding of project appraisal documents.
Note: FY = fiscal year.

Box B.1. What Is Cascade Training?

Cascade training is training that relies on those receiving the training to become trainers in the next phase. The number of cascade phases used depends on the context, the design of the cascade, and the number of teachers targeted. An advantage of the cascade is that it can reach a large number of teachers rapidly with fewer resources.

A concern with cascade training is the lack of consistency in the content from one phase to the next. Factors that mitigate this problem include ensuring that training is based on teachers' needs, selecting experts and the first phase of trainers carefully, monitoring each phase of training, and providing comprehensive education materials for trainers and trainees.
Sources: Karalis 2016; Orr and others 2013.

Research is mixed on whether the training mode or the quality and fidelity of the training matters more. Some studies have found that the cascade approach is associated with negative outcomes (Orr and others 2013; Popova and others 2018), partly because of dilution of content over multiple phases (Orr and others 2013; Karalis 2016), though it is possible to minimize that effect. Observations from task team leaders and fieldwork in Vietnam suggest that outcomes depend on how the cascade model is implemented. For distance methods to work effectively, reliable access to technology is needed (Orr and others 2013); however, some authors have found this mode associated with negative outcomes (Popova and others 2018). Case studies of professional development across a

variety of countries also showed that a variety of modes can be successful, suggesting that it is not the mode but other factors that affect outcomes (Reimers and Chung 2018).

Training is provided for all educational levels, but primary teacher training is emphasized more (figure B.4), consistent with the World Bank's focus on primary education in low-income countries. Nearly half of all projects provide training to primary teachers. Generally, operations contained training programs that addressed multiple levels simultaneously. For example, 24 percent addressed both primary and secondary levels. Recipients of training programs were teachers and instructional leaders in the system. This is important to ensure coherence with the feedback provided by school managers and principals, head teachers, or inspectors.

Figure B.4. In-Service Training Programs Approved FY13–18, by School Level and Percent
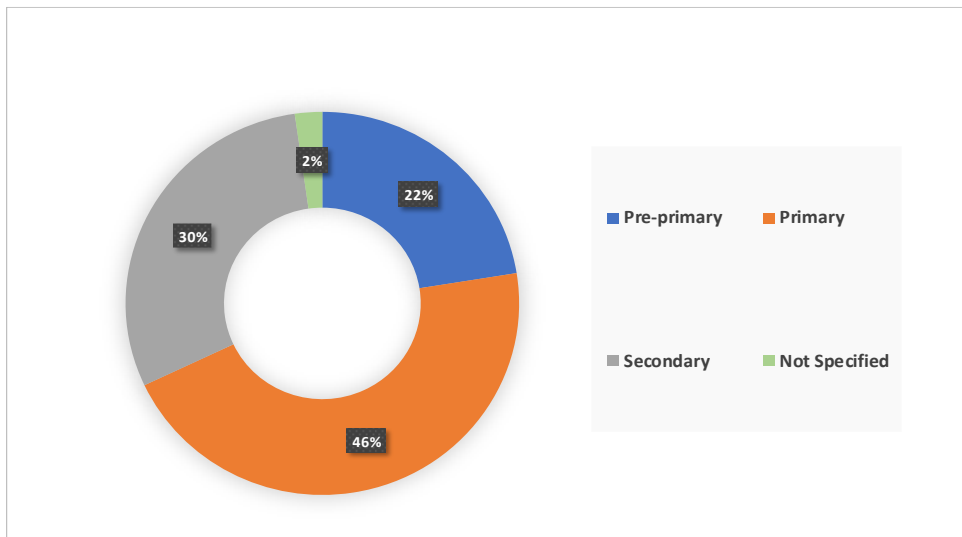
Source: Independent Evaluation Group coding of project appraisal documents.
Note: Categories are not mutually exclusive. FY = fiscal year.

More often than not, World Bank lending for teacher training was not at the national level. The larger the country, the less likely that the program was at the national level, consistent with subnational responsibilities for education in some countries. For example, in Brazil, only one of five projects was at the national level. Instead, 46 percent of operations focused on geographic areas, subnational, or less-than-national coverage, and 39 percent were at the national level. Smaller countries, like Tanzania, Uruguay, or Vietnam, often had a national program. However, national training programs typically do not provide countrywide universal coverage, as with Uruguay's training program that occurs in 20 percent of full-time public schools. By contrast, the early childhood education project in Vietnam provided training to all preschool teachers in every province.

Pedagogy was the focus in about half of the appraisal documents (figure B.5) regardless of the school level. The expectation would be to see more discipline-focused training for secondary teachers, but this was not the case. The heavy focus on pedagogy might be to address shortcomings in preservice preparation. Task team leaders reported a need to focus much more on discipline-specific content, given teachers' limited capacity to teach numeracy, literacy, and science.

Figure B.5. Content Focus of Training Programs (Percent)
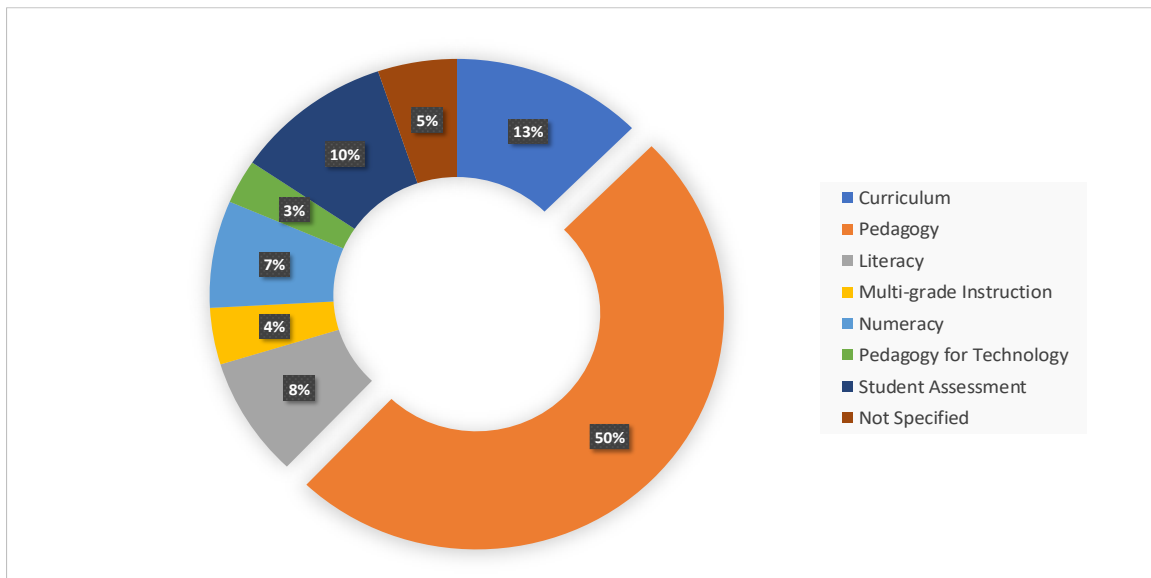
The incentives found in PADs were predominantly professional rather than accountability or financial related. The most common professional incentives were certification (18 percent) and mastery of professional knowledge (19 percent). This finding is consistent with the World Bank's use of in-service training programs to address qualification of teachers who are unqualified or undertrained.

Figure B.6. Incentives, by Type and Percent

Legend:
- ■ Accountability Pressure: Job stability
- ■ Accountability Pressure: Principal/head teacher/inspector feedback
- ■ Financial Incentives: Bonus pay
- ■ Financial Incentives: Career Opporutnities
- ■ Financial Incentives: Differential salary
- ■ Financial Incentives: Per diem during training
- ■ Professional Incentives: Certification
- ■ Professional Incentives: Intrinsic motivation
- ■ Professional Incentives: Mastery and professional growth
- ■ Professional Incentives: Recognition and prestige
- ■ Not Specified

Source: Independent Evaluation Group coding of appraisal documents.

The World Bank's experience with preservice training is limited. Of the 40 appraisal documents examined, 35 discussed features regarding preservice, and fewer of those provided details of the features supported. In addition, documents provided limited contextual discussion of the preservice institutional features, which raises the question of whether adequate attention was devoted to contextualizing the assessment of preservice institutions. The drivers covered in each Region are shown in table B.1.

Table B.1. Preservice Operations by Driver Disaggregated by World Bank Region (number)

Region (projects *[no.]*)	Entry and/or Exit Examination	Coursework	Projects with Financed Activities	Practicum Quality Assurance
Africa (18)	4	13	9	3
South Asia (4)	1	4	1	1
East Asia and Pacific (5)	1	5	2	2
Latin America and the Caribbean (5)	2	4	1	2
Europe and Central Asia (7)	0	2	1	1
Middle East and North Africa (1)	1	0	1	0
Total (40)	9	28	15	9

References

Karalis, Thanassis. 2016. "Cascade Approach to Training: Theoretical Issues and Practical Applications in Nonformal Education." *Journal of Education and Social Policy* 3 (2): 104–108.

Orr, David, Jo Westbrook, John Pryor, Naureen Durrani, Judy Sebba, and Christine Adu-Yeboah. 2013. *What Are the Impacts and Cost-Effectiveness of Strategies to Improve Performance of Untrained and Under-Trained Teachers in the Classroom in Developing Countries?* London: EPPI Centre, Social Science Research Centre, Institute of Education, University of London.

Popova, Anna, David K. Evans, Mary E. Breeding, and Violeta Arancibia. 2018. "Teacher Professional Development around the World: The Gap between Evidence and Practice." Policy Research Working Paper 8572, World Bank, Washington, DC.

Reimers, Fernando M., and Connie K. Chung, eds. 2018. *Preparing Teachers to Educate Whole Students: An International Comparative Study.* Cambridge, MA: Harvard Education Press.

Appendix C. Key Features of Preservice Training from Review of Literature

This appendix discusses the key drivers of quality teacher preparation in detail. Policy discussions about effective preservice training regimes tend to focus on three variables: screening measures to get the best people into the field, coursework and preparation that is coupled tightly with the actual work of future teachers in the classroom, and effective practicums that provide exposure to actual teaching experiences.

The data sources used here include high-profile summaries of teacher education preparation policy and teacher development. These studies are augmented with smaller, more focused studies that tend to be from developing countries and often concentrate on a single topic (like the practicum) or country context.

The literature has several limitations. First, the widely cited discussions on teacher preparation tend to come from industrialized settings like Europe or the United States, which potentially limits their applicability to the low- and middle-income countries that are the focus of this review. Second, the empirical basis for what works in teacher preparation is very limited. The data challenges in linking features of the preservice training experience with teacher capacity and student outcomes are considerable, and surprisingly few studies have taken up this question with measures that go beyond basic indicators of teacher education or practicum exposure (Boyd and others 2008). As a result, the evidence base for effective teacher preparation relies mainly on expert opinion. It also relies on somewhat functional depictions of effectiveness where the approaches taken by countries with high levels of student achievement are used as reference points (or contrasted with countries with low scores on international tests, that is, functional contrasts).

Table C.1 is a global summary of the results from the literature review for effective features of preservice training. The indicators are grouped into four categories; subcategories are used to provide additional details. The key data sources constitute the seven columns.

Table C.1. Key Indicators for Evaluating Teacher Preservice Training Quality

Indicator	Preservice Training Program						
	SABER (2012)	TEDS studies	Bruns-Luque (2012)	World Bank (2019)	AMTE (2018)	UNESCO (2012)	Lewin (2004)
1. Attracting the best candidates							
Screening mechanisms	X						
Minimum education level (for example, ISCED 5A), subject matter specialization	X	X	X		X	X	
Academic achievement (such as marks)		X		X			
Entrance exam(s) focusing on specific skills (problem solving, interpersonal skills, and similar qualifiers)			X	X		X	
Interviews				X		X	
Alternative routes	X					X	
Nontraditional entrance (such as Teach for America)			X				
Shortened initial training with more on-the-job training							X
Stipends, tuition, and scholarships		X	X		X	X	
Mechanisms for underrepresented groups (such as top 10 percent of class)					X	X	
2. Teacher preservice preparation							
Institutional features							
Engagement with partners (researchers and community)					X	X	
Working partnerships with schools					X	X	
Closely linked with universities				X			
Research engagement (original research, conferences, and collaborations)					X	X	

Key Features of Preservice Training from Review of Literature

Indicator	Preservice Training Program						
	SABER (2012)	TEDS studies	Bruns-Luque (2012)	World Bank (2019)	AMTE (2018)	UNESCO (2012)	Lewin (2004)
Duration of program		X					
Quality of faculty							X
Balance of content and teaching knowledge						X	
Transparent selection mechanisms (of teacher educators)						X	
Infrastructure (space and local learning materials)							X
Support structure (counseling and exam preparation)					X		
Self-assessment and monitoring functions (data and case studies)					X		
Coursework and preparation							
Basic coursework support ("catch up")						X	
Adequate balance of subject matter							X
Pedagogy and pedagogical content knowledge based on local conditions and materials, "embedded in curriculum," evidence based				X	X		X
Education studies (local materials and curriculums, linked to local context)				X	X		X
Sustained learning experiences (coherence)					X		
Learner-centered						X	X
"Teacher-as-researcher" and DDDM							
Preservice practical experience							
Duration (and availability)	X			X	X	X	X
"Increasingly comprehensive" responsibilities		X			X	X	

Indicator	Preservice Training Program						
	SABER (2012)	TEDS studies	Bruns-Luque (2012)	World Bank (2019)	AMTE (2018)	UNESCO (2012)	Lewin (2004)
Close collaboration with schools					X		
Time for reflection and dialogue							X
Regulation support (avoid "sink or swim")							
Training Institute or capable/experienced teachers/school director engagement					X	X	X
Formative assessment							X
Materials							
3. Exit from training/entrance into profession							
Graduation/Postgraduation filtering							
Weak: Just graduate; Strong: exam plus teaching proficiency/portfolio			X				
Practice requirement							
Accreditation-Certification exams			X				
Mentoring, induction, and coaching programs, probationary periods	X	X	X				X
4. Sector control features							
National curriculum for teacher education (standards, core competencies, and other measures)			X		X	X	X
Tight coupling with actual work in classroom		X		X	X	X	
Widespread participation and ownership					X		

Appendix C
Key Features of Preservice Training from Review of Literature

Indicator	Preservice Training Program						
	SABER (2012)	TEDS studies	Bruns-Luque (2012)	World Bank (2019)	AMTE (2018)	UNESCO (2012)	Lewin (2004)
Evaluation mechanisms to verify						X	
Policy guidance on practicum responsibilities, supervision requirements		X					
Supply and demand "balance"							
Sufficient specialists across all areas							
Centralized control over entrants		X					
Teacher education institute quality assurance							
Effective accreditation oversight				X			
Ability to disaccredit/close institutions		X	X				
Regularly evaluate all institutions						X	
Support function						X	
Regulation mechanisms for creation of new careers						X	
Standardized mechanisms for evaluation of Graduates						X	
Competitive funding for programs			X				
Policies for teacher trainer recruitment, deployment, Development, and career opportunities						X	X
Setting up national training center(s)			X				

Note: AMTE = Association of Mathematics Teacher Educators; SABER = Systems Approach for Better Education Results; TEDS = Twins' Early Development Study; UNESCO = United Nations Educational, Scientific and Cultural Organization.

The table is exhaustive and can be used to inform World Bank task team leaders (and others) about effective approaches to preservice preparation. Unfortunately, the indicators in the table are not very specific and thus provide only general guidance. This reflects the lack of empirical evidence on what works with teacher education programs. However, it also reflects the difficulty of relying on simple indicators to capture inherently qualitative processes related to factors like teacher educator capacity or the effectiveness of practicums.

The remainder of the appendix reviews the four quality drivers for preservice training in detail.

Quality Driver 1: Screening and Filtering

Teachers encounter screening points throughout their professional life cycle where it is possible to impose conditions to restrict the number who pass to the next stage (figure C.1). For example, developed countries that use high-stakes screening mechanisms at multiple points along this sequence, such as Japan and Korea, tend to outperform countries with less stringent screening on international assessments.[1]

Figure C.1. Screening Points in the Teacher Education and Development Pipeline
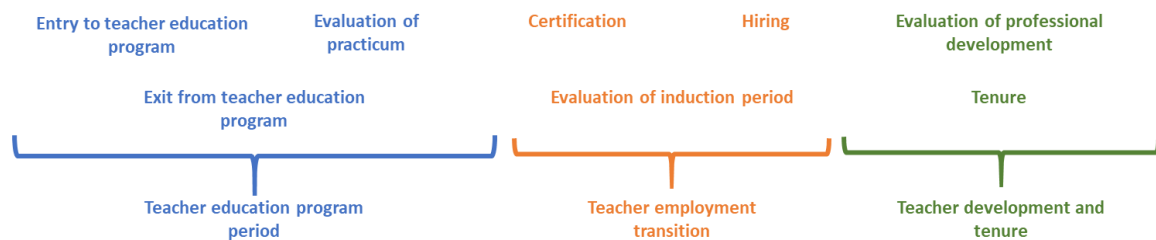
Source: Adapted from Wang and others 2003.

Countries (and institutions) that restrict participation to trainees with university-level training or specific coursework by subject (for future specialists), use filters related to student performance (marks) or entrance examination results, or assess potential based on interviews are more likely to bring in high-quality candidates. A sizable amount of (noncausal) evidence backs up this point, including from the Teacher Education and Development Study in Mathematics (TEDS-M) data analysis (appendix D).

However, what does this mean for the rest of the elements in the evaluation's conceptual framework (figure 1.2)? One implication of the screening is that how it is done with candidates is less important than what candidates bring into the system. This is a version of the screening versus human capital arguments about the role of schooling in developing skills. Given the widespread concern about teacher education system curriculums and effectiveness and the lack of evidence about specific features that

improve teacher capacity and student learning, the importance of initial screening in many discussions is understandable. It is also noteworthy that there is little empirical guidance—and, in fact, some disagreement among experts—about the specific ingredients that go into effective training systems.

The position of this review is that training and screening matter and need to be viewed as complementary features of effective preparation regimes. However, the screening argument is important and has implications for two additional aspects of initial selection.

Alternative routes. Alternative training routes can be used on an emergency basis or as a mechanism to attract a different kind of candidate. In the United States, this topic is controversial, and a large amount of literature is debating the relative effectiveness of programs like Teach for America. Some of the rhetoric associated with that program suggests that teacher education programs are obsolete and unnecessary, although over time, Teach for America supporters recognize the challenges of creating excellent teachers in a short period (Schneider 2014). The important point is that alternative routes can introduce dynamism into the teacher training system and provide new evidence on ways to best produce effective teachers. This is especially true in the poorest countries where traditional teacher training systems are overwhelmed and the exigencies of developing new teachers argues for looking into radical new approaches that bypass the traditional routes (Lewin 2004). This can include flexible hiring regimes that are accompanied by extended on-the-job training, for example.

Broadening participation. Broadening the participation of teacher education programs can help create a more diverse teaching corps that better reflects the student population with which they will work. However, this goal often comes with an inherent challenge related to standards and selectivity. One message from the literature is that teacher preparation systems need to be proactive and use incentives like scholarships to target candidates from underrepresented groups. Rules like the 10 Percent Rule in the state of Texas in the United States should also be considered—the student's relative standing in their school is used instead of their absolute standing (based on exam results and other factors). The underlying argument is that standout candidates from underrepresented groups are likely to have the potential to be excellent teachers, even if they do not appear to be as strong (on paper) as other candidates.

Quality Driver 2: Coursework

There are three broad categories of coursework for teacher trainees: content or subject matter knowledge, methods (including pedagogical content knowledge preparation),

and more general education studies topics related to child learning and development, student assessment, and the like.

Unfortunately, no clear guidance exists on the best mix of coursework for preparing teachers, especially given the variation in subjects and levels. In countries like the United States, this is (again) a controversial topic; mathematicians and other subject matter experts often argue for very strong preparation in subject matter while education experts emphasize the importance of knowledge of the curriculum teachers will actually teach. Additionally, initial selection mechanisms have far-reaching consequences for teacher coursework because a more capable trainee cadre is likely to have already completed higher-level coursework in the subject matter.

Teacher educator curriculums are often criticized for being outdated, placing too much emphasis on theory and not enough practical knowledge to prepare teachers for their future work in the classroom. In the poorest countries, these criticisms tend to be even more strident (Lewin 2004), and descriptions of teacher training programs fall far short of what is required to adequately prepare teachers. To back up this point, it is not unusual to find extremely low levels of subject matter knowledge (especially in mathematics) among teachers in developing countries. (For more information, see the Service Delivery Indicator results from the World Bank at http://datatopics.worldbank.org/sdi/.) This is clearly an aspect of teacher preparation that requires close attention.

Insuring basic levels of content knowledge. It is obviously important for subject specialist teachers to have a deep understanding of the topics that are their classroom responsibility, especially topics like mathematics and science. However, generalist teachers working in primary grades must have minimum levels of content knowledge across the core subjects, and again, mathematics often stands out as a weakness. Exit examination mechanisms are potentially important to verify that teachers have at least these basic levels of preparation.

Methods preparation embedded in the curriculum. Methods preparation includes general theory about teaching and specific teacher knowledge elements like pedagogical content knowledge. One of the main concerns about teacher preparation is that the curriculum and materials (when they exist) are too general, abstract, theory-based, or contextually inappropriate. Teacher training must be embedded in the curriculum that teachers will be using in their actual work and in the contextual reality of their future classrooms and communities. Boyd and others (2008) provides some empirical support for the embeddedness argument in their analysis of New York City teachers and learners.

Access to materials. The materials teachers use as part of their training must also be embedded in their eventual work in the classroom. The lack of materials—or lack of locally relevant materials—is a problem that comes up repeatedly in developing country descriptions of teacher training experiences (Lewin 2004 and Sorto and Luschei 2009).

Learner-centered versus hierarchical delivery. Caution is required when deciding on the best approach to teach at any level, and there are wide-ranging debates about the relative merits of constructivism versus more teacher-centered teaching. Nevertheless, there are reasons to be concerned about the predominant approach to preparing teachers in low-income countries and the way that teacher educator methods trickle down into the classroom work of future teachers. Teacher trainees need to be able to ask questions, have time for reflection and learning by trial and error, and receive instruction that accounts for their level of preparation. This includes the classroom as well as the practicum experiences (discussed below). But the evidence from case studies and other summaries shows a marked tendency toward hierarchical training of teachers that in effect treats them like children, which limits its effectiveness and tends to narrow the trainees' vision of how to work with children (Akyeampong 2017; Lewin 2004; and UNESCO 2012).

Teachers as researchers. High-performing systems both prepare (or select) teachers with high levels of capacity and engender a culture of continuous improvement and searching for solutions to address problems in the classroom and beyond. The teacher-as-researcher concept must begin during the teacher preparation phase. This is not just about developing tools (such as content knowledge, pedagogical content knowledge, and student assessment skills) to diagnose individual children and classrooms to inform change and adaptation; it also relates to the vision in place about how teachers perceive their work and responsibilities.

Quality Driver 3: Practicum

One of the critical phases of teacher preparation is the practice teaching phase, in which teacher trainees enter classrooms to learn through experience. Like other aspects of preservice preparation, this feature has multiple dimensions and could be the subject of a separate study. The goal is to highlight some key aspects of the practice to consider when reviewing the sector and possible entrance points for support.

Practicum policy and regulations. This is related to larger concerns about teacher preparation curriculums and their ability to effectively prepare teachers and adapt to ongoing changes in educational missions. For practicums, there seems to be a general lack of guidance on how this aspect of teacher education is managed, which by default leaves it up to individual training centers to decide. A national policy on practicums

should be considered that lays out minimum duration for classroom experiences along with rules about supervision, support, and responsibilities for both the teacher training center and the school where the practice takes place.

This type of policy or regulatory guidance is related to features of the education system. An underlying tension in this dimension concerns the best source for policy directives. There is always an argument for decentralized approaches, in which individual training institutes decide on the best practice based on their own experiences, characteristics of students, and local contextual features. However, national standards can be disconnected from classroom context and seen as too unwieldy. The important point is that some guidance is required for effective practicums, and the guidance should be based on a mix of best practices and research evidence. In low-income countries, capacity constraints could limit the ability of individual institutions to define these goals, which is why a national directive should be considered.

Availability and duration. The evidence suggests that most countries offer some practice teaching as part of the teacher preparation program. As with training program duration, there is no agreed-on minimum number of days or hours for effective practicums. This is partly because experiences during these practices vary, which can include teachers simply observing classrooms and schools in a support or observer role or actually leading an individual class. In theory, longer practicums are better, and very short practicums of days or a few weeks should be flagged. However, long practicums can be costly, and a poorly designed or supported practicum can be a frustrating experience for the teacher.

Trainee responsibilities and activities. The TEDS-M data include two categories of practicum: extended teaching, in which trainees spend two or more weeks being prepared to teach a class with students, and introductory field experiences, which are more short term or episodic and not focused on classroom work. Both versions have positive features, with the latter providing teachers with opportunities to observe schools and speak with staff, students, and others. The distinction is important, however, because it is necessary to clearly define what is happening when trainees are placed in schools. It is also important to explain that field experience visits are not a substitute for extended teaching practice.

Furthermore, within the extended teaching practice category is a range of roles and responsibilities that teachers may or may not be assigned. This is intricately related to the kind of support and supervision they receive during the practicum experience and regulations about practicums. Practicum experiences need to strike an effective balance between providing teachers with real-life experience in classrooms so they can learn, while ensuring that the experience is productive for them and the students. The best

description of what is required for effective practicums is from the Association of Mathematics Teacher Educators, which describes a process of "increasingly comprehensive" responsibilities for teachers who slowly take on more responsibilities as part of their practice teaching experience.

Support and supervision. This critical aspect was discussed concerning policy and trainee responsibilities. Discussions of effective teacher practice regimes focus on productive interaction between participants and supervisors from training institutes or schools, or both. This begins by forming productive partnerships with schools and creating a practicum system with input from many stakeholders (schools, training institutes, and others). Experienced mentors who are familiar with the needs of beginning teachers are crucial for creating a trainee-centered experience. Formative assessments are required where feedback is provided and adjustments are made, accompanied by reflection and dialogue. Teacher trainees also need quality curriculum and teaching material aids that are aligned with their work in the classroom.

The reality of the practicum experience can often fall short of the ideal description, and the more serious problems seem likely in the poorer countries (Akyeampong 2017). A common description is a "sink or swim" model in which trainees are given too much autonomy too early and have to deal with a range of issues for which their coursework—often quite theoretical—provided little or no guidance. These issues are not just about teaching the content, providing explanations, and designing lesson activities. Even basic elements, like controlling the class or dealing with behavioral issues, can pose sizable challenges to teachers in training. There is often a lack of materials to provide further support, and the assessment regime is based on summative, end-of-cycle reviews that may be based on a single observation or some type of test or review. This is not a recipe for effective use of the practicum time, and as a result, teachers enter their full-time positions with a limited set of tools to deal with the same issues and problems that arise during the practicum.

One of the critical deficiencies in the practicum is the lack of qualified teacher mentors in the schools where trainees spend their practicum time. It is difficult for training institutions to mobilize supervisors to monitor trainee practicums closely, which means that schools need to fill this supervision and support gap. However, when these teachers are poorly equipped to act as mentors or have a hierarchical or top-down approach to supporting teacher trainees, then critical teacher learning opportunities are lost. This is not an easy situation to correct in countries with widespread deficiencies in quality, where teacher practicums necessarily take place in average schools with low capacity, or trainees are grouped together in special or exceptional schools where there are fewer

opportunities to practice, and the context is likely to be very different from what they will face in their full-time work.

The practicum and lessons from in-service training. One indirect source of information on effective practicums—and teacher preservice training in general—is the professional development literature. These are not identical activities, but there is clearly some overlap because teachers have to implement new ideas in a classroom setting. Importantly, some of the key features of effective interventions identified in the professional development literature by Popova and others (2018) are relevant to the previous paragraphs on practice teaching, including the following:

- Content embedded in the curriculum

- Focus on a specific method with detailed instructions on implementation

- Significant and sustained in-person follow-up support

- Involved teachers in a co-learning model

In their review of empirical studies of teacher training, Popova and others (2018) also identified features that are associated with sizable or significant impacts on student achievement, including providing training participants with materials, a specific subject focus (versus general topics), incorporating lesson enactment into practice sessions, more consecutive days of face-to-face training, practicing with other teachers, and follow-up visits that focus on material covered in trainings.

Quality Driver 5: Quality Assurance

The final set of factors reference specific quality assurance mechanisms that affect preservice teacher training in many ways. This reflects the importance of the institutional setting for determining preservice quality—high-quality teacher training centers do not develop overnight but are often the result of a sustained set of policies that can have deep roots in the education sector.

National curriculum and standards for teacher education. The potential role of curriculums and standards has already been referenced in relation to training center content and the practicum. The critical feature is the coupling (or embeddedness) of the standards with the actual work of teachers. They must reflect the actual needs of students and teachers and the working conditions where these activities take place.

The national teacher education curriculum and standards should be defined based on widespread participation and input from stakeholders, including teachers, communities, subject matter experts, administrators, and others. The standards should be informed by

research and be adaptable to new developments, especially related to technology. Evaluation mechanisms should be in place to monitor progress on implementation and actual use.

Teacher training institution quality assurance. The TEDS-M study on quality assurance (Ingvarson and others 2013) is an excellent summary of the role of accreditation in affecting teacher preparation quality. Countries with weak accreditation systems have no effective control over training institutions or they rely on voluntary participation mechanisms. At the other extreme, countries with strong accreditation have external agencies with the power to disaccredit (or shut down) training centers.

Accreditation is a potentially important lever in determining quality, but like any institutional feature, it needs to be based on real power and capacity. Just having powers in name only will not matter if these are not actually used, and uneven regulation activities—including political interference—can undermine the work. Capacity issues are also crucial because the accreditation agency must have the ability to effectively monitor and regulate training institutions, which includes verifying that filtering mechanisms, like examinations, are being implemented transparently and efficiently.

The accreditation process is not just about holding training institutes accountable to a standard of quality. This process should also have a support component for developing institutions to the necessary standard. This requires regular interaction and consistently applied evaluations that are iterative and build on each other.

The Bruns and Luque (2015) review of teachers (*Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*) in Latin America identified another potentially interesting quality assurance mechanism for teacher training providers: competitive funding of activities. This is not a direct quality assurance mechanism but one that works to improve capacity through competition. An added benefit is that the new activities funded could be used to test new ideas related to training.

References

Akyeampong, K. 2017. "Teacher Educators' Practice and Vision of Good Teaching in Teacher Education Reform Context in Ghana." *Educational Researcher* 46 (4): 194–203.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Teacher Preparation and Student Achievement." National Bureau of Economic Research (NBER) Working Paper 14314, NBER, Cambridge, MA.

Bruns, Barbara, and Javier Luque. 2015. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean.* Washington, DC: World Bank.

Ingvarson, Lawrence, John Schwille, Maria Teresa Tatto, Glen Rowley, Ray Peck, and Sharon L. Senk. 2013. *An Analysis of Teacher Education Context, Structure, and Quality Assurance in TEDS-M Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M).* Amsterdam: International Association for the Evaluation of Educational Achievement.

Lewin, Keith M. 2004. "The Pre-Service Training of Teachers: Does It Meet Its Objectives and How Can it Be Improved?" Background Paper commissioned for the Education for All Global Monitoring Report 2004, The Quality Imperative.

Popova, Anna, David K. Evans, Mary E. Breeding, and Violeta Arancibia. 2018. "Teacher Professional Development around the World: The Gap between Evidence and Practice." Policy Research Working Paper 8572, World Bank, Washington, DC.

Sorto, M. A., and T. F. Luschei. 2010. "Teachers' Education, Supervision, and Evaluation in Costa Rica." In *International Handbook of Teacher Education World-Wide: Issues and Challenges, Volume II*, edited by K. G. Karras and C. C. Wolhuter, 727–746. Athens: Atrapos Editions.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2012. *Antecedentes y criterios para la elaboración de políticas docentes en América Latina y el Caribe.* Paris: UNESCO.

---

[1] There are several caveats regarding the connection between teacher education and student outcomes. First, teacher education is not a consistently significant predictor of outcomes like student achievement. It tends to be moderate in size, even when significant, and explains little of the effect individual teachers have on students.

Appendix D. Secondary Data Analysis: Preservice Training

Reducing a large amount of literature to a core set of lessons that provide guidance across a range of contexts and varying the messages by levels of teacher preparation (primary and secondary) are two limitations for any review of preservice training. However, the biggest limitation is the lack of causal evidence based on experimental designs with random assignment (Goldhaber 2019). The fundamental complication is the self-selection that is inherent in initial entrance into the profession (Harris and Sass 2011). However, other aspects of preservice training, such as the practicum, have little use for random assignment to inform policy, which is a major difference between the in-service training literature.

The lack of empirical evidence means that the most referenced summaries of preservice preparation effectiveness tend to be based on experiences in developed countries or they rely on functional reference points where high scores on international achievement tests implicitly validate the approaches taken, such as in Taiwan, China, or Singapore. Lessons from developed countries can be applied to other countries when findings are adequately contextualized.

This review uses international and regional assessment data to deepen the analysis of the large number of factors identified in the preservice training literature. This includes descriptive summaries, comparisons, and multivariate analysis that digs deeper into the underlying factors associated with outcomes like teacher content knowledge. In some cases, the results provide empirical support for a feature that is highlighted in the general summaries. However, the more common result is a lack of robust guidance on the topic, with inconsistent (or insignificant) results in the multivariate analysis or comparisons across countries that do not show a clear relationship between the feature and teacher capacity or student achievement outcomes. Therefore, the results are generally consistent with the larger empirical literature on preservice training and provide a general checklist of features of effective preservice training regimes.

The three data sets that were incorporated in the analysis are described in the next sections, including summaries of the key variables that are available, limitations, and the statistical modeling that was carried out. This is followed by a summary of the main findings categorized by quality driver.

Teacher Education and Development Study in Mathematics

The 2008 Teacher Education and Development Study in Mathematics (TEDS-M) data from 17 countries are used to assess the relationship between features of teacher training

programs and teacher capacity. The data have three features that are especially useful. First, the study subjects are students enrolled in teacher training institutions. This makes it possible to analyze teacher training outcomes—or teacher education program production—based on training experiences without the delay that comes with analyzing data from in-service teachers who might have completed their training 20 years earlier.

Second, the data have two measures of teacher capacity related to mathematics content knowledge (MCK) and mathematics pedagogical content knowledge (MPCK). The tests were developed to capture meaningful aspects of what teachers need to know to be effective in the classroom (Tatto 2013), and build on important lines of inquiry into what makes some teachers more effective than others (Shulman 1986).

Third, the data include a detailed summary of teacher training experiences. This refers to individual trainee experiences with coursework and practicums, and institutional features related to admissions, teacher trainer qualifications, and autonomy, among many others.

The data also have limitations. The information is dated 2008 and includes only measures of mathematics teacher capacity. The MCK and MPCK measures are based on constructs about what teachers should know to be effective rather than actual links with teaching quality in the classroom or student achievement. The cross-sectional nature of the data limits the ability to establish causation between X and Y, and therefore limits the discussion of key features of preservice training to statistical associations.

The estimation strategy regresses outcome Y onto vectors of individual teacher background (X) and selection (S) variables, and individual and institutional indicators (T) that capture different features of the teacher education program experience:

$$Y_{in} = \beta_X' X_i + \beta_S' S_i + \delta T_{in} + (\epsilon_{in}, \pi_n) \tag{1}$$

When the teacher education feature is an individual trainee measure (see T_{in} in equation 1), then the estimation includes institutional fixed effects (π_n); when the feature is measured at the institutional level, then the fixed effects are relaxed. The TEDS-M data rely on a replicate weighting setup that is incorporated in all estimations.

Southern and Eastern Africa Consortium for Monitoring Educational Quality

Fifteen countries participated in the third regional assessment organized by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) III in 2006–11. In each country, representative samples of grade six students

completed tests in reading, mathematics, and health and HIV knowledge. Grade six teachers also completed tests in these three subjects.

The unique feature of SACMEQ is the availability of teacher test scores for a number of low- and middle-income countries. The statistical analysis focuses on the teacher test scores, with comparisons by preservice training and education experiences, subject specialization, in-service training, and teaching materials. This was done using both descriptive and multivariate methods. The regression work was based on an equation similar to equation 1, with extensions for random and fixed effects designs, although these were complicated by the small number of teachers in many schools.

Like TEDS-M, the SACMEQ III data are somewhat dated, and again the cross-sectional design limits causal inference. However, the availability of teacher test scores in the Africa Region makes it possible to tentatively extend the analysis of preservice training effectiveness to a quite different set of countries compared with TEDS-M.

Third Regional Comparative and Explanatory Study

The Latin American Laboratory for the Assessment of the Quality of Education conducted its third regionwide assessment of student achievement, the Third Regional Comparative and Explanatory Study (TERCE), in 2013 in 15 countries. Representative samples of third grade and sixth grade students completed standardized tests in reading and mathematics, and they were linked with their teachers, who completed detailed background questionnaires.

The TERCE data were chosen as a complement to the TEDS-M and SACMEQ data because they include information on teachers' preservice training experiences, including practicums, and because they cover another region of low- and middle-income developing countries. There is also no information on teacher capacity as measured by pedagogical content knowledge, in addition to the standard limitations of large-scale assessments that cannot clearly establish causation. The analysis again relies on a mixture of descriptive comparisons and multivariate regression work, but in the latter, the dependent variable is for student achievement rather than teacher knowledge.

Results

The secondary data analysis generated an enormous amount of output, especially for the TEDS-M work, given the availability of dozens of variables related to the key preservice training features that were identified in the literature review. Overall, the descriptive comparisons and multivariate modeling did not identify very many variables that are consistently associated with teacher capacity measures (or student achievement for

TERCE). This result is itself important and highlights the inherent challenge in this line of inquiry.

The results are divided into the main quality drivers (figure 1.2 in the main report). The bulk of the results are from TEDS-M, but additional findings from SACMEQ and TERCE are also incorporated.

Screening of teacher candidates. Among the quality drivers in this review, screening receives the most empirical support. Table D.1 summarizes the covariates of MCK for TEDS-M secondary-level teacher trainees with institutional fixed effects. The results show higher levels of pretraining mathematics coursework, and the trainee's self-reported performance in high school is significantly associated with content knowledge in most countries.

Figure D.1 shows that among the lowest-scoring quintiles for TEDS-M trainee MCK, a sizable percentage of trainees completed only grade 11 or lower mathematics, while a majority of the highest scoring trainees (about 64 percent) completed to grade 12 (advanced). Figure D.2 makes a similar argument based on SACMEQ III data comparisons of high- and low-scoring countries based on teacher content knowledge: Most of the Kenyan teachers (high scoring) have completed senior secondary or higher levels of school, whereas in the lowest-scoring countries, significant proportions of teachers have only primary or lower secondary schooling.

Appendix D
Secondary Data Analysis—Preservice Training

Table D.1. Covariates of Secondary Teacher Trainee MCK: Base Model with Institutional Fixed Effects

Independent Variables	CHL	TAI	GER	MAL	PHI	POL	RUS	OMAN	SING	SWI	THA	USA	NOR
Age (in years)	−1.19	−1.7	2.0**	1.3	−1.2	−5.8	−0.4	−6.6	−3.0**	−2.1	1.1	−0.5	−0.5
	(−1.02)	(−0.78)	(2.66)	(0.73)	(−0.81)	(−1.65)	(−0.33)	(−1.60)	(−3.64)	(−1.44)	(0.51)	(−1.08)	(−0.99)
Gender = male	17.4+	14.8*	28.2**	10.7+	15.6**	32.9**	−6.6	−13.7	22.9**	12.4	6.2	37.7**	29.2**
	(2.76)	(2.00)	(3.73)	(1.73)	(3.15)	(3.29)	(−1.39)	(−1.58)	(3.65)	(1.42)	(1.44)	(5.90)	(4.58)
Books in home	0.7	0.9**	−0.2*	−1.7	−0.8*	2.5	0.3	−0.09	4.5+	−7.9	−0.08	6.1*	7.3*
	(0.19)	(5.16)	(−2.19)	(−0.64)	(−2.15)	(0.65)	(0.22)	(−0.60)	(1.90)	(−1.45)	(−0.42)	(2.31)	(2.35)
Parental education	1.4	0.1	2.1	4.8*	−1.8	0.02	−1.4	−2.7	7.0**	−0.7	2.7*	−1.3	−1.7
	(0.63)	(0.03)	(0.88)	(2.28)	(−1.57)	(0.01)	(−0.97)	(−1.13)	(3.12)	(−0.29)	(2.41)	(−0.46)	(−0.73)
Highest math level (excluded = year 12 advanced)													
Year 12 (nonadvanced)	−20.7**	−71.1**	3.0	11.2	2.4	−5.8	—	−28.2+	−41.5**	7.0	−16.1**	−32.6*	−10.6
	(−2.86)	(−5.44)	(0.22)	(1.16)	(0.10)	(−0.68)		(−1.72)	(−3.66)	(0.47)	(−3.95)	(−2.57)	(−0.68)
Year 11	0.6	—	−97.7**	−11.0*	−8.1	—	—	−22.5	−4.5	—	—	−18.2	−15.2
	(0.02)		(−4.65)	(−2.33)	(−0.47)			(−1.09)	(−0.59)			(−1.56)	(−1.52)
Year 10	−21.7	—	40.5+	−39.4**	25.5	—	—	—	−18.2*	−21.2	24.7+	—	−21.1*
	(−0.73)		(1.81)	(−2.63)	(1.35)				(−2.20)	(−1.40)	(1.96)		(−2.24)
Below year 10	−3.1	−40.6+	—	—	10.0	—	—	—	—	−37.9*	−19.4	−33.3	−30.5
	(−0.19)	(−1.77)			(0.60)					(−2.16)	(−1.62)	(−0.92)	(−0.58)
Missing or country specific	−122.1*	−42.2**	—	−231.7*	−0.2	—	—	—	—	—	−75.4**	—	—
	(−2.11)	(−2.69)		(−25.90)	(−0.01)						(−5.30)		
Usual level of grades in high school (excluded = Always at the top)													
Usually near the top	−7.9	−15.1+	−21.6	−3.8	−4.6	−0.17	−2.5	−2.4	4.8	−26.7	−14.7	−5.7	−28.0*
	(−0.77)	(−1.70)	(−1.10)	(−0.84)	(−0.50)	(−0.01)	(−0.55)	(−0.31)	(0.51)	(−1.33)	(−1.36)	(−1.07)	(−2.15)
Generally above average	−11.0	−14.3	−36.2+	−14.7*	−20.9*	−8.2	−14.9**	4.8	−16.3+	−46.0*	−25.3*	−21.9*	−31.2**
	(−1.04)	(−1.58)	(−1.77)	(−2.26)	(−2.56)	(−0.55)	(−2.70)	(0.31)	(−1.90)	(−2.08)	(−2.30)	(−2.15)	(−2.74)
Generally about average	−6.9	−24.5+	−50.8*	−17.7*	−16.9*	−22.9	−5.4	—	−16.5	−69.0**	−34.8*	−6.5	−45.2**
	(−0.78)	(−1.78)	(−2.36)	(−2.12)	(−2.36)	(−1.39)	(−0.77)		(−1.62)	(−2.65)	(−2.57)	(−0.80)	(−3.38)
Generally below average	−27.2	−13.4	−97.8**	—	—	—	−93.2**	—	−12.9	−29.1	−33.7+	−31.4**	−55.8**
	(−1.37)	(−0.78)	(−3.44)				(−3.04)		(−0.77)	(−1.23)	(−1.77)	(−5.51)	(−2.64)

Independent Variables	CHL	TAI	GER	MAL	PHI	POL	RUS	OMAN	SING	SWI	THA	USA	NOR
Had prior career before teacher training	1.5	−48.4**	−31.6**	−7.9+	−11.5**	−27.2	−7.0	−1.7	−7.9	12.1	−6.7	−18.5**	3.2
	(0.22)	(−2.65)	(−2.89)	(−1.67)	(−2.67)	(−1.40)	(−0.89)	(−0.10)	(−1.17)	(0.98)	(−0.99)	(−3.07)	(0.46)
Chose teaching for love of mathematics (excluded = not a reason)													
A minor reason	13.3+	31.4**	103.7**	−19.0	−25.1**	−2.8	26.4	5.6	−8.7	−18.3	1.1	21.3*	4.6
	(1.81)	(2.85)	(2.58)	(−1.65)	(−3.71)	(−0.19)	(1.65)	(0.20)	(−0.97)	(−1.53)	(0.11)	(2.66)	(0.66)
A significant reason	18.8*	40.0**	108.7**	5.8	−25.5**	1.6	30.3*	16.3	0.5	−12.8	3.6	29.8**	30.1**
	(2.32)	(4.24)	(2.79)	(0.59)	(−4.40)	(0.10)	(2.18)	(0.71)	(0.06)	(−1.08)	(0.43)	(3.53)	(3.93)
A major reason	10.4	35.7**	134.2**	13.0+	−22.6**	15.1	31.6*	17.4	−0.7	−7.4	16.1	42.9**	46.2**
	(0.79)	(3.58)	(3.45)	(1.77)	(−4.91)	(0.82)	(2.53)	(0.68)	(−0.07)	(−0.72)	(1.64)	(4.67)	(3.95)
Teacher preparation level (excluded = lower secondary, grade 10)													
Lower and upper secondary (grade 11)	—	—	72.3**	—	—	40.9**	—	—	34.5**	—	—	54.6**	—
			(8.59)			(3.21)			(6.14)			(10.03)	
Institution fixed effects	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country mean (SD)	354.2	667.3		493.4	441.5	540.3	593.5	472.0		531.1	479.0	505.4	435.3
	(84.4)	(75.3)		(50.8)	(49.0)	(66.0)	(96.2)	(47.2)		(50.2)	(58.6)	(66.6)	(61.0)
Sample size	702	364	746	375	672	293	2,096	209	385	140	645	471	336
Explained variance (R^2)	0.16	0.29	0.46	0.35	0.28	0.29	0.58	0.14	0.32	0.25	0.40	0.62	0.29

Note: — = not available; CHL = Chile; GER = Germany; MAL = Malaysia; MCK = mathematics content knowledge; NOR = Norway; PHI = the Philippines; POL = Poland; RUS = Russian Federation; SD = standard deviation; SING = Singapore; SWI = Switzerland; TAI = Taiwan, China; THA = Thailand; USA = United States.
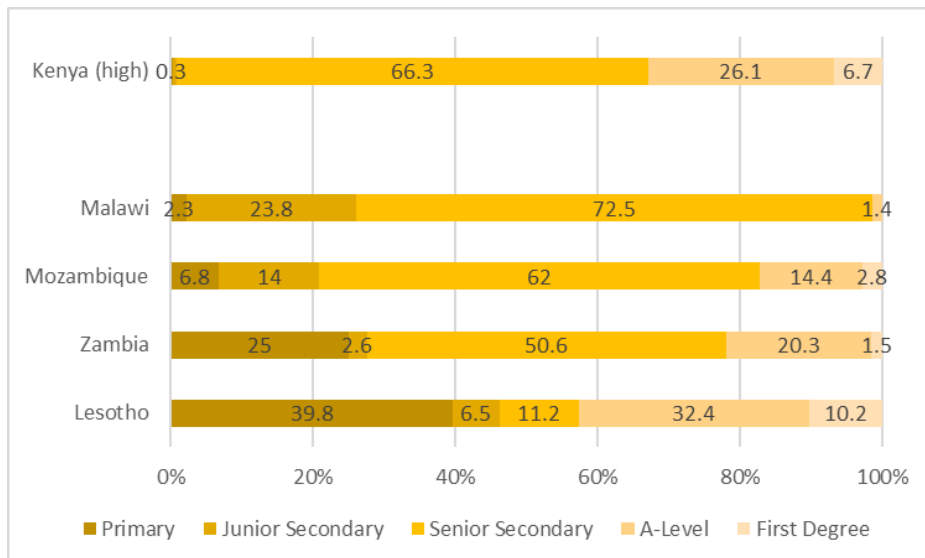*p <.05
**p <.01

Figure D.1. Highest Mathematics Coursework Grade Level by Secondary Teacher Trainee MCK Quintile (global)
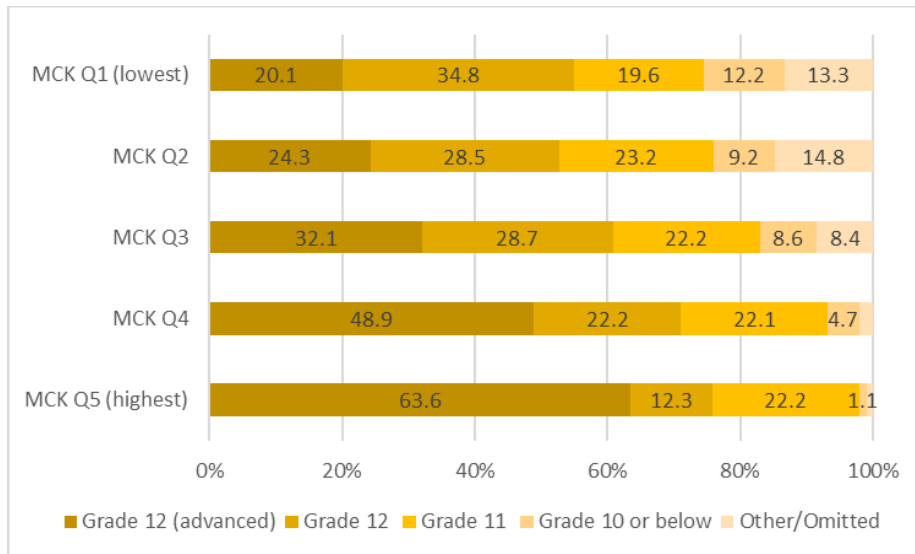
Source: Teacher Education and Development Study in Mathematics (2008).
Note: MCK = mathematics content knowledge.

Figure D.2. Teacher Education Level by Country, SACMEQ III

Source: Southern and Eastern Africa Consortium for Monitoring Educational Quality III.
Note: SACMEQ = Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Figure D.3 summarizes the results for the multivariate work with TERCE data from Latin America. The results are relevant for the screening argument because the indicators capture more selective features of teacher preservice preparation. Student achievement—at least in several countries—is significantly higher in classrooms with the following factors: teacher completed a training program (certification), attended a university, and is a subject specialist.

Nevertheless, these indicators are insignificant in the majority of the regressions and are negative (and significant) in a handful of cases.

Figure D.3. Teacher Variable Results in Grade Six Student Achievement Regressions, TERCE
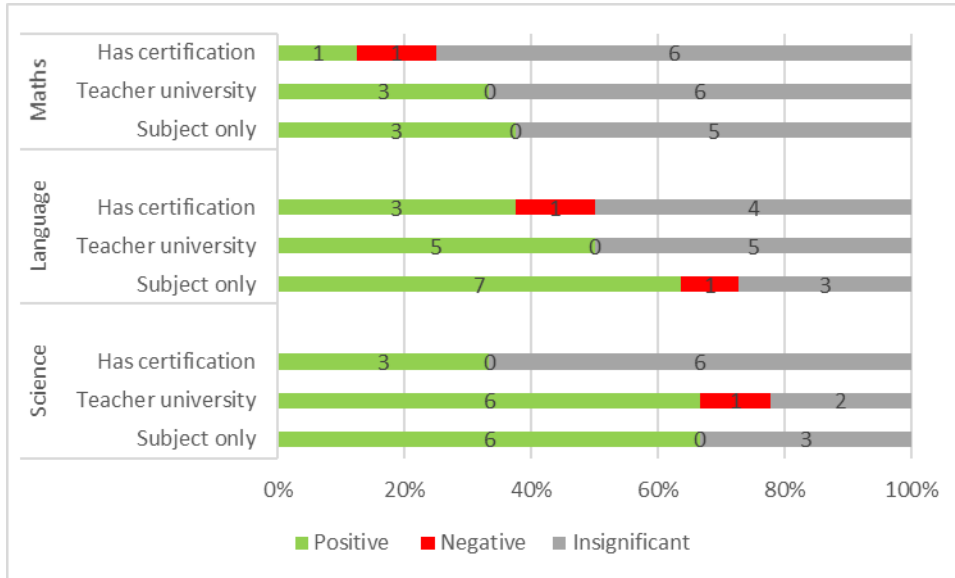
Source: Third Regional Comparative and Explanatory Study data (2013).
Note: Numbers refer to number of countries in each category based on student achievement regressions that include controls for student background. TERCE = Third Regional Comparative and Explanatory Study.

Figure D.4. Average Number of Training Institution Graduation Requirements by Level and Country, TEDS-M
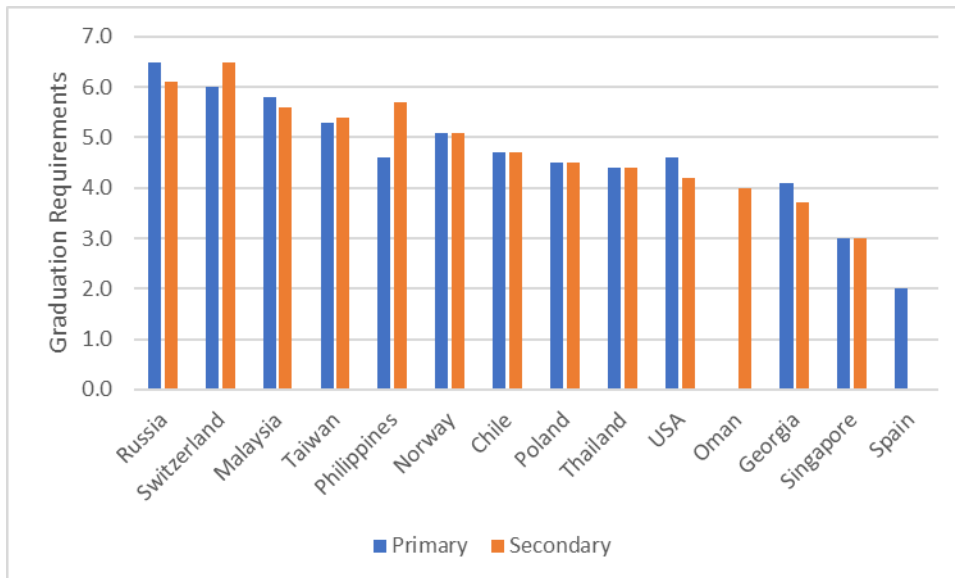
Source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

Graduation requirements are an additional screening mechanism that is captured in the TEDS-M data. Training institutions could select up to eight graduation requirements related to

89

classroom performance, different examinations, and things related to the practicum or teaching. Relatively high-performing countries or economies like the Russian Federation, Switzerland, and Taiwan have more than five requirements for graduating, which is consistent with the filtering and screening argument (figure D.4). However, Singapore has only three requirements. In the multivariate analysis, the number of graduation requirements was positively associated with student knowledge levels in five countries at the secondary level, although the results were more mixed in the primary level. Regarding individual requirements, receiving a passing grade in field experience and writing a thesis were the most consistently positive predictors of trainee mathematics content and pedagogical content knowledge, but this was true in only three or four countries. The positive effect for a thesis is similar to what Boyd and others (2008) found in New York City, where teachers who had completed a capstone project were associated with higher student achievement levels later on.

Coursework and Institutional Features

The TEDS-M data have extensive information on coursework. Table D.2 begins with a summary of coursework variables in the fixed effects multivariate analysis of primary and secondary teacher trainee MCK. University-level mathematics coursework is positively associated with primary-level trainee MCK in nine countries and is statistically significant in four of these. However, it is negatively associated (and significant) in two countries and insignificant in eight countries. The results for the other measures of coursework are mostly insignificant, with some positive and negative effects. The results do not vary much by level (primary or secondary) or dependent variable (content or pedagogical content knowledge).

The remaining results for coursework also provide little in the way of guidance regarding specific aspects of training that improve teacher trainee content or pedagogical content knowledge (in mathematics). Each block of variables offers some interesting individual results, but overall the findings fall far short of providing solid evidence about key training features.

The following is a summary of the additional variables that were analyzed, with individual findings that stand out.

- **University-level mathematics coursework, class by class.** The individual classes that are the most significant predictors of MCK include linear and abstract algebra, calculus, statistics and probability, and mathematical logic courses.

- **Mathematics pedagogy learning opportunities and methods.** For the mathematics pedagogy courses, the following activities were most frequently associated with higher pedagogical content knowledge: listen to a lecture, ask questions during class time, and participate in whole class discussion. Reading about mathematics research and writing mathematical proofs were activities most frequently negatively associated with pedagogical content knowledge.

Table D.2. Covariates of Teachers Trainee Math Content Knowledge: Teacher Education Program Coursework (Institutional Fixed Effects)

Independent Variable	CHL	TAI	GEO	GER	MAL	PHI	POL	RUS	SING	Oman	ESP	SWI	THA	USA	NOR
Primary Trainees															
Math (university level)	21.4	49.5**	−5.6	143.1*	17.3+	−22.5*	22.8*	−33.8*	17.9	—	8.0	8.2	−38.3	−16.0	10.4
	(1.42)	(3.23)	(−0.28)	(6.55)	(1.89)	(−3.53)	(2.04)	(−2.87)	(1.16)		(0.84)	(0.64)	(−1.30)	(−1.42)	(0.60)
Math curriculum	4.1	27.4**	11.5	118.2*	4.9	−0.8	7.1	−0.3	−3.3	—	−3.0	−4.9	−2.3	−36.3*	8.5
	(0.30)	(3.43)	(0.65)	(9.91)	(0.43)	(−0.06)	(1.38)	(−0.03)	(−0.18)		(−0.33)	(−0.60)	(−0.15)	(−3.56)	(0.33)
Math education/pedagogy	0.8	8.0	4.0	108.7*	6.5	−7.2+	−0.7	12.7	17.6	—	−10.1+	17.8	−21.1+	4.4	15.1
	(0.09)	(0.64)	(0.25)	(8.62)	(0.53)	(−1.68)	(−0.07)	(1.19)	(0.98)		(−1.76)	(1.43)	(−1.94)	(0.44)	(1.13)
General pedagogy	−5.7	3.3	−5.4	29.0**	−2.2	−2.7	0.03	−2.0	−2.6	—	−1.0	−3.6	−3.0	−5.7*	−0.9
	(−1.29)	(0.59)	(−1.06)	(4.84)	(−0.40)	(−0.41)	(0.01)	(−0.56)	(−0.34)		(−0.33)	(0.72)	(−0.62)	(−2.17)	(−0.16)
Education foundations	−6.8	2.7	−26.9	38.2+	−2.2	−21.1	1.3	3.7	−27.3+	—	−5.8	20.5	−5.2	4.6	−0.3
	(−0.37)	(0.20)	(−1.48)	(1.67)	(−0.16)	(−1.21)	(0.16)	(0.22)	(−1.70)		(−0.64)	(0.82)	(−0.30)	(0.47)	(−0.02)
Teaching for diversity	−5.1+	−8.9+	4.4	16.2*	−7.2+	−7.5	−7.0+	−3.3	−2.6	—	2.2	−6.9	−11.9*	−3.7	−11.6+
	(−1.74)	(−1.93)	(0.87)	(2.05)	(−1.82)	(−1.10)	(−1.81)	(−1.12)	(−0.43)		(0.81)	(−1.54)	(−3.03)	(−1.05)	(−1.86)
Secondary Trainees															
Math (university level)	44.0*	39.6	46.8+	—	17.8	4.3	−81.5*	132.0*	−7.7	−37.7	—	−60.1*	0.2	17.5	26.7
	(2.94)	(1.42)	(1.73)		(1.22)	(0.20)	(−2.20)	(2.37)	(−0.64)	(−0.69)		(−2.12)	(0.01)	(0.93)	(1.62)
Math curriculum	3.3	−22.2	−13.7	—	−3.6	10.7	−36.1	53.1*	−6.4	−4.9	—	−27.2	12.7	14.9	1.3
	(0.19)	(−1.28)	(−0.68)		(0.17)	(0.53)	(−1.09)	(2.45)	(−0.38)	(−0.15)		(−0.94)	(0.82)	(1.05)	(0.23)
Math education/pedagogy	28.9*	−11.8	1.6	—	2.2	−11.3	9.8	28.9	−22.2	−6.9	—	−11.3	2.8	−2.1	18.1
	(2.58)	(−0.78)	(0.12)		(0.22)	(−0.77)	(0.53)	(1.49)	(−1.15)	(−0.37)		(−0.54)	(0.21)	(−0.18)	(0.87)
General pedagogy	7.9	9.6	−4.8	—	−0.7	−3.0	−3.1	7.1+	−3.8	4.4	—	−5.5	0.02	−9.1*	5.0
	(1.19)	(1.46)	(−0.65)		(−0.17)	(−0.21)	(−0.28)	(1.79)	(−0.61)	(0.51)		(−0.63)	(0.01)	(−2.81)	(0.81)
Education foundations	−2.1	−8.9	27.4	—	−3.1	9.9	32.9+	−18.1	11.3	6.2	—	38.6	−13.6	8.0	−34.7*
	(−0.11)	(−0.50)	(1.30)		(−0.39)	(1.05)	(1.97)	(−0.91)	(0.83)	(09.42)		(1.58)	(−1.14)	(0.45)	(−2.13)
Teaching for diversity	3.8	−0.04	−8.2	—	−6.6+	−9.2	−12.4	3.7	−4.8	10.2*	—	−3.9	−5.3	−10.7*	−4.5
	(0.75)	(−0.01)	(−1.25)		(−1.70)	(−1.05)	(−1.63)	(1.29)	(−0.92)	(2.63)		(−0.38)	(−1.64)	(−2.32)	(−1.14)

Note: — = not available; CHL = Chile; GER = Germany; MAL = Malaysia; MCK = mathematics content knowledge; NOR = Norway; PHI = Philippines; POL = Poland; RUS = Russian Federation; SD = standard deviation; SING = Singapore; SWI = Switzerland; TAI = Taiwan, China; THA = Thailand; USA = United States.
*p <.05 **p <.01

General education learning opportunities and methods. The most consistently significant (positive) predictors of trainee outcomes include the following: build on pupils' existing mathematics knowledge skills (MPCK), create learning experiences that make the central concepts of subject matter meaningful to students (MPCK), explore how to apply mathematics to real-life problems (MCK and MPCK), explore how to use manipulative materials or physical models to solve mathematics problems (MCK and MPCK), locate suitable curriculum materials and teaching resources (MPCK), and integrate mathematical ideas from across areas of mathematics (MCK and MPCK).

Within this same block, a number of activities are also somewhat consistently negative (and significant) predictors of trainee knowledge outcomes, including analyze pupil assessment data to learn how to assess more effectively (MCK), deal with learning difficulties so that specific pupil outcomes are achieved (MCK), help pupils to learn how to assess their own learning (MCK and MPCK), use assessment to give effective feedback to parents or guardians (MPCK), and use standardized assessments to guide your decisions about what and how to teach (MCK and MPCK). The negative relationship between assessment-related training and trainee learning is an interesting result that, at the very least, suggests that trainee capacity development is slower in programs that are more active in using assessment to guide teaching activities.

This review takes a broad view on coursework and includes characteristics of the training program that affect the delivery of coursework, including overall program length. The TEDS-M data show that high-performing countries (or economies) do not always have the longest programs (figure D.5). Taiwan's training programs are all 50–60 months (4 or more years) long, but in Singapore, a sizable proportion of trainees are trained in programs of 18–24 months, which is shorter than the average program in Chile and the Philippines.

Figure D.5. Length of Preservice Preparation Programs in Select Countries, TEDS-M
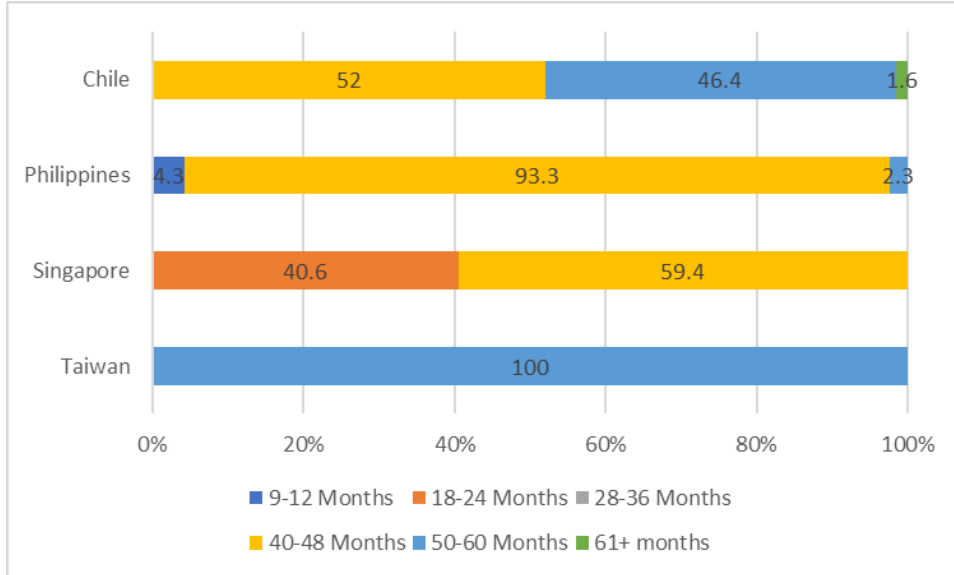
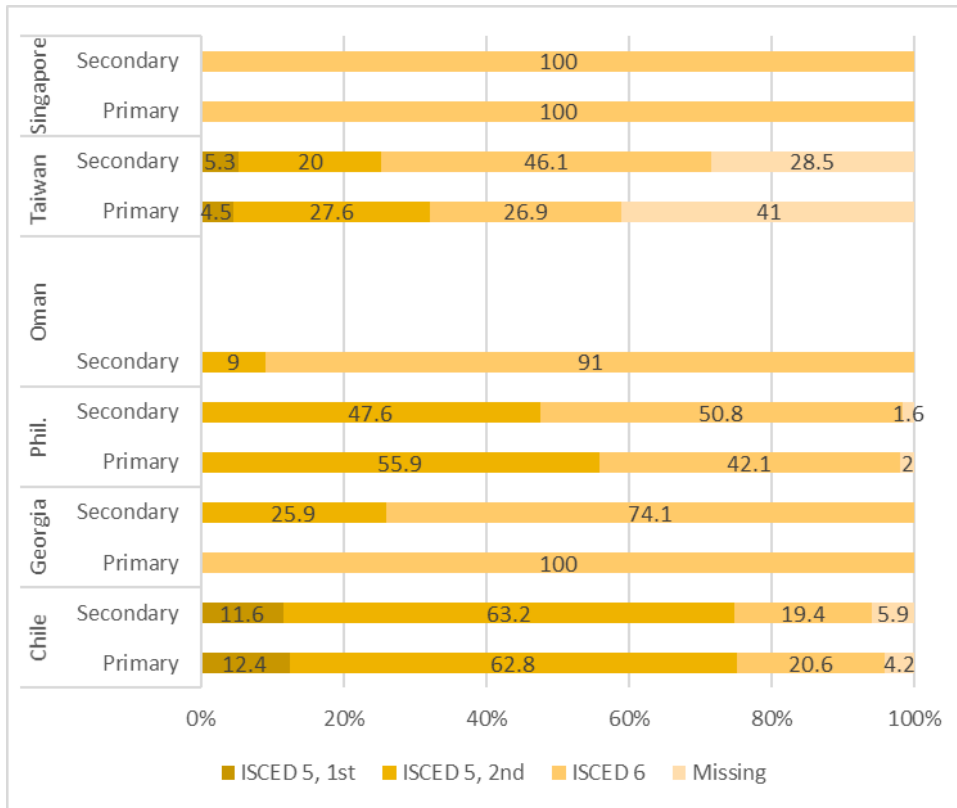Data source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

Statistical guidance on length is mixed. In four of the TEDS-M countries, primary teacher trainee content knowledge is significantly higher in programs that are longer, but for pedagogical content knowledge and secondary education, the results are more mixed, with significant negative associations in two countries. The TERCE data also have information on preservice training length. The statistical results are also somewhat uneven, but in three to five countries, there is evidence that longer preparation programs are associated with high levels of student achievement.

Most teacher training institutes in TEDS-M countries require their mathematics content course instructors to have an International Standard Classification of Education 6 level degree (figure D.6).

In Singapore, this is 100 percent for both primary- and secondary-level training. However, in Taiwan, a sizable proportion have the five level credential requirements. Additionally, some of the lower-scoring countries (Georgia and Oman) have relatively high standards for mathematics content teachers. There is also a lot of variation related to teaching credentials, experience and joint appointments in training institute, and actual school (figure D.7). In Taiwan, the mathematics pedagogy educators are less likely to have teaching credentials or experience than those in Singapore, and there is no clear pattern when comparing high- and low-scoring countries.
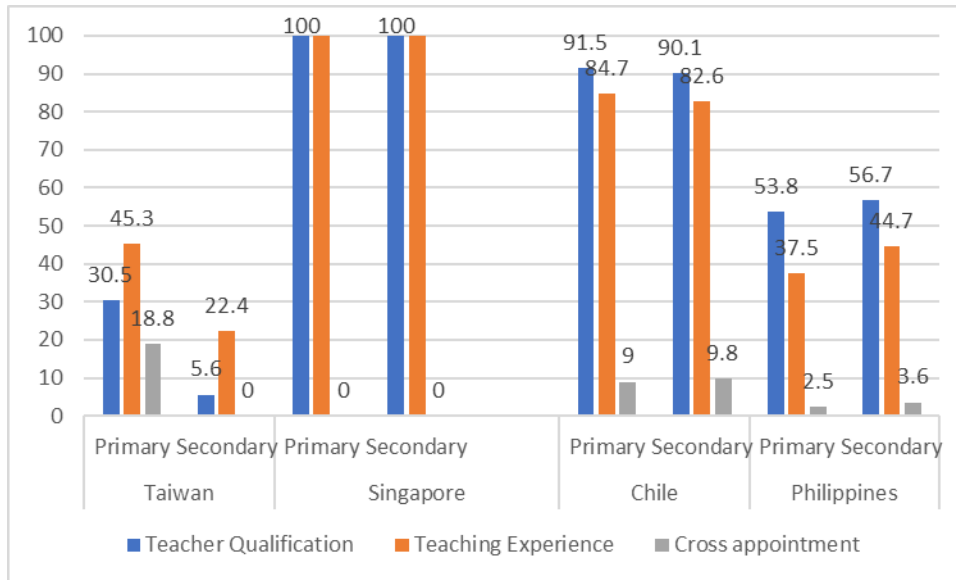
Figure D.6. Mathematics Content Educator Credentials by Country, TEDS-M

Source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

The multivariate results provide only a little guidance on this question of effective teacher educator credentials. For content knowledge, there is some evidence that higher-educated mathematics content course educators produce higher levels of MCK among trainees: The mathematics credentials were positively associated with trainee content knowledge in four countries in primary and secondary level, but they were negative (and significant) in two countries in primary. For pedagogy course instructor credentials, there are even fewer significant results, which includes education level and teaching credentials and experience.

Figure D.7. Mathematics Pedagogy Teachers with Teaching Qualification, Teaching Experience, and Cross Appointment in Schools by Country, TEDS-M (percent)
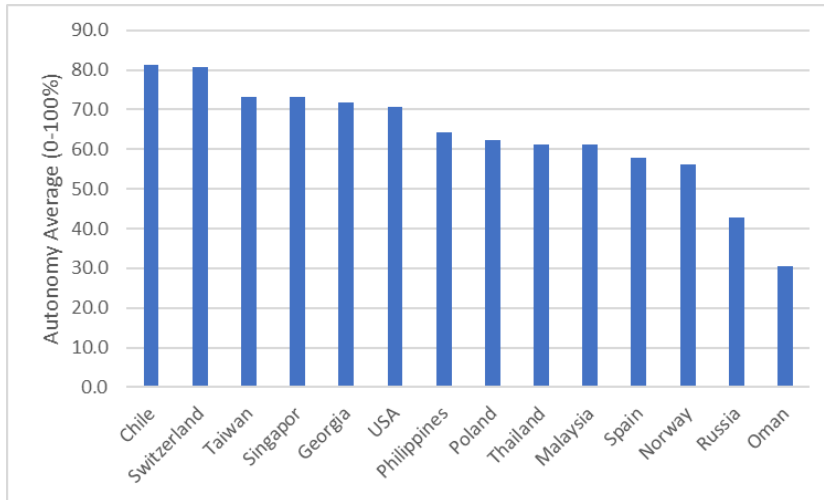

Source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

The TEDS-M data include a table of questions related to institutional autonomy. The literature offers no guidance on this topic because autonomy can either allow schools to adapt to needs of students and provide a better experience or lead to low quality when capacity is low and there is not much oversight. Training institute control over the following decisions was positively (significant) associated with trainee knowledge levels in at least five of the countries: program goals and emphasis (MCK), standards of classroom performance expected of graduates (MCK), mathematics pedagogy curriculum (MPCK), and the liberal arts curriculum (MCK and MPCK). Overall autonomy (average control across 15 aspects) was positive and significantly related to content knowledge in seven countries and five countries for MPCK; however, average autonomy was negative (and significant) in three countries for the MCK outcome and one country for the MPCK measure.

The results from TEDS-M are only suggestive regarding autonomy, and this is clearly a variable where interaction is likely (especially with local capacity and systemic oversight). Nevertheless, it is noteworthy that, on average, institutions that report having more control over features of teacher training tend to have better outcomes for teachers. Additionally, there is some evidence that teacher training institutions in higher-scoring countries have more autonomy (figure D.8). Taiwan and Singapore have two of the highest autonomy averages overall in TEDS-M. However, the pattern is not

very clear because Chile (low-scoring) also has high autonomy while relatively high-scoring countries like Norway and the Russian Federation report much lower averages.

Figure D.8. Average Autonomy (Percentage of Decisions Made by Institution) by Country, TEDS-M

Source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

The practicum. The first key feature of the practicum experience is simply having access to this training resource. Figure D.9 shows that among TERCE countries in Latin America, there is a lot of variation in the percentages of teachers who report a practicum as part of their training, and these results are not much different when restricted to new teachers (less than five years of experience). The standouts are the extremely low practicum participation rates in some of the Central American countries.

Figure D.9. Grade Three and Six Teachers Reporting Practicum Experience, TERCE (percent)
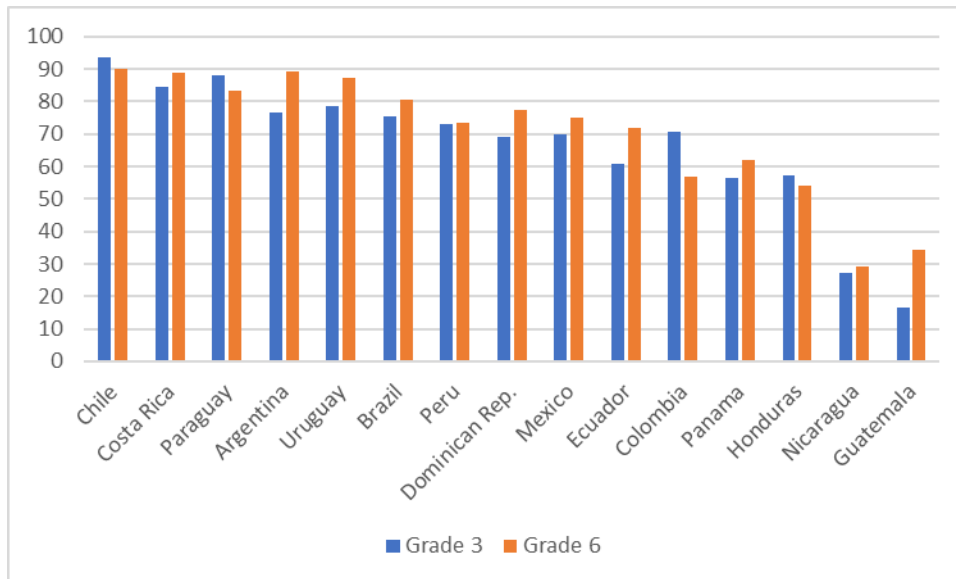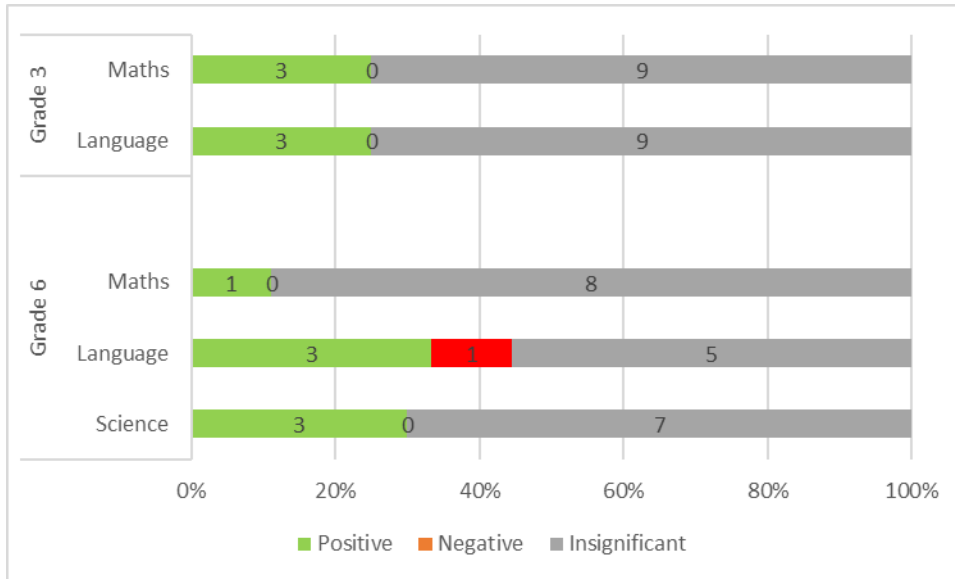

Source: Third Regional Comparative and Explanatory Study data (2013).
Note: TERCE = Third Regional Comparative and Explanatory Study.

Among the TEDS-M participants, the practicum is nearly universal, but there is a lot of variation in terms of length. High-performing Taiwan has a very long practicum period, but so do lower-performing countries like the Philippines and Thailand. The multivariate analysis also looked at this question within each country to see if teacher outcomes (namely pedagogical content knowledge) were affected by length of the practicum experience. In both the primary- and secondary-level training, there were three countries where longer time in practice was significantly associated with higher pedagogical content knowledge. However, in two countries, the measures of practicum length (especially the Introductory Teaching form) were negatively associated with pedagogical content knowledge.

Figure D.10 summarizes the multivariate results for teacher practicum in TERCE. In several countries (by grade and subject), there is evidence that teacher participation in practicum leads to better student achievement levels. This was analyzed in a separate analysis for just newer teachers (< 5 years of experience), and the results suggested larger effects in some countries but with more variability.

**Figure D.10. Teacher Practicum Results in Student Achievement Regressions by Grade and Subject, TERCE**

Source: Third Regional Comparative and Explanatory Study data (2013).
Note: TERCE = Third Regional Comparative and Explanatory Study.

In addition to length of the practicum, a key feature identified in the literature is the degree of support during the practicum. The main concern is often referred to as "sink or swim," an approach in which teaching students are placed into situations with little supervision or support. Figure D.11 summarizes the respondents from high- and low-performing countries (by level of study) for the question of what percentage of time they alone were in charge of the class. Note that in Georgia (primary only) and the Philippines, about 6 percent of trainees report not having access to any practicum. In Taiwan, the teacher trainees report relatively low percentages of time in charge, which suggests that they spend a lot of time with direct supervision during the lengthy periods they report in practicums. However, for the rest of the countries, the trainees are much more likely to report being in charge of the teaching activities. This is especially true in Singapore, which has a different profile than Taiwan does on this indicator.

Figure D.11. Percentage of Practicum Time Teacher Trainee Was in Charge of Teaching Class by Study Level and Country or Economy, TEDS-M
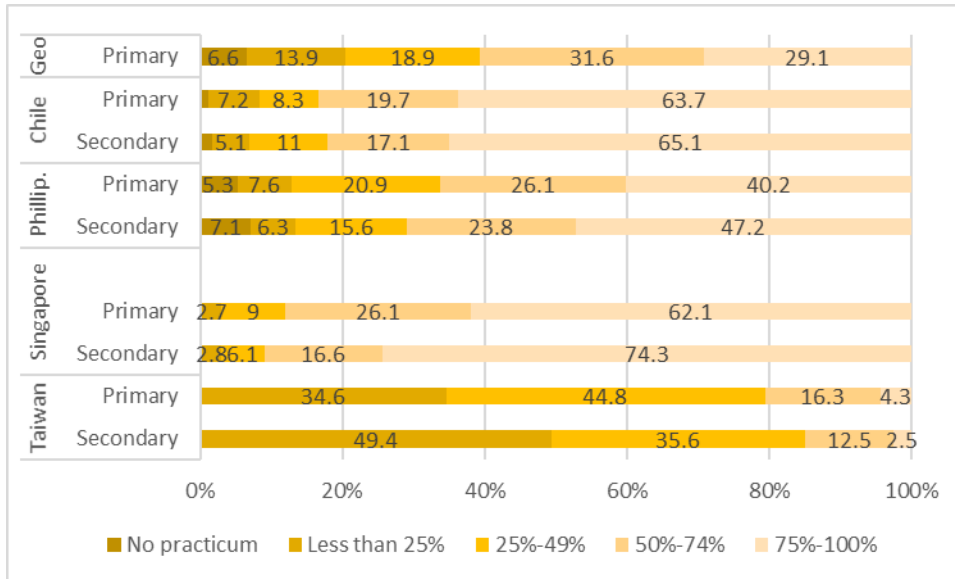
Data source: Teacher Education and Development Study in Mathematics data (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

Figure D.12 continues with an institutional-level variable for how frequently the institution expects teacher trainees to be observed during their practicum teaching experience. Once again, Taiwan is an outlier, with a very high percentage of training institutions (85 percent) reporting that their trainees are observed every day. This is again notable given the large number of hours that trainees report working in practicums and suggests that the practicum in Taiwan is closely supervised. In Singapore, the goal is instead to supervise teachers at least once a week. In the other countries, the supervision is more varied across the various institutions. In Chile and the Philippines, a sizable proportion reports daily supervision (the Philippines) or once per week (Chile), but there are also institutions that follow-up on practice teachers only every 2–3 weeks or even just once a month.

Figure D.12. Frequency Teacher Trainee is Observed by Teacher Educator/Supervisor During Practicum by Study Level and Country or Economy, TEDS-M
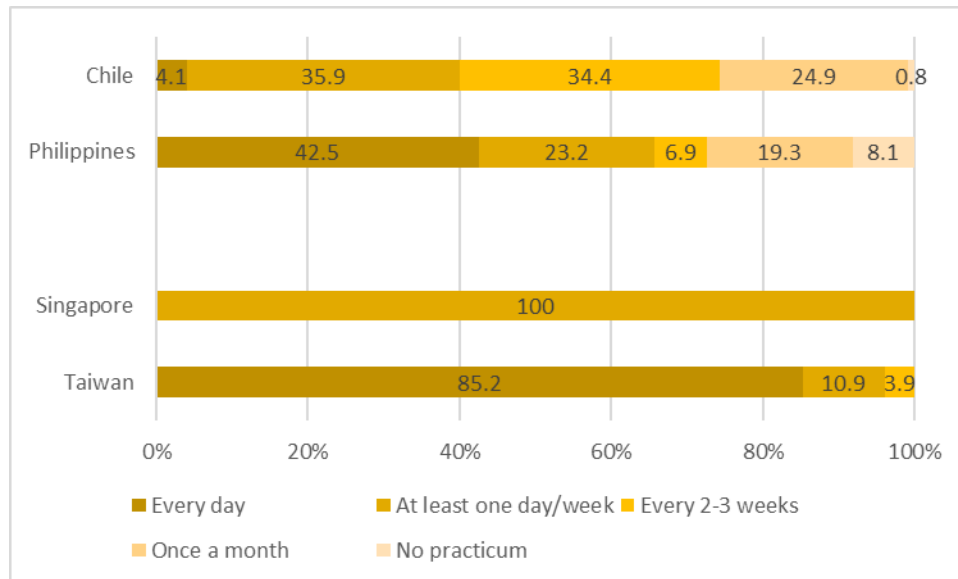

Source: Teacher Education and Development Study in Mathematics (2008).
Note: TEDS-M = Teacher Education and Development Study in Mathematics.

In addition to basic measures of length, supervision, and autonomy, the TEDS-M data asked trainees about the practicum experience itself. Many of these indicators were included in the multivariate analysis of pedagogical content knowledge (one by one) to test for practicum experiences that stand out as significant predictors of MPCK. Once again, the analysis yields little insight into best practice. One result that stands out is that in eight of the 14 countries, the degree to which trainees reported using different methods in their practicum from what they were taught in class is negatively associated with pedagogical content knowledge (and significant). Additionally, the most consistently positive predictor of pedagogical content knowledge (significant in four countries) was the frequency trainees reported having to demonstrate that they could apply teaching methods learned in coursework in their actual classes. This is a useful counterpart result to the finding that inconsistency in these two dimensions is negatively associated with pedagogical content knowledge. Overall, the message is consistent with the belief that close-on links between what is learned in classes and what is applied in the practicum (with supervision) are necessary for developing skills.

Overall, the message from this line of inquiry is that, despite the unprecedented detail for an international data set, the TEDS-M indicators for the practicum experience are not likely to capture the core qualitative features related to the support that teachers receive and the degree to which the practicum experience is structured in a way that builds capacity. This is similar to a black box challenge that researchers face when trying to unpack classroom processes that are effective or ineffective.

References

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Teacher Preparation and Student Achievement." National Bureau of Economic Research (NBER) Working Paper 14314, NBER, Cambridge, MA.

Goldhaber, D. 2019. "Evidence-based teacher preparation: Policy context and what we know." *Journal of Teacher Education* 70:2: 90–101.

Harris, Douglas N., and Tim R. Sass. 2011. "Teacher Training, Teacher Quality, and Student Achievement." *Journal of Public Economics* 95 (7–8): 798–812.

Shulman, Lee S. 1986. "Those Who Understand: Knowledge Growth in Teaching." *American Educational Research Association* 15 (2): 4–14.

Tatto, Maria Teresa, ed. 2013. *The Teacher Education And Development Study In Mathematics (Teds-M): Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries.* Amsterdam: International Association for the Evaluation of Educational Achievement.

Appendix E. Scaling Theory of Change

The literature identifies different types of scaling that might be pursued, often in parallel, including horizontal scaling that focuses on the breadth of coverage of an intervention; vertical scaling, which involves a deeper embedding of the scaling up process within the policy making and implementation system; and functional scaling that pertains to the expansion of the type of activities or areas of engagement—for example, expanding the range or level of subject matter offered in existing training or including functional aspects of the education system (Perlman and others 2016). The type of scaling effort pursued has significant implications for what is done regarding strategy and priorities (planning), resourcing (implementing), and monitoring and evaluation (M&E). Thus, all of these components may not be required to the same degree for other, less complex forms of scaling that focus more on elaboration of content (for example, functional scaling) or enlargement (horizontal scaling) without seeking to affect system-level change. It is necessary to know from the start which type of scaling up is intended so that steps are taken to support scaling potential at the planning and design stage, and through implementation and monitoring and evaluation.

The theory of change (figure E.1.) derived from the literature and case studies presents conditions under each of the three phases—planning, implementing, and M&E—with reference to the two types of scaling (vertical and horizontal), which were examined in the case studies. Given the dynamic, process-driven nature of scaling, aspects of the theory of change (such as communication, political support, and M&E) span the entire process but are discussed under specific stages in the next section.

Figure E.1. Theory of Change

Planning for scale up	Implementation of scale up	M&E	Outcomes
Relevant stakeholders consulted on scale up	Regular, ongoing communication with key stakeholders	M&E systematically deployed to support its viability	Vertical scale up
Scale up plan developed:	Political support	M&E used to provide a feedback loop for adaptation	Horizontal scale up
Clear sequencing	Influential champions	Training quality monitored	
Financial planning beyond initial (project) period	Sufficient financing to support the intended scale up	Anticipated number/level of beneficiary related outputs achieved	
Plan includes M&E	Financing likely to support sustainability	Training program evaluated	
	Adequate logistical support and human resources		

ASSUMPTIONS
Identified need, clarity regarding nature of scale up (how much, for how long, and in what direction), continuity of support (political, financial), long-term commitment

Planning Scaling

The need for consultation and communication with stakeholders spans the entire scaling process. The literature emphasizes the need for ongoing communication and coordination among stakeholders (Hardee and others 2012). In examining the experience of bringing Escuela Nueva to scale in Colombia, Colbert and Arboleda (2016) emphasizes the importance of close interaction with teachers and students (beneficiaries) who, the authors suggest, need to be the key actors of the change.

Among the areas of weakness identified in the design of scaling is the lack of a long-term target or plan beyond the life of the project or program, effectively deactivating an important potential driver of scaling (Begovic and others 2017). Hassler, Hennessy, and Hofmann (2018) point to the need to plan for scaling from the outset and argue that scalable, sustainable, and effective models for in-service teacher training are required. To meet the need for large-scale, systemic, and ongoing development opportunities for teachers, programs need to explicitly focus on scaling and sustaining while maintaining effectiveness, and need to do so during the life of the operation. The difference between failure and success in large-scale education system improvement is associated with the choice of drivers of reform that inform the intervention. In that regard, the "wrong" drivers in interventions focus on individualism, technology, punitive accountability, and fragmented policies. The "right" drivers associated with planning for scaling include capacity building, pedagogy, and systemic policies—drivers that support ongoing depth and scale (Fullan and Quinn 2016).

Implementing Scaling

The need to secure political support or influential champions throughout the scaling up process is heavily emphasized in the literature. Scaling up is characterized as a dynamic process that requires a force, or driver, to provide momentum. Once the innovative idea has been formed and demand identified, a leader or champion is needed (Hartmann and Linn 2007). The literature also suggests the need for both deep and wide systemic support for gains to be sustained and scaled up. Innovative programs are often required to show quick gains while also devoting energy to building trust and capacity among key stakeholders in the system whose support will be required to sustain the innovation beyond a pilot phase (Christina and Vinogradova 2017).

Sufficient financial support to execute the planned scaling phase is necessary, as is sufficient human resource input to support logistics and delivery. Successful scaling requires careful balancing of desired outcomes with practical realities and constraints. Along with taking care of current realities in that regard, attention needs paid to

securing financial support for the future (sustainability), an added burden for promoters of the scaling effort (WHO 2010).

Monitoring and Evaluation

The greatest degree of emphasis in the literature is on the need for systematic evidence, to include monitoring, evaluation, and research. The question of the scalability of an intervention—the capacity of an individual intervention to be scaled up—is underpinned by the quality of the evidence to support its claims (Milat and others 2016). The literature asserts the need for rigorous evaluation (external, and mixed methods, including experimental and quasi-experimental research designs) to add to the knowledge base and that scaling strategies should have robust M&E systems (clear indicators of progress, systems to track service delivery, and agreed-on outcomes) linked with stakeholders, noting that good M&E can promote accountability, transparency, and ownership of policy initiatives (Hassler, Hennessy, and Hofmann 2018; Hardee and others 2012).

The literature recommends a more systematic operational approach to scaling up supported throughout the process by robust M&E that provides feedback to allow for adaptation, as necessary, in the design of the intervention and of the scaling up (Begovic and others 2017). For example, regarding the assessment of training, Nielsen's (2013) review of the scaling of the Early Grade Reading Program in the Arab Republic of Egypt identifies many lessons, including the importance of science, that is, the credibility of the cognitive scientific framework that the project lent valuable credibility.

Related to the systematic use and availability of evidence is the need to demonstrate effectiveness, which can attract partners and build support for innovation. The literature emphasizes the need for feedback into the policy process to test and determine what works and to identify necessary additional reforms in operational policies. At the same time, there is a need to emphasize the importance of simple steps in innovation and early and easily demonstrable gains, noting that innovative programs are often required to demonstrate quick gains while promoters also have to expend energy on building stakeholder support, trust, and capacity. Stakeholder support will be required to sustain the innovation beyond the early or pilot phase or phases (Hardee and others 2012; Nielsen 2013; Christina and Vinogradova 2017). Among the lessons identified by Colbert and Arboleda (2016) in examining the experience of bringing Escuela Nueva to scale in Colombia is that innovations have to be easily replicable within existing conditions, and the attitudinal change of teachers is positively affected through demonstrating that the model worked and was a good fit for its beneficiaries.

References

Begovic, Miliça, Johannes F. Linn, and Ratislav Vrbensky. 2017. "Scaling Up the Impact of Development Interventions: Lessons from a Review of UNDP Country Programs." Global Economy and Development Working Paper 101, Brookings Institution, Washington, DC.

Christina, Rachel, and Elena Vinogradova. 2017. "Differentiation of Effect across Systemic Literacy Programs in Rwanda, the Philippines, and Senegal." *New Direction of Child Adolescent Development* March (155): 51–65.

Fullan, M., and J. Quinn. 2016. *Coherence: The Right Drivers in Action for Schools, Districts, and Systems.* Ontario, Canada: Corwin Press.

Hardee, Karen, Lori Ashford, Elizabeth Rottach, Rima Jolivet, and Rachel Kiesel. 2012. *The Policy Dimensions of Scaling Up Health Initiatives.* Washington, DC: Futures Group, Health Policy Project.

Hartmann, Arntraud, and Johannes F. Linn. 2007. "Scaling Up: A Path to Effective Development." 2020 Focus Brief on the World's Poor and Hungry People, International Food Policy Institute, Washington, DC.

Hassler, Bjoern, Sara Hennessy, and Riikka Hofmann. 2018. "Sustaining and Scaling Up Pedagogic Innovation in Sub-Saharan Africa: Grounded Insights for Teacher Professional Development." *Journal for Learning and Development* 5 (1): 58–78.

Milat, Andrew J., Robyn Newson, Lesley King, Chris Rissel, Luke Wolfenden, Adrian Bauman, Sally Redman, and Michael Giffin. 2016. "A Guide to Scaling Up Population Health Interventions." *Public Health Research and Practice* 26 (1): e2611604.

Nielsen, H. Dean. 2013. Going to Scale: The Early Grade Reading Program in Egypt: 2008–2012. Case Study Report for the U.S. Agency for International Development, H. Dean Nielsen, Maplewood, NJ.

Robinson, Jenny Perlman, Rebecca Winthrop, and Eileen McGivney. 2016. *Millions Learning: Scaling Up Quality Education in Developing Countries*. Washington, DC: Brookings Institution.

WHO (World Health Organization). 2010. *Nine Steps for Developing a Scaling Strategy.* Geneva: WHO.

Burns, Mary. 2014. "The Myths of Scaling-up: How Misconceptions about Scaling-Up Can Hurt High-Quality Implementation." *Global Partnership for Education* (blog), January 14. https://www.globalpartnership.org/blog/myths-scaling.

Frake, April N., and Joseph P. Messina. 2018. *Toward a Common Ontology of Scaling Up in Development.* ResearchGate. https://www.researchgate.net/publication/283417477_Towards_responsible_scaling_up_and_out_in_agricultural_development_An_exploration_of_concepts_and_principles.

Appendix F. Conditions for Scaling in the Case Studies

Conditions for Planning of Scaling

Table F.1 shows how the analysis of the six in-service training cases map onto the conditions for planning of scaling in accordance with the theory of change.

Table F.1. Presence of Conditions—Planning of Scale-Up in Case Studies

	All Relevant Stakeholders (including beneficiaries) consulted		Scale-Up Plan Developed		Scale-Up Plan Includes					
					Clear sequencing		Financial planning beyond initial (project) period		Robust monitoring and evaluation	
Condition present	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Cases (no.)	4	2	4	2	3	3	1	5	4	2

Broad-based consultation with stakeholders was evident in all cases, but consultation with teachers was a notable omission in Vietnam. Consultation in relation to the content and purpose of in-service training and about the ongoing scaling process was particularly strong in Ghana and Uruguay. In Uruguay, in-service training is demand driven and supported by ongoing consultation between and among the ministry, the Council of Initial and Primary Education, the Central Board of Directors of the National Administration of Public Education, teachers and principals, trainers, inspectors, and teacher unions. In Ghana, consultation is inclusive and involves ministry officials, teachers, technical teams, bilateral donors, other donors, and academic institutions, among others. Feedback was solicited from teachers and parents through forums in the relevant communities. There was some level of consultation with most stakeholders in Vietnam; however, teachers were not consulted directly regarding training content or the scaling process. Parents were informed about the rollout of new pedagogic approaches, particularly under the Vietnam Escuela Nueva Project (VNEN), but they were not consulted.

In most cases, in-service training was supported by a scaling plan (that is, logistics for progress and rollout backed by adequate resources) that contributed to broadly successful scaling within the project cycle. Under the VNEN, a carefully designed scaling-up plan was developed after the initial pilot that ring-fenced support for 1,447 schools supported by a clear plan regarding identification, rollout, and associated supports. The approach also anticipated voluntary uptake by nonproject schools, and

this materialized. In Ghana, an explicit scaling plan was developed for the Untrained Teachers Diploma in Basic Education (UTDBE) training that built on what the pilot had achieved. The UTDBE scaling was premised on increasing engagement with teacher colleges, improved certification requirements, and data on the numbers of unqualified teachers. The in-service training in Ghana was not supported by an explicit plan but was built out of existing commitments to in-service within schools. Expansion of in-service training in Ghana is ad hoc and linked to a variety of donor-funded grant projects that are intending to improve teaching capacity. There is no real strategy as to how a larger professional development and in-service model will be instituted in the long term.

There was little evidence of planning beyond the immediate project cycle. In Uruguay, in-service training has been institutionalized for about 20 years and has been supported by successive governments. In Vietnam, the School Readiness Promotion Project (SRPP) provided budget support for training expenses reimbursed as delivery linked indicators based on the number of people trained, but there was no planning beyond the immediate training provision. Under the VNEN, a late and unsuccessful effort was made to secure funding from the Global Partnership for Education for a follow-on project to help further mainstream the innovation; however, this was not accompanied by a planned budgetary commitment beyond the project, despite the promoters' aspirations. Funding for in-service training programs in Ghana is built into budget calculations, but the district offices have struggled to maintain momentum because funds have been more difficult to attain. Regarding in-service training at the secondary level in Ghana, there was no evidence of a careful approach to financing of the particular type of training in the longer term.

The rollout of in-service scaling was well sequenced in all cases. The SRPP had a clear plan for the rollout of an expert to deliver Hanoi-based training to provincial staff. The next level of the cascade was then a matter for the provincial authorities, although provincial capacity was not homogenous, and more tailored support may have been merited in certain instances. The VNEN was supported by the experience of a pilot project, and the sequencing of the scaling was informed by that experience using an adapted cascade model that ultimately led to a greater number of teachers trained than originally anticipated. In Uruguay, each training cycle has a carefully managed two-year duration—during the first year, the targeted schools receive training in math and sciences, followed by training in language and social sciences in the second year. In Ghana, there was a commitment to delivering UTDBE in line with demand and while it filled a clear gap (large proportion of unqualified teachers in the system)—the training was ultimately stopped given the aim (a dramatic increase in the number of qualified teachers now looking for teaching positions) was attained.

Conditions for Implementation of Scaling

Table F.2 shows how the analysis of the six in-service training cases map onto the conditions for implementation of scaling as per the theory of change.

Table F.2. Presence of Conditions—Implementation of Scaling

	Regular, Ongoing Communication with Stakeholders		Political Support or Influential Champion		Sufficient Financing to Support Initial Scaling Plan		Financing Likely to Support Sustainability		Adequate Logistical Support		Human Resources Available for Training Program and Follow-Up	
Condition present	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Cases (*no.*)	4	2	6	0	6	0	1	5	6	0	5	1

Noting the previous discussion of communication—which spans the scaling process—in-service training scaling efforts that the World Bank supported also tended to be underpinned by significant stakeholder backing throughout the process. In all cases, the scaling up efforts had institutional support and, in certain instances—most notably with reference to the VNEN project and support for in-service training in Uruguay—influential champions who supported the scaling. Because scaling up can require significant mobilization of resources and logistical planning, it is necessary to secure reliable support to see the effort through. The support of influential champions can be particularly important where there are ambitions to progressively scale up and where vertical scaling is desired. For most of the in-service scaling efforts assessed here, that type of support was not critical, perhaps because all of the in-service training was in tune with relevant government policy and met an identified need about which there was little argument, that is, key stakeholders were already committed to change. For example, teachers had been recruited into early childhood education in Vietnam to meet increased enrollment. Many of these were inadequately trained during preservice, and the in-service training sought to address some of that deficit. In Ghana, the UTDBE was designed to fill the gap in formal certification for many teachers in isolated areas where filling positions proved a challenge. The VNEN is perhaps the exception in that the approach adopted to meeting the agreed, identified need—to reform pedagogy at primary level so that it was more child centered—was quite radical in context and had been adopted from a completely different culture (Colombia). In that instance, the innovation required (and had) strong and influential champions, but despite that support, no financing was available for the next stage of the scaling effort.

Although planning included enough financial support to successfully implement scaling within the confines of the various projects, the lack of planning for further scaling up meant no ongoing resources with which to sustain most of the various efforts. The issue of forward planning was not relevant in the case of UTDBE (Ghana) because most of the demand was met after the scaling up of the pilot. However, in other instances, the end of the project supporting the scale-up effectively meant the end of the in-service training, at least in terms of planned, ongoing, structured provision. The Ministry of Education in Vietnam sought additional funding from the Global Partnership for Education late in the project cycle for furthering the VNEN scaling but was unsuccessful in obtaining the additional funding. In-service under the SRPP ceased with project closure, and financing was not available to continue face-to-face training or to refine e-learning introduced during the project. Expansion of in-service training in Ghana is ad hoc and linked to a variety of donor-funded grant projects that are intending to improve teaching capacity, and there is no real strategy as to how a larger professional development and in-service model will be instituted in the long term. The situation in Uruguay was different because funding for in-service benefited from government and policy support across time.

In-service scaling was well supported by adequate logistical efforts in all cases and by input of appropriate levels of human resources in most cases. In Vietnam, the rollout of in-service provision to almost all early childhood education teachers in Vietnam was smoothly executed, as was the rollout of the VNEN, where a modified cascade approach was adopted. In Uruguay, in-service training was originally developed in adequate venues and supported by the provision of relevant materials. The more recent adoption of school-based modality has simplified the logistical requirements and responded to teachers' concerns. In Ghana, the rollout of in-service was smooth, organized, and well-designed. The human resource aspect was equally well managed in almost all cases. For example, under the VNEN, international experts were available to support core teachers and were frequently active at the local level. In Uruguay, the move to school-based training was accompanied by the need to expand the pool of trainers to cover half of full-time school in the first year—more trainers were trained to meet the need. In Ghana, specialized trainers were made available to deliver the UTDBE, while in-service training relied on the available resources and approaches from the district offices. To deliver training to second level teachers, overall design allowed for engaging relevant teacher colleges, including those with outreach and in-service experience. There was some underprovision evident regarding the SRPP, including limited follow-up support for training.

Monitoring and Evaluation

Table F.3 shows how the five in-service training programs performed in monitoring and evaluation (M&E).

Table F.3. Conditions for Monitoring and Evaluation in Scaling Up

	M&E Systematically Deployed to Support Viability/Success, and for Feedback/Adaptation		Training Quality is Monitored		Anticipated Number/Level of Beneficiary Related Outputs Delivered		Training Program is Evaluated	
Condition present	Yes	No	Yes	No	Yes	No	Yes	No
Cases (no.)	2	4	1	5	6	0	2	4

Note: M&E = monitoring and evaluation.

There are mixed findings regarding the extent to which M&E was used systematically to support scaling up through the generation of evidence to support sustainability or ensuring the quality of training. Impact evaluations were associated with the two scaling up efforts (UTDBE and VNEN), that is, scaling up based on lessons from an earlier pilot that was evaluated to generate lessons to inform implementation at scale. M&E was used extensively and applied for the UTDBE training in Ghana. An impact evaluation collected extensive information that at various times informed changes made to the approach; an impact evaluation undertaken for the VNEN project (published postproject) measured cognitive and noncognitive improvement in children exposed to the VNEN by teachers trained under the project. Core monitoring and evaluation under VNEN was basic (largely focused on numbers trained and number of participating schools) and, M&E was not used to provide a feedback loop for adaptation.[1] In Uruguay, data on teachers' enrollment and completion was monitored, and some qualitative feedback from teachers was collected, but there were no systematic M&E arrangements. Under the SRPP, numbers of teachers trained are tracked, but M&E was not used for feedback purposes. In Ghana, the numbers of teachers trained and the number of in-service training courses run were tracked, but M&E was not used to support success or strategically improve the nature and approach for the training.

All the monitoring systems for the various projects supporting in-service training could produce data to support basic indicators such as numbers trained. For example, it is known that under the SRPP in Vietnam, 97 percent of all preschool teachers were trained, and teachers in 1,447 schools received in-service training and associated supports under VNEN. In Ghana, the UTDBE supported in-service training (along with accreditation and certification) for most of the target group.

Evidence of assessment of the quality of in-service training was limited. Even where there is no ambition for further scaling up, the assessment of quality is important for future learning. For example, under the SRPP, the fidelity of implementation of the cascade approach was not monitored to ensure homogenous and quality training. In Uruguay, there are no regular assessments of teacher practice. Teacher satisfaction with the training and other informal feedback is collected, but there is no direct, systematic quality assurance in place. In Ghana, there was a concerted effort to formulate quality standards for the tertiary teaching courses, but the in-service training courses were not carefully monitored, and only limited effort was made to understand what effect the training was having on results in different regions. There are feedback reports for participants in the more recent rollout of in-service at second level in Ghana, and the training team is looking at ways to improve the courses, but the approach is not systematic.

---

[1] However, a video study was undertaken and used as a training tool that provided a form of feedback into the learning cycle. The video study was a one-off effort.