Assessment in Large-scale Teacher Professional Development Programs

Pearl Boateng and Freda Wolfenden
2022

Executive Summary

- There is a distinct gap in the literature concerning assessment and teacher professional development (TPD) programs. Assessment of TPD (i.e., overall TPD program evaluation) and assessment in TPD are often conflated in existing literature though both are crucial in measuring the effectiveness of any TPD program.

- This paper focuses on assessment *in* TPD, which refers to the explicit and specific assessment of and for teachers' professional learning observed through changes in their individual beliefs, knowledge, and practices. Assessment in TPD is inherently complex; often lacks clarity; and intersects with various discourses, power relations, and hierarchies.

- Implementing effective assessment of and for teachers' learning in large-scale TPD programs without compromising quality and equity is challenging. But some promising examples from the Global South highlight where assessment at scale can be enhanced using technology. These examples also illustrate typical resource demands for different types of assessment at scale as well as other considerations for policymakers and practitioners.

- Reflective frameworks can provide useful prompts in designing assessment components in large-scale TPD programs, taking into account system-level considerations, assessment design, feedback loops, teacher recognition/motivation, and evaluation of the assessment design itself.

Acronyms

BLF   Building Learning Foundations
CPE   Computadores para Educar
EGRS   Early Grade Reading Study
ICT   information and communications technology
LTCEP   Learners' Teacher Competency Enhancement Program
PLE   personal learning environment
PRIMR   Primary Math and Reading
GEEAP   Global Education Evidence Advisory Panel
TCT   Teacher Competency Tests
TCTP   Tu Clase, Tu País
TEJAS   Technology Enabled Education through Joint Action and Strategic Initiatives
TESS-India   Teacher Education through School-based Support in India
TISS-ITE   Tata Institute of Social Sciences – Integrated approach to Technology
                     in Education
TPD   teacher professional development
UNESCO   United Nations Educational, Scientific and Cultural Organization
ZEST   Zambian Education School-based Training

Table and Boxes

Introduction

This briefing note focuses on issues pertaining to assessment within large-scale teacher professional development (TPD) programs. Conventionally, there are two main purposes for assessment — for certification and for supporting learning (Newton, 2007). In professional learning programs, we might add a third purpose: to promote career-long professional learning through the development of skills associated with an identity as a self-evaluating and self-regulating professional (Boud, 2000).

Large-scale TPD programs are concerned with improving the quality of classroom teaching through practice-based learning, and programs are designed around the idea that specified outcomes for participating teachers focus on changes in their classroom and school practices. However, teachers' practice is deeply bound up with cultural discourses; competing perspectives on good professional performance; the politics of knowledge, institutions, and systems (including tensions with regulatory regimes); as well as systems of professional accountability. These shape how the outcomes are formulated. In addition, such outcomes often focus on the short term rather than equipping teachers to adapt and respond effectively to longer term changes in the education ecosystem: they do little to support career-long professional learning.

How these outcomes are assessed is equally complex and influenced by the dominant views of learning, knowledge, and the role of the teacher in each context. These will shape how the outcomes are understood by teachers and assessors. When knowledge is seen as objective, transferable, and individually-based, assessment in TPD will foreground the reproduction of knowledge — the learning of content — and utilize modes of assessment, such as multiple-choice questions and examinations, which check what teachers know. These forms of assessment are often seen as neutral; they do not impact teachers' assessment performance. But when professional learning is understood as a social activity — socially generated and moderated — assessment will be more likely to involve approaches such as learning portfolios and peer observation that are situated in teachers' historical and social contexts.

However, research undertaken for this briefing note found that there are significant gaps in the assessment of teachers within most TPD programs, including large-scale programs in the Global South. Several papers highlight assessment, or evaluation, *of* the overall impact of large-scale TPD programs but few papers address how teachers' learning is assessed *within* such programs. This neglect of assessment is problematic; it risks ignoring the learning potential in assessment, in particular the opportunity to influence learning practices and develop teachers' abilities to self-assess their work.

This briefing note hopes to add to limited discussions in this area and begin moving towards a reflective framework that aims to ask important questions about how teachers' professional learning, including changes in professional practice, can be usefully, authentically, and reliably assessed in large-scale programs. It seeks to emphasize the importance of assessment being equitable for teachers in terms of access, content, and/or modality. It also underscores the importance of assessment in communicating priorities in quality teaching, in particular equity in learning for pupils.

The briefing note starts by providing an overview of our working definition of assessment and why it is important. We then present a typology of common types of assessment in TPD programs and their characteristics. In reviewing examples for the typology, we prioritized approaches that aspire to recognize teacher agency, avoid excessive atomization of practice, and could be used and adapted across different contexts. This reflects our position that teachers' practice is context-derived, emergent, and a collective endeavor with students rather than an individual activity of acquiring new knowledge and skills (Fenwick, 2009).

The briefing note concludes by offering some reflective questions for consideration by policymakers, practitioners, and implementers of large-scale TPD programs.

The purposes of assessment in TPD programs

Adapting the UNESCO definition, we define assessment *in* TPD as the process through which the progress and achievements of a teacher are measured or evaluated in line with a defined quality criteria to support the development of their professional practice (UNESCO-IBE, n.d.). *Therefore, our focus is on the developmental trajectory of teachers' continuous professional learning rather than program evaluation and/or performance management.* For the purposes of this briefing note, assessment *of* TPD refers to the overall evaluation of TPD programs and whether the program has achieved its stated objectives and/or goals. Assessment *in* TPD (and assessment *for* TPD) refers to the explicit and specific assessment of the impact of program participation on teachers' beliefs, knowledge, and practices.

Both types of assessment are crucial to assess the overall effectiveness of any TPD intervention, build the evidence base of what works and under what conditions, and provide evidence of impact on student learning. For the teacher, assessment *in* TPD provides a marker of what is valued in the professional sphere, be it professional competencies, pedagogy-related competencies, or otherwise. It also offers possibilities for external validation of progress in these competencies (Mardapi & Herawan, 2018). In many low-income contexts, engagement with TPD has a direct effect on promotion, progression through national career frameworks, and at times, salary. Where TPD is solely and explicitly related to career progression, this can create adverse incentives and a focus on compliance rather than transforming practice. Holistic assessment in TPD can contribute to a process and culture of lifelong/continuous professional learning, motivation, and improved practice. It is important that this assessment is contextualized, relevant, and positioned so that its purposes and intent are clearly communicated to relevant stakeholders. However, assessment frameworks and tools are never neutral and will undoubtedly be interpreted and understood differently by different people.

ASSESSMENT *OF* TPD

There is limited evidence on what constitutes effective or successful TPD in low-income contexts (Boateng & Wolfenden, 2021; Haßler et al., 2018). Calls for additional research are dominated

by a focus on student achievement (Mardapi & Herawan, 2018), which although extremely important, still does not address issues of how explicit changes in teaching and learning processes occur, particularly those that result in improved student outcomes. Much of the existing evaluation studies from low-income contexts focus on small-scale, one-off initiatives and interventions with variable methodologies (Sims & Fletcher-Wood, 2021) and recent literature call for better quality assessment of such programs and better linkages to student learning outcomes (Popova et al., 2021). Evidence from high-income contexts demonstrates an increasing number of focused, rigorous, and causal impact research studies exploring the link between TPD programs and learner outcomes. In parallel, studies on large, donor-funded TPD programs are increasingly focused on student learning outcomes.

However, school-level evaluation of TPD is often neglected (King, 2014), is considered problematic (Hampton et al., 2004), and is seldom carried out in a "systematic and focused manner" (Muijs & Lindsay, 2008, p. 196). Where it does occur, it appears to lack explicit reference to clearly articulated learning outcomes for both teachers and students and is therefore more instinctive and less systematic in approach (King, 2014).

ASSESSMENT *IN* TPD

Where assessment *in* TPD occurs, it has traditionally focused on teacher satisfaction/ perceptions (often using self-reported data as evaluation) or on classroom observations (Hampton et al., 2004; O'Sullivan, 2011). A particular challenge in assessment in TPD is "gathering data about what teachers actually do and how they learn in the program; what roles they play as learners and the nature and extent of their actual opportunities for learning" (Ingvarson et al., 2003, p. 30). For example, although teacher self-assessment is acknowledged to be important for teachers because it strengthens the process of identity formation (Wenger, 1998), there are few recorded instances of the use of robust self-assessment in TPD programs. Similarly, although both giving and receiving peer feedback are widely recognized as helpful for learners to develop critical learning skills and an understanding of feedback as a dialogue about practice, few examples of this are found in TPD programs.

Other scholars, such as Merchie and colleagues (2018), highlight that TPD is inherently complex, with blended purposes leading to a lack of clarity in many existing assessment models. These issues are compounded and even more complex in large-scale programs. As such, some authors argue that it may be best to begin studying the impact of a TPD intervention on teacher knowledge only (King, 2014). However, an emphasis on specific codes of knowledge rarely aligns with or affirms the messy actuality of classroom practice and problems or dilemmas encountered by teachers on a daily basis. Likewise, it seldom recognizes that professional learning is constructed within the complex conditions of the specific education setting, with its particular discourses, power relations, and hierarchies (Billet, 2004).

As a result of the above, few studies or programs address how teacher beliefs, knowledge, behaviors, and practices change through participation in TPD over time. There is also a paucity of studies on the tools needed to assess teachers' learning in ways that acknowledge the complexities of practice and experience. Moreover, there is a dearth of research that challenges

the idea that professional learning is acquisitive, non-political, and representable (Fenwick, 2009). There is a particular lack of examples within large-scale TPD programs largely due to the number of additional considerations needed. This includes various dimensions of system capacity, quality, equity, and resourcing, to name a few.

Table 1 outlines some of the forms of assessment used within large-scale TPD programs and describes some of their characteristics. These examples were found through rapid literature searches/reviews as well as previous in-depth research conducted via a complementary briefing note, "TPD@Scale Briefing Note: Moving Towards Successful Teacher Professional Development in the Global South" (Boateng & Wolfenden, 2022) and the "TPD@Scale Compendium" (Foundation for Information Technology Education and Development, 2022). This is followed by an emerging framework of reflective questions for policymakers, program implementors, and practitioners to strengthen approaches for the assessment of large-scale TPD programs.

TABLE 1. FORMS OF ASSESSMENT IN LARGE-SCALE TPD PROGRAMS IN THE GLOBAL SOUTH[1]

| Type of assessment | Frequency of use | Resource demands (at-scale) | | | Proximity to classroom practice | Level of personalization/ feedback to teachers | Advantages | Disadvantages | Ways in which assessment can be enhanced at scale | | Project examples |
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	
1. Discussions in focus groups/teacher professional communities	Medium	High/High	Low	High	Dependent on content of questioning/ discussion	High/Medium	Can gather a variety of data evidencing changes in teachers' knowledge, beliefs, values, and practices	• Resource-intensive • Can be reliant on self-reported changes and will need to be triangulated with other evidence to make significant conclusions	ICT can be used to make data collection/analysis more effective.	Ensure inclusion of teachers/ schools from various locations, and with different experiences, backgrounds, needs, etc.	• Technology Enabled Education through Joint Action and Strategic Initiatives (Tejas) (*see TPD@ Scale Compendium Profile 4*) • Early Grade Reading Study II (EGRS II) – Wits • Zambian Education School-based Training (ZEST) (*see TPD@Scale Compendium Profile 13*)

[1] The examples provided here are not exhaustive but are based on those found during literature review and searches.

Type of assessment	Frequency of use	Resource demands (at-scale)			Proximity to classroom practice	Level of personalization/ feedback to teachers	Advantages	Disadvantages	Ways in which assessment can be enhanced at scale		Project examples
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	**Project examples**
2. Lesson observations/ micro-teaching/ practice	Medium	Medium to high/Medium	Low/ Medium depending on tools used to conduct observations & give feedback	High	High	Medium to High/High	• Directly related to classroom practice and can provide evidence on actual changes in practice over time if done at regular intervals • Allows real-time assessment of improvements/ challenges in teaching and learning • Can link TPD inputs to changes in practice	• Human resource-intensive even if primarily school-based • Can be time-consuming and requires expert staff across multiple locations to facilitate the lesson observation process and feedback • Requires careful scheduling to be sustained long term	• ICT can be used to support observers to collect data more efficiently and across multiple locations. • A more peer-led approach (e.g., video clubs, communities of practice), coupled with training and resources can support teachers to conduct their own observations, reflect, and share feedback.	• Recognize different teachers' starting points and emphasize inclusive teaching practices • Can be led by (trained) peers to support mutual learning while conducting observations	• Primary Math and Reading (PRIMR)/ Tusome • Tejas • Ceará Teacher Feedback and Coaching Program (*see TPD@Scale Compendium Profile 7*) • Building Learning Foundations (BLF) (*see TPD@Scale Compendium Profile 5*) • ZEST • Tu Clase, Tu País (TCTP) (*see TPD@ Scale Compendium Profile 11*)

| Type of assessment | Frequency of use | Resource demands (at-scale) | | | Proximity to classroom practice | Level of personalization/ feedback to teachers | Advantages | Disadvantages | Ways in which assessment can be enhanced at scale | | Project examples |
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	
3. Lesson plans	Medium	Low/Low	Low	High	Medium	High/High	Allows assessment of teachers' changing approach to teaching and learning	• On its own, doesn't address actual lesson delivery • Can be time-consuming for school leadership to check and difficult to do for all teachers at scale	ICT can be used by teachers to draft lesson plans and by reviewers to assess them remotely and more efficiently via an online platform.	Ensure lessons explicitly address different types of students and their specific needs and that teachers receive differentiated support	• Peking University X-Learning Centre (*see TPD@Scale Compendium Profile 9*) • Tata Institute of Social Sciences – Integrated approach to Technology in Education (TISS-ITE)

| Type of assessment | Frequency of use | Resource demands (at-scale) | | | Proximity to classroom practice | Level of personalization/ feedback to teachers | Advantages | Disadvantages | Ways in which assessment can be enhanced at scale | | Project examples |
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	
4. Multiple choice tests/quizzes/ questionnaires/ entry/exit tests	High	High/High	High	Medium	Dependent on content	Low/Medium	Can be useful for needs analysis and identifying areas for improvement at scale but needs comprehensive follow-up support to enhance it further	• If not coupled with adequate follow up support/ feedback, can be quite limited in supporting teachers to actually change their practice • Can be quite removed from the contexts in which teachers teach if not carefully designed	ICT can be used as a core medium to host quizzes/tests, etc.	• Ensure online/offline modalities for varying levels of connectivity and access • Provide differentiated content where possible	• TCTP • Learners' Teacher Competency Enhancement Program (LTCEP) Indonesia • Teacher Education through School-based Support in India (TESS-India) (*see TPD@Scale Compendium Profile 1*) • TISS-ITE • Teachers for Teachers (*see TPD@Scale Compendium Profile 6*) • ZEST • Computadores para Educar (CPE) (*see TPD@Scale Compendium Profile 3*)

Type of assessment	Frequency of use	Resource demands (at-scale)			Proximity to classroom practice	Level of personalization/ feedback to teachers	Advantages	Disadvantages	Ways in which assessment can be enhanced at scale		Project examples
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	
5. Online portfolios (An online platform which hosts a variety of evidence on teacher knowledge, learning and practice, e.g., example lesson plans, online tasks, videos, uploads of examples of student work, etc.)	Low	High/High	High	High	Dependent on content	High/Medium to High	Can provide a sustainable platform where teachers can store evidence and reflect on their learning and progress over time, thus increasing their ownership of their professional learning journey	• Can be more challenging in contexts with weaker ICT infrastructure • Costly to set up • Requires ongoing technical know-how and support to maintain in the long term	Not applicable	Include differentiated pathways and content to support teachers with varying skills, knowledge, and at various stages of their career	• CPE • TCTP

Type of assessment	Frequency of use	Resource demands (at-scale)		Human resource level of expertise required	Proximity to classroom practice	Level of personalization/ feedback to teachers	Advantages	Disadvantages	Ways in which assessment can be enhanced at scale		Project examples
		Financing (Start-up costs/ ongoing costs)	*ICT*						*By ICT*	*For equity*	
6. Peer assessment (Online and offline forms of peer collaboration. This might include peer review of lesson plans, peer lesson observations, peer review of tasks via MOOC platforms, etc.)	Low	Low	Low to High depending on the medium used	High	Dependent on content of tasks	High/High	Can be a sustainable way of building a cadre of teachers who are able to conduct their TPD activities (e.g., lesson observations) and engage in ongoing mutual learning with peers	Resource-intensive in the initial instance and would require high-quality training, guided materials, resources, etc., to ensure that teachers' knowledge and capacity are strengthened to conduct peer tasks effectively	• ICT can be used to support observers to collect data more efficiently. • Online platforms can also be used to enable peer assessment across multiple locations leading to greater sharing of more diverse ideas, learning experiences, etc.	Ensure all teachers' experiences and practice are acknowledged and leveraged to provide/ challenge/support teacher motivation at different stages of their career/ professional development	• TESS-India • TCTP • TISS-ITE • CPE

| Type of assessment | Frequency of use | Resource demands (at-scale) | | | Proximity to classroom practice | Level of personalization/ feedback to teachers | Advantages | Disadvantages | Ways in which assessment can be enhanced at scale | | Project examples |
		Financing (Start-up costs/ ongoing costs)	*ICT*	*Human resource level of expertise required*					*By ICT*	*For equity*	
7. Reflective videos/video assignments (e.g., tasks requiring teachers to reflect on their knowledge/ learning via video upload; assignments that require teachers to use video to demonstrate classroom practice learning, etc.)	Low	High/High	High	High	Dependent on content	High/High	• An innovative way of teachers reflecting on their practice and gaining feedback from a tutor/ coach • Could take a peer-led approach as mentioned above	Requires stable ICT infrastructure, connectivity, and expert human resource to review video content and provide feedback	Not applicable	Include differentiated pathways and content to support teachers with varying skills, knowledge, and at various stages of their career	• CPE • TCTP • BLF

| Type of assessment | Frequency of use | Resource demands (at-scale) | | | Proximity to classroom practice | Level of personalization/ feedback to teachers | Advantages | Disadvantages | Ways in which assessment can be enhanced at scale | | Project examples |
		Financing (Start-up costs/ ongoing costs)	ICT	Human resource level of expertise required					By ICT	For equity	
8. Self-assessment tests	High	Low	Medium/ High	Medium	Dependent on content	Low/Medium	Can be useful for needs analysis and identifying areas for improvement at scale but needs comprehensive follow-up support to be made more effective	• If not coupled with adequate follow up support/ feedback, can be quite limited in supporting teachers to actually change their practice • Can be quite removed from the contexts in which teachers teach if not carefully designed	ICT can be used as a core medium to host self-assessment tests.	• Ensure online/offline modalities for varying levels of connectivity and access • Provide differentiated content where possible	•TCTP

Each type of assessment should be addressed holistically and reviewed in terms of how it works alongside other key enabling factors in the system as well as other core TPD program components. As discussed in our complementary briefing note (Boateng & Wolfenden, 2022), high-quality, large-scale TPD programs should 1) promote teacher agency and inclusion of both teachers and students; 2) be authentic and classroom-based; 3) actively encourage reflective practice; 4) provide ample opportunity for collaborative peer-learning; and 5) be sustained over time.

A reflective framework to begin rethinking assessment in large-scale TPD programs

This final section focuses on important questions to guide the design and development of assessment in large-scale TPD programs. As outlined in Table 1, while each individual type of teacher assessment has its advantages and disadvantages, any selected combination of these types of assessments are most effective when carefully chosen and integrated into a holistic TPD framework. This is evidenced by the cases presented in Boxes 1-3. These examples of promising practices highlight how assessment in large-scale TPD programs is being attempted in Indonesian, Indian, and Colombian contexts using competency mapping approaches, open digital badges, and reflective videos as part of their respective wider TPD frameworks.

In the example from Indonesia (Box 1), the repeated use of the Learners' Teacher Competency Enhancement Program (LTCEP) is an attempt to overcome the noted limitation of competency lists in capturing the complexity and entirety of teachers' knowledge-in-action in their classrooms. Meanwhile, the linking of a small number of assessments to an open digital badge has been well received by teachers in a pilot project in India (Box 2).

Box 1. Promising practice: Nationwide teacher assessment and competency mapping framework in Indonesia

In 2012, Indonesia launched its Teacher Competency Tests (TCT) as part of its national teacher competency mapping process, with a teacher workforce of approximately three million. The TCT assesses the professional and pedagogic competence of teachers and uses both online/ offline modalities (depending on local connectivity). In 2015, the TCT had reached approximately 2.7 million teachers. Its aims were: 1) to obtain information on the number of teachers who have reached the required level of professional and pedagogical competency; 2) to create a teacher competency map, which informs the State's provision of subsequent teacher education and training; and 3) to assess teacher performance and inform subsequent teacher-related policy. In terms of design, the TCT took the form of a two-hour test comprised of 60-100 multiple choice questions, with four options for each question.

The TCT in Indonesia is a key component in teacher career development. Once the test is completed and a competency profile created, this is used as guidance for the self-development of teachers as well as for support provision through TPD. At regular intervals, the TCT is conducted again to assess progress. The new results of the TCT are then used as the basis for further guidance and teacher development.

In 2016, the government launched a follow-up policy on TCT results through the Learners' Teacher Competency Enhancement Program (LTCEP). The LTCEP provides learning opportunities that aim to improve the ability and competence of teachers in performing professional duties. The TPD activities are aimed at developing teachers' abilities, attitude, and skills with the view of impacting teacher performance and the teaching and learning process in the classroom. These activities are offered via face-to-face training, online training, or a combination of both depending on teachers' competency profile and the intensity of support needed.

Source: Mardapi & Herawan, 2018

Box 2. Promising practice: Pre- and post-COVID-19 TPD using digital badges in India

The COVID-19 pandemic further emphasized the need to support teachers at scale. In this vein, the Government of India has been considering the use of open digital badges to support at-scale TPD efforts. While open digital badges have received global recognition as a means for designing, recognizing, rewarding, and monitoring teachers' professional learning, badges are currently rarely used in many countries. But workshops and pilot projects implemented by The Open University, UK and Tata Institute of Social Sciences in Assam, India have demonstrated how digital badges can be used to support professional learning.

Their research suggested that open digital badges linked to small, short, and standalone achievements give status and recognition to teachers' efforts. Badges offer wide-ranging flexibility in what can be awarded, evidenced, and assessed; clarity for teachers in terms of articulating what is expected; scalability by providing a potential common digital currency for local, regional, and national TPD systems; engagement by allowing teachers' choices in which badges they pursue to support their professional learning trajectories; and shareability by allowing teachers to share their achievements with others. This in turn builds teachers' confidence and self-esteem; enhances community recognition; and provides a marker for individual teachers' progress and future TPD planning.

Source: Cross et al., 2021

As a final example, the Computadores para Educar (CPE) program in Colombia (Box 3) is attempting to scale the reification of practice in a teacher-created video. The video potentially plays two roles: as an assessment form and as a tool that itself mediates learning.

Box 3. Promising practice: Reflective videos as teacher assessment in Colombia

Computadores para Educar (CPE) is a program run by the Government of Colombia that promotes educational innovation through the access, use, and adoption of technology in the country's schools. The CPE program offers ICT-mediated learning through the provision of computers and internet to Colombian schools. Opportunities for TPD are offered in partnership with local universities that have a presence in regions of operation.

The program provides differentiated pathways through the Ministry of Education-defined "Educational Innovation Route," which aims to develop teachers' ICT competencies. This is offered in a blended learning modality and consists of four modules of progressive complexity.

In terms of teacher assessment, program tutors administer an entry and exit ICT skills test for teachers. There are four knowledge tests at the end of each module of the course that must be passed. The final product — the capstone assessment — is a video made by the teacher, in which the learning process across the four modules is documented. The video remains as evidence in the teacher-student's personal learning environment (PLE).

Source: SUMMA, 2021

The examples demonstrate some promising initial approaches to assessment in TPD. Assessment in large-scale TPD programs is often more challenging due to the sheer number of the teaching workforce; uneven distribution of resources and human capital; logistical challenges; financing; policy coherence; and how these relate to teacher motivation, career progression, teacher choice, equity considerations, and so on. Nonetheless, a more reflective approach to design can move us towards more effective at-scale approaches.

Box 4 offers an emerging framework of reflective questions/design prompts that can begin to support policymakers, practitioners, and program implementers of large-scale TPD programs. Of course, many of these questions need to consider a number of issues in a holistic and interconnected manner.

Box 4. Reflective questions for the design of teacher assessment in large-scale TPD programs

System context

- How does the assessment design fit within current system capacity?
- What opportunities and challenges might affect assessment implementation in different contexts (e.g., north/south, urban/peri-urban/rural)?
- Which roles at the school/district/regional/national levels are needed to support the sustainability of the assessment approach?
- How does the assessment approach align with existing teacher competency frameworks?

Assessment design

- How can it be ensured that all teachers and their various starting points (of knowledge/skills/ practice) are considered in assessment design, content, and access?
- How are different authentic classroom contexts within the country being considered in assessment design?
- What kind of practitioner-led assessment strategies can teachers use?
- Does the assessment relate to student learning outcomes?
- How does the assessment take different forms of inclusion for both teachers and students into account?
- What proportion of the assessment focuses on classroom practice?
  - Is this being given sufficient emphasis?
- In what ways are teachers' knowledge/practice/skills expected to change and be applied in the classroom?

Feedback

- What opportunities do teachers have to reflect on their practice, whether individually or collaboratively?
- In what ways are teachers being offered opportunities to share and collaborate with others at the school and local levels to strengthen learning?
- How do teachers receive feedback?
  - How does the feedback teachers receive link to further professional learning?
- What kind of follow-up support will teachers receive?
  - How will this follow-up support be assessed and sustained?

Recognition

- How is progress in teachers' professional competence being validated/recognized/awarded/ celebrated?
  - What different mechanisms are being used?
  - What criteria are being used?

Evaluation of the assessment in the TPD program

- What mechanisms are being used to measure teachers' experience of the assessment modalities and tasks?
  - How do these inform program improvement?
- What assessment is most useful to teachers and why?
- How is assessment being evaluated in relation to quality, cost efficiency, and equity?
- Do the assessment design/tools provide relevant data that could help teachers improve the practice of their profession?

Areas for further research

To support more evidence in this area, we suggest the following as key areas for further research:

- Characteristics of effective large-scale TPD programs in low-income countries
- Assessment of teacher professional learning and teacher change in the classroom

References

Boateng, P., & Wolfenden, F. (2022). *TPD@Scale briefing note: Moving towards successful teacher professional development in the Global South.* Foundation for Information Technology Education and Development.

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*. https://doi.org/10.1080/713695728

Cross, S., Adinolfi, L., Charania, A., Wolfenden, F., Ramchand, M., Sen, S., Paltiwale, S., Coughlan, E., Shende, S., Sarkar, D., Chatterjee, A., & Balli, O. (2021). *Teachers' professional learning during and after Covid: A role for open digital badges.* The Open University, UK.

Fenwick, T. (2009). Making to measure? Reconsidering assessment in professional continuing education. *Studies in Continuing Education, 31*(3), 229-244. https://doi.org/10.1080/01580370903271446

Foundation for Information Technology Education and Development. (2022). *TPD@Scale Compendium.*

Hampton, G., Rhodes, C., & Stokes, M. (2004). *A practical guide to mentoring, coaching, and peer-networking: Teacher professional development in schools and colleges* (1st ed.). Routledge Falmer. https://doi.org/10.4324/9780203591949

Haßler, B., Hennessy, S., & Hofmann, R. (2018). Sustaining and scaling pedagogic innovation in Sub-Saharan Africa: Grounded insights for teacher professional development. *Journal of Learning for Development, 5*(1), 58-78.

Ingvarson, L., Meiers, M., & Beavis, A. (2003). Evaluating the quality and impact of professional development programs. In M. Meiers (Ed.), *ACER Research Conference Proceedings 2003. Building teacher quality: What does the research tell us?* (pp. 28-34). ACER.

JISC. (2015). *Transforming assessment and feedback with technology.*

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

Mardapi, D., & Herawan, T. (2018). Assessing teacher competence and its follow-up to support professional development sustainability. *Journal of Teacher Education for Sustainability, 20*(1), 106-123. https://doi.org/10.2478/jtes-2018-0007

Merchie, E., Tuytens, M., Devos, G., & Vanderlinde, R. (2018). Evaluating teachers' professional development initiatives: Towards an extended evaluative framework. *Research Papers in Education, 33*(2), 143-168. https://doi.org/10.1080/02671522.2016.1271003

Muijs, D., & Lindsay, G. (2008). Where are we at? An empirical study of levels and methods of evaluating continuing professional development. *British Educational Research Journal, 34*(2), 195-211. https://doi.org/10.1080/01411920701532194

Newton, P. (2007) Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice.* https://doi.org/10.1080/09695940701478321

O'Sullivan, H. (2011). Leading and managing professional learning in schools. In H. O'Sullivan, & J. West-Burnham (Eds.), *Leading and managing schools* (pp. 111-125). Sage.

Popova, A., Evans, D.K., Breeding, M.E., & Arancibia, V. (2021). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, 1-30. https://doi.org/10.1093/wbro/lkab006

Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: A critical review. *School Effectiveness and School Improvement, 32*(1), 47-63. https://doi.org/10.1080/09243453.2020.1772841

SUMMA. (2021). *ICT-mediated teacher training models: A multi-case study of initiatives in Latin America.* Unpublished manuscript.

UNESCO International Bureau of Education. (n.d.). *Assessment.*

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* (Learning in doing: Social, cognitive, and computational perspectives). Cambridge University Press. https://doi.org/10.1017/CBO9780511803932