rppl
Research Partnership
for Professional Learning

Measuring Teacher Professional Learning:

Why It's Hard and What We Can Do About It

October 11, 2023

AIR
Advancing Evidence.
Improving Lives.

ANNENBERG
BROWN UNIVERSITY

CITY TEACHING ALLIANCE

Leading Educators

NEW TEACHER CENTER

STUDENT ACHIEVEMENT PARTNERS

Teach For America

Teaching Lab

UnboundEd

About RPPL

The Research Partnership for Professional Learning (RPPL) is a collective of professional learning (PL) focused organizations and researchers committed to advancing educational equity for students historically pushed to the margins of our education system.

Acknowledgments

The authors would like to thank the RPPL members, affiliates, research advisors, and funders who provided thought partnership in creating this report. Thank you to Dionn Brown, Karen Cornetto, Chris Engledowl, Alexandra Maus, Aizat Nurshatayeva, Carter Romansky, Krysten Wendell, Bonnie Williamson-Zerwic, and the communications team at RALLY. In particular, we thank Elizabeth Chu, Nadya Chinoy Dabby, Heather Hill, Nicole Michelle Joseph, John Papay, Jim Short, and Sola Takahashi for their helpful review and suggestions.

This report was made possible by support from the Bill & Melinda Gates Foundation and Charles and Lynn Schusterman Family Philanthropies. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders.

Authors
(in alphabetical order)

Stacey Alicea
Research Partnership for Professional Learning

Katie Buckley
Teach For America

Diana Cordova-Cobo
Student Achievement Partners

Aliza Husain
UnboundEd

Laura Meili
Leading Educators

Lisa Merrill
American Institutes for Research

Krista Morales
Research Partnership for Professional Learning

Lisa Schmitt
New Teacher Center

Nathaniel Schwartz
Annenberg Institute at Brown University

Tim Tasker
Leading Educators

Viticia Thames
City Teaching Alliance
**now at Instruction Partners*

Shaye Worthman
Teaching Lab

CONTENTS

Introduction

The Research Partnership for Professional Learning (RPPL) is a consortium of practice organizations and researchers working together to improve the quality of teaching and learning through evidence-based professional learning (PL). Over the past five months, we have convened a group of representatives from 10 PL organizations to engage with a key question for our work: **How can researchers and practitioners practically and equitably measure progress of teacher learning, identify areas for continuous improvement, and demonstrate the impact of PL experiences on teachers and students?**

Teacher PL is ultimately in service of student learning. In recent years, policy leaders at all levels of the education system have voiced calls for more data collection and greater accountability measures to gauge the student learning outcomes of professional learning efforts—both from the districts that run their own programs and from external vendors contracted to provide teacher development services. Such data could aid the field in demonstrating transparency and responsible use of resources in support of teachers' growth and development. Stronger data collection could also serve many purposes beyond accountability; it could help communicate the value and impact of PL initiatives to various stakeholders (including educators, administrators, policymakers, students, families, and the education community) and play a crucial role in fostering equitable and inclusive educational environments by evaluating the extent to which PL initiatives address issues of equity and culturally responsive teaching and meet the diverse needs of diverse students.

We need practical, equity-advancing, valid, and reliable measures that help PL organizations and their partner systems improve their services and communicate progress and impact.

Calls for more rigorous measurement of teacher professional learning, however, often ignore the challenges—both technical and substantive—of measuring PL outcomes in ways that meet the needs of multiple users and groups. Because PL organizations work on the ground year-round with educators in schools, we look to them as a critical resource for solving measurement challenges. They are experienced and innovative practitioners and have a line of vision across many different districts and schools. They implement programs, services, and solutions in real-time, test out innovations to examine results, and can surface the voices and needs of educators, leaders, and students to inform future offerings. All of this allows them to identify common challenges and needs and play a critical role in translating between school systems, researchers, and the PL field at large. Leveraging their expertise and experience has the potential to lead to better measurement solutions and more robust study outcomes.

This report, collectively produced by a working group across RPPL's network, centers the voices and practices of organizations working alongside districts, schools, and teachers, understanding that they are well-situated to surface the challenges and needs of practitioners. We describe the measurement challenges that PL organizations confront and offer views from inside these organizations about where there are opportunities for improvement. Throughout the paper, we also highlight real-world examples from each organization that show how they are using and developing innovative, practical, equity-focused measures to address the challenges they face.

CHALLENGES IN THE FIELD

→ The View from RPPL Organizations

Our PL organizations take data and measurement seriously. Each of our organizations has built comprehensive frameworks for data collection, employs researchers and analysts to make sense of the data they collect, and actively seeks out new innovations for measuring improvement in outcomes. Most of our organizations also partner directly with academic researchers to gather additional data for effectiveness studies and test new innovations and measurement tools. **Why, then, is the consensus among us that we still do not have the measures necessary to capture the data we need to evaluate the implementation and outcomes of PL programs?**

We use PL data for multiple purposes. Specifically, the data we collect—through teacher and sometimes student surveys, classroom observations, and various forms of student assessment—gets used formatively, to provide just-in-time information back to teachers and school leaders about areas for improvement and to inform our organizations' ongoing practices. Ideally, the data are also used to summatively capture teacher growth, to assess program effectiveness, to inform reports to districts, funders, and other external stakeholders, and to contribute to broader field-building research.

In a perfect world, we would have aligned measures suited to all of these purposes. But we are working in environments where teacher and leader time is short, where privacy rules rightfully require considerable protections for sharing information, and where the capacity for data collection and analysis both within districts and within our organizations is limited. This means that we try to develop and pick instruments that can meet multiple needs simultaneously.

Here, we highlight some of the central challenges and tensions that make it more difficult for our organizations to achieve the kinds of data-driven improvement work that we believe will help us best support students, teachers, and leaders.

CHALLENGES IN THE FIELD

Misalignment of Research Priorities

One key challenge is that **researcher–developed and validated measurement instruments are not aligned to the needs of PL organizations and their school partners.**

The strongest research-developed measures of teacher perception and practice—surveys about teachers' views on their self-efficacy, for example, or classroom observation tools to gauge instructional shifts—have the benefit of validation studies that ensure these tools return reliable results that offer information about meaningful, underlying constructs that directly relate to long-term outcomes.

The reality is that, even within the research world, such measures are not always easy to come by (e.g., the RAND American Mathematics Educator Study Survey). Most of our organizations work to develop teachers' knowledge of equitable teaching practices and to collect data about teachers' responses to this training, but the researchers that support RPPL know of only one existing survey measure of Common Core-aligned practices that aims to study this topic.

Even when research-developed measures exist, they are frequently inflexible, lengthy, and not built to translate into practical, actionable feedback for teachers or school leaders, therefore producing unclear and broken throughlines to classroom improvement. For example, Teaching Lab (TL) and many of our PL organizations prioritize culturally responsive and sustaining education (CRSE) practices within their teacher learning sequence. Research-developed surveys to measure teachers' understanding of these practices do exist, such as the Culturally Responsive Teaching Self-Efficacy Scale (CRTSE) and Culturally Responsive Teaching Outcome Expectancy (CRTOE). However, these surveys consist of 40 and 26 items, respectively, making them impractical to add to organizational surveys that aim to capture multiple constructs at once. TL conducts a biannual

CASE STUDY

Teaching Lab

How to Build Comprehensive Frameworks for Data Collection

PL organizations require overarching evaluation frameworks to guide their measurement cycles and determine the impact of their services on teachers and students. The Guskey Framework (2016) is a useful starting point for PL organizations to begin to develop their own evaluation plan that aligns with their theory of action. While the framework is simple and linear, disentangling the correlational and causal relationships between the different levels and the impact they have on one another is not.

The first four aspects of Guskey (participant perceptions, knowledge, mindsets, and enabling conditions) can be reasonably discerned from surveys and assessments, effectively evaluating the most important outcomes, teacher practice and student learning, is increasing complexity and resource-intensive; classroom observations and student data, including surveys, student work samples, and formative and summative assessment data, all require time and important human resources collect, analyze, and interpret.

Measurement cycles at PL organizations vary depending on the scope of work, specifically the types of services and the timeline of partnerships. Many PL providers, such as Teaching Lab (TL), collect data on two broad schedules.

Teaching Lab

→ READ FULL CASE STUDY

CASE STUDY

UnboundEd

Establishing the Reliability and Validity of Self-Created Tools

In January 2022, Northwest Evaluation Association (NWEA) partnered with CORE Learning, a subsidiary of UnboundEd, to offer collaborative professional learning to educators with one of NWEA's large school system clients in Alaska. The program was designed as a three-credit college course and included synchronous professional learning and asynchronous coaching from NWEA and CORE's Online Elementary Reading Academy (OERA) for approximately 100 teachers. To assess changes in teacher knowledge, NWEA and CORE collected responses from approximately 100 teachers before and after they completed the course. An end-of-program report measuring teacher knowledge using a modified version of the Teacher Knowledge of Early Literacy Skills (TKELS) showed improvements in teacher knowledge. TKELS was selected as an external instrument because it was developed as a third-party measure of the impact of an early literacy initiative and it measured a wide selection of early literacy knowledge in which NWEA's client was interested.

Before implementing the modified version of TKELS, NWEA and CORE considered how well the modified assessment aligned with the collaborative professional learning program they were delivering. Since it was used solely for internal purposes, the assessment was considered a close enough match to the goals of the program.

nwea UnboundEd

→ READ FULL CASE STUDY

educator survey that asks questions about educators' mindsets (e.g., recognition of the importance of race and culture, holding high expectations for all students), enabling conditions such as school environment (e.g., teacher trust and social capital) and CRSE. If TL was to use one of the research-developed surveys of CRSE along with similarly long surveys of its other target areas, teachers would spend their valuable and limited learning time answering survey questions instead of working together on learning that could improve student outcomes. More likely, they simply would stop responding to survey questions, invalidating TL's data collection efforts.

Length is not the only concern. According to organizations within our working group, researcher-developed measures are usually grounded in a framework or set of principles that school stakeholders may not be aware of and/or could be at odds with their understanding of the issue (e.g., differing conceptions of what constitutes high-quality instruction). Measures developed by researchers are also not typically created in collaboration with teachers, often resulting in data collection that feels irrelevant to partner schools because they do not center teachers' needs or experiences as educators and classroom leaders. For the PL field, specifically providers, an additional challenge is that measures validated in other contexts might not actually correspond well to their organization's theory of change, programming, or products. Further compounding this challenge, PL providers are incentivized to have distinct theories of change and programming from each other, making it even more complex to use previously validated measures across the suite of PL organizations. In other words, the existence of research-developed measures is not enough. The process by which measures are developed must bridge the gap between the measurement needs of practitioners, PL organizations, and researchers.

CHALLENGES IN THE FIELD

Lack of Practical Measurement Tools

A related challenge is that the field **lacks sufficient practical measures of PL to support ongoing progress assessment and continuous improvement.**

Our organizations need data and measurement systems to build our knowledge and our partner districts' knowledge of what is required for ongoing improvement. Yet many of the current research-developed measures, such as classroom observation tools, are primarily used as summative assessments. Although summative assessments are crucial in measuring change over time, they do not give teachers in-the-moment feedback on how to improve their practices. Formative assessments that provide data that can be used for continuous improvement are core to the work of our organizations. To further complicate this challenge, our organizations also leverage formative assessments to obtain feedback on how well they have delivered their programming for their own improvement efforts, making capturing participants' experiences a tricky balancing act.

To build this formative knowledge, organizations generally develop custom measures and tools or combine items from existing assessments in ways that likely violate the integrity of the broader scales and invalidate any established psychometric properties. The result is a series of measures that are more applicable to the relevant contexts, but their psychometric properties are unexamined.

A true redesign would require additional effort to investigate item and construct validity in the context and among the target population for which it was redesigned. Most PL organizations do not have the time, expertise, or financial resources to conduct such psychometric studies. These measurement needs, coupled with the prohibitive cost of engaging in measurement validation, often result in the use of tools that are shared widely but don't meet research measurement standards and thus are difficult to use as a source of precise and generalizable knowledge.

These examples highlight PL organizations' need for shorter, more targeted instruments that they can compile to determine continuous change across different areas related to the focus of their services. This echoes a broader call across the field for what some researchers call "practical measures" to support continuous improvement (Yeager, Bryk, et al., 2013; Carnegie Foundation for the Advancement of Teaching, n.d.). Such practical measures are short tools, meant to be used frequently to gauge improvement in proximal indicators of change that can inform immediate decision-making as well as predict distal outcomes of interest.

Practical measures are widely defined as having three core properties:

- **Useful:** Yield meaningful, actionable, and relevant data for practitioners
- **Easy:** Minimally burdensome to collect and analyze; easily embedded into practitioner routines, but predictive of longer-term outcomes
- **Consequential:** Connected to experiences that matter for learning and to outcomes we care about for students, especially those who have been traditionally underserved (WestEd, n.d.)

We see promise in efforts such as the Carnegie Foundation for the Advancement of Teaching's Practical Measurement for Improvement Library; the Practical Measures, Routines, and Representations project; WestEd's Math Practical Measurement Project; and the EdInstruments database. These research organizations created repositories after bringing together researchers and practitioners and building consensus around how to best support educators in improving teaching and learning. They serve as models for RPPL to build upon as we continue to grapple with the measurement challenges of our PL organizations and the broader PL field.

CHALLENGES IN THE FIELD

Insufficient Attention to Equitable Measurement

Even when measures are practical, they often do not consider **equity in classroom instructional practices and teacher and student experiences.**

Across various frameworks, several core principles consistently emerge in relation to equitable measurement, often situated in the larger context of equitable or culturally responsive evaluation (Equitable Evaluation Initiative, n.d.; Hood, Hopson, & Kirkhart, 2015; We All Count, n.d.). Despite these frameworks, however, measures and data collection processes do not routinely address how power dynamics and structural inequities manifest in schools, classrooms, and teacher-student relationships.

Currently, off-the-shelf research-based measures run the risk of being used in ways that do not center students' needs or experiences. This can result in the generation of practices grounded in insufficient data that can be used in inequitable and harmful ways with students, particularly historically marginalized students. PL organizations seeking to support equitable teaching practices but using off-the-shelf measures must develop and adopt both innovations in measurement and a continuous improvement strategy that specifically aims to increase the learning of students historically pushed to the margins of our education system. An additional tension at the heart of the challenges for PL providers' measurement issues is that they need their work to "center" educators, or at least pass muster with them, for it to have a chance of showing up in classrooms. However, it is also possible that what passes muster with teachers is misaligned with student needs, which may not always align with the focus of PL providers. Our PL organizations are actively working to meet this challenge by incorporating equity-grounded frameworks and practices in their PL services. Yet, they struggle to find and leverage the right tools and partnerships to assess their continuous improvement efforts and impact.

CASE STUDY

City Teaching Alliance
(formerly known as Urban Teachers)

Embedding Transformative Social and Emotional Learning into Measurement

There is evidence that teachers may operationalize guidance for developing students' social and emotional learning (SEL) skills through a deficit lens when students have disabilities or are students of color. For example, Kaler-Jones (2020) cited an instance when classrooms in a predominantly Black and Brown school had posters on the wall that defined social awareness as "keep your hands to yourself." In response to similarly reported (mis)uses of SEL with students with disabilities and students of color, the Collaborative for Academic, Social, and Emotional Learning (CASEL) refined its definition of high-quality systemic SEL to include "a specific form of SEL implementation that concentrates SEL practice on transforming inequitable settings and systems, and promoting justice-oriented civic engagement—we are calling this *transformative SEL*." However, to date, there have yet to be research-based measures of transformative SEL (T-SEL) and its implementation as of yet in the literature.

So while City Teaching Alliance had developed a T-SEL curriculum and T-SEL classroom observation rubrics, there were no ready-to-use, off-the-shelf teacher and student surveys aligned to teachers' and students' experience that City Teaching Alliance's external research partner could use to evaluate its effectiveness.

CITY TEACHING ALLIANCE

→ READ FULL CASE STUDY

To center and provide ownership to the individuals and communities most proximate to the problem of practice we want to address, measurement tools should:

1. Be selected, adapted, or created with involvement from school leaders, teachers, and/or students.

2. Be contextualized to local school systems and communities and/or teacher and student cultural experiences and needs.

3. Assess factors core to educational equity, such as equitable access to resources, representation, inclusive instructional practices, culturally responsive teaching, and the reduction of achievement gaps among student subgroups.

4. Support the act of sharing back data and learning in meaningful, accessible, and actionable ways to school leaders, teachers, and students.

For example, according to our organizations, few protocols and virtually no tools exist that put student academic data alongside student experiences in one place, and help leaders and teachers take coherent action on that data. ANet has sought to solve this problem through a series of small scale changes. First, by defining the elements that make up an "Empowering Learning Environment" for students, then by gradually incorporating measures aligned to those in their partner schools. ANet is now in the first year of piloting their student experience surveys alongside student academic data to support teachers and leaders to take action based on these combined data.

All data and learning, whether formative or summative, should be routinely and consistently shared and contextualized in ways that are meaningful, accessible, and actionable to school leaders, teachers, students, and the PL field at large. This aligns with continuous improvement efforts grounded in practical measurement that provide just-in-time data used to inform and improve equitable instruction practices and teacher and student experiences. It also ensures that summative research that is truly evaluating the impact of PL toward improving educational equity for all students reaches the audiences best positioned to enact evidence-based practices and tools to support change efforts.

CHALLENGES IN THE FIELD

Variation in Data Collection across District Partnerships

A fourth challenge in moving toward better and more standardized data collection on PL is **inconsistency across school and district partnerships**, both in terms of what gets collected and how the data are shared.

To ensure they are responsive to the needs of school communities and philanthropic expectations, PL providers often use different measures with different partners to capture their impact. Consequently, inconsistent measures present challenges when comparing and evaluating service effectiveness across various schools and districts. Not only is this due to differing partnership goals or existing school/district measurement systems, but also because PL providers face increasing challenges when using common survey and observation instruments due to laws and regulations prohibiting the use of certain words or concepts within the educational setting. For example, words or phrases like "inclusivity," "diversity," "community," and "social and emotional learning" are prohibited in some states or districts. Measures are revised accordingly, replacing phrases like "attends to students' emotional, intellectual, or physical well-being" with "attends to students' well-being." In some cases, the edits substantially alter the meaning of a survey item or observational indicator. The priority is placed on providing high-quality PL while recognizing that measures may not align with data collected elsewhere.

PL organizations also encounter significant obstacles when accessing and analyzing student-level data from their partners. Accessing student data typically requires signing data-sharing agreements with districts, which can be time-consuming and cumbersome. Data is often shared on timelines that allow for impact analyses but do not allow for ongoing monitoring of program effectiveness. Moreover, while the process is sometimes expedited when student data disaggregation (which allows for quicker data access) is not requested, it potentially leads to less meaningful analyses. For example, assessing equity in educational outcomes requires the disaggregation of data by teacher and student attributes, such as gender, race/ethnicity, and language in order to determine which groups of teachers and students are benefiting most or least from PL services. This challenge is further complicated by the fact that students' academic outcomes are often narrowly interpreted as scores on standardized assessments. Widely recognized and accepted measures and rubrics that assess student growth beyond test scores do not currently exist in the field. The good news is, PL organizations are working toward addressing this challenge. TL, for example, does collect and score student work as an alternative (if test scores aren't available) or complement (when test scores are available) to achievement tests.

CHALLENGES IN THE FIELD

Capacity Constraints

Lastly, members of our working group highlighted **internal and external capacity issues** that hinder attempts to collect and analyze data that informs PL efforts.

For instance, continuously collecting survey data can burden participants and lead to measurement fatigue. The collection of observational data requires training and re-training observation raters, ensuring consistently high inter-rater reliability, and conducting observations during times of the school year when classrooms are operating in "business as usual" mode (i.e., not within the first six weeks of school start, not during state testing windows, not directly before or after long breaks), which is therefore typically high-effort, high-cost, and time-constrained. Other capacity challenges include building the measurement and analysis capacity of school partners, data privacy infrastructures that are often prohibitive when collecting or sharing individual student data, the implications of staff turnover, and training and norming measures among program staff, particularly raters. For example, due to the PL provider's proximity and established relationships with the school partner, data collection often becomes the PL provider's responsibility. It adds to or even competes with their other responsibilities for time and resources. Additionally, introducing an observation instrument that differs from the state, district, or school's tool can create questions and confusion regarding the expectations for teaching excellence. While such conversations may be fruitful and even important, they may divert time and energy from the contracted services.

CASE STUDY

Teach For America

Addressing Internal and External Capacity Issues

Teach For America (TFA), which recruits and develops a diverse corps of leaders who make an initial two-year commitment to teach in high-need schools, created a **measurement instrument that allows them to assess how teacher training on students' experiences of learning conditions is associated with improved educational outcomes**.

TFA began exploring Cultivate for Coaches (Cultivate), a survey and framework developed by Dr. Camille Farrington and the University of Chicago Consortium on School Research (Consortium), because it aligns with the TFA's coaching model as it requires embedded coaching support for individual teachers (1:1 or in professional learning communities) using actionable tools and strategies. Cultivate is also aligned with a progress monitoring instrument, Elevate by PERTS, allowing teachers to monitor progress between survey administrations as they test new methods and techniques.

Following positive results from a small pilot, a team at TFA established a research-practice partnership with the Consortium to fully embed Cultivate into the work of TFA. This partnership was centered on the conviction that Cultivate's year-long system of support cannot simply be an add-on to TFA's programming but instead needs to be authentically embedded and aligned across TFA's system of supports, including current programming for corps members (CMs) delivered during pre-service (the summer before entering the school year).

Teach For America

→ READ FULL CASE STUDY

Organizations face challenges in accessing student data and in making sense of the data when obtained. The challenges are partly due to the misalignment between classroom-based assessments (CBAs), districts' PL visions, and PL organizations' primary objectives. Additionally, there is currently no guidance or support in the field for analyzing student data related to PL services, not to mention the question of how much time and intensity it requires for a PL organization to work with a partner to see shifts in student data, and what student data to specifically target. For example, PL organizations often collect student data, but districts are unable to share teachers' names. Therefore, organizations cannot compare teachers they supported versus teachers they did not support. Organizations also run into issues when meaningfully analyzing the data longitudinally when datasets are not necessarily linked from year to year (e.g., to follow the same students' trajectories). PL organizations require overarching evaluation frameworks to guide their measurement cycles to determine what to measure to evaluate PL effectiveness and what to measure when. Many of our organizations use the Guskey Framework (2016) as a starting place for this work since it provides a clear vision for the types of data our organizations might collect.

WHAT'S MISSING IN THE FIELD

→ Opportunities to Innovate

While we recognize that there will never be a single perfect set of measurement tools that captures the needs of all our organizations and partners, we are optimistic about the possibilities for innovation in the field. RPPL is positioned to help convene PL stakeholders to identify, share, and enact the features of PL that improve equitable instructional practices shown to positively impact students' classroom experiences, well-being, and academic growth. We see tremendous opportunity to address some of the challenges listed above by focusing on equitable and practical measurement, combining measures that elevate marginalized needs and voices with a process orientation around quick turnaround data for continuous improvement. In the following sections, we outline in more detail areas of opportunity that we want researchers and practitioners to take up and act on in the coming years.

Content of the Measures (i.e., the "what")

We need a greater **library of measurement tools that prioritize student well-being across cultures and contexts**, allowing for research on student outcomes outside of traditional measures of academic achievement. While assessment scores help measure some aspects of teacher practice and student learning, our organizations resoundingly believe that when measuring the success of PL, the field should also consider student outcomes beyond test scores. A challenge of PL research is that not enough teacher PL attempts to measure student outcomes and instead focus on earlier Guskey levels. What, then, are measures of student achievement and teacher practice that we could consider beyond test scores? How can we think more broadly about what student and teacher outcomes mean beyond student achievement?

According to Gutiérrez (2009), the field often views equity as separate from excellent teaching, rather than conceptualizing excellence in teaching *with* equity. Measures such as the Panorama, Tripod, Leaps Student Voice Survey, Cultivate for Schools, and Elevate surveys help get us closer to that conceptualization and can provide meaningful information when aligned to PL objectives.

For example, Teach For America (TFA) is administering the Cultivate survey, which asks students in grades 5-12 about their learning conditions (e.g., meaningful work), learning mindsets (e.g., growth mindset), learning strategies (e.g., organization and time management), and motivation in their classroom. In between the fall and spring survey administrations, teachers have the option to administer Elevate, an aligned progress monitoring tool, to support continuous improvement of the classroom conditions they've prioritized based on the Cultivate fall results. Using the two instruments together informs an organization-wide metric on the impact of PL on equitable classroom conditions and supports teacher improvement in real-time. This approach represents a shift from focusing solely on changing how students respond to and navigate pre-existing classroom conditions and instruction to concentrating on systemic changes that are core to creating more equitable learning environments.

There is also a need to **couple observational data from the observer's point of view with direct evidence from students' experiences of the learning environment, with a focus on historically marginalized student groups**. Because classroom observation instruments generally are not sensitive enough to inform us about the learning environment for marginalized student groups that appear in small numbers within classrooms, Wilson (2022) developed classroom observation rubrics known as the Equity and Access Rubrics for Mathematics Instruction (EAR-MI) aimed at identifying and observing instructional practices that support equity and access in successful mathematics learning environments. In this process, Wilson highlights the importance of designing rubrics to capture practices that promote conceptually oriented mathematics instruction while also addressing dimensions of equity outlined by Gutiérrez (2009), including access, achievement, identity, and power. For example, rubrics often ask raters to code lessons according to a 4-point scale, where the distinction between a 3 and a 4 is whether "most students" or "all students" are acting in the desired

way, and a 3 is considered proficient. What if there are one or two students who represent historically marginalized student populations, and those are the ones not included in the "most" or "all"? We don't have the information needed from measures to interrogate data in ways that allow us to drive equity in the learning experience for students.

Additionally, our organizations share a need for observation tools that can be 1) easily normed across schools and raters, whether used virtually or synchronously, 2) easily used to train school leaders to observe instructional practices and students' experiences of the learning environment; and 3) used by schools and districts when PL organizations are no longer providing support. Existing observational tools that focus on teachers' CRSE practices, such as the Culturally Responsive Instruction Observation Protocol (CRIOP), are complex, require significant norming, and often require much more information than what is observed in a single lesson.

Finally, organizations' observation tools should align with the district's observation tools. Organizations such as New Teacher Center (NTC) and Achievement Network (ANet) regularly create crosswalks to show how their measurement instruments, such as Student Achievement Partners' (SAP) Instructional Practice Guides (IPGs) are aligned with other tools used in their partner districts (i.e., Danielson Framework, T-TESS, etc.).

Triangulating data sources is also critical when using surveys and other **self-report measures for teachers and students**. A common concern with some survey measures is that they are biased. For example, Leading Educators (LE) uses an organization-developed survey with its partners regarding the school and system enabling conditions that support effective PL for instructional improvement. Teachers and leaders at various levels share their ratings about the extent to which each condition is in place in their school or system. Analyzing such data can come with worries of inaccuracy or biased individual perceptions. However, surveys also offer leaders, teachers, and students a chance to share their voices and gather direct data from them. Also, when survey content includes demographic information, it provides an opportunity to disaggregate information and check for any meaningful differences in sub-groups that could reinforce inequities.

CASE STUDY

New Teacher Center

The Learning Environment: Measuring Beyond Teachers' Instructional Practice

New Teacher Center (NTC) recognized a need to supplement existing tools, like the Instructional Practice Guide (IPG) or the NTC-enhanced version created in partnership with Student Achievement Partners (SAP), with **observation indicators that would address classroom conditions to support social and emotional learning and learning (SEL) differences**. In 2017, NTC convened a panel of leading researchers/experts in SEL to participate in a series of conversations over several months to articulate the conditions that constitute an Optimal Learning Environment and observable indicators of those conditions. Together, NTC crafted a classroom observation rubric piloted in several U.S. sites. NTC then sought input from additional expert reviewers with specialties in SEL, learning differences, and Mind Brain Education (MBE) research.

The rubric was further refined and launched nationally in 2018 with partners who use NTC tools. The rubric includes indicators of teaching practices, student actions, and classroom interactions, and NTC crosswalked it with most major observation tools (e.g., Danielson Framework for Teaching) to demonstrate alignments that exist. Due to project delays, the rubric's psychometric properties have not been established, but NTC welcomes the opportunity to examine the reliability and validity of the instrument.

NEW TEACHER CENTER

→ READ FULL CASE STUDY

CASE STUDY

Leading Educators

Supporting Systemic Coherence and Impact

Leading Educators (LE) exists to ensure excellent and equitable teaching for all students. Successful implementation of **PL systems focused on equity and excellence requires considerable alignment and coherence within and across schools, which Leading Educators defines as a set of enabling conditions: a clear, widely-held vision for high-quality teaching and learning, curriculum and assessment materials aligned to that vision, skilled instructional leaders to guide the learning, resources (e.g., time) to support it, and the data to understand its impact and to make adjustments.**

Few districts have yet had the support necessary to achieve consistent levels of coherence by putting these conditions in place, which threatens effective implementation of collaborative professional learning. Additionally, focusing on these conditions at the school and district levels is a way of ensuring alignment and equitable resourcing across schools while also creating the opportunity for teachers and leaders to learn from each other through a collectivist (systems) approach rather than individualistic (personal) approaches to improvement.

Leading Educators

→ READ FULL CASE STUDY

In order to realize the promise of survey data while also striving to **disrupt bias, students and teachers should be at the center of measurement**. Nicole Joseph and Elizabeth Anderson's new project, Measuring Inclusive Constructs of Mathematics Identity (MICMI), exemplifies this concept by interviewing students and creating measures of Black girls' mathematical identities. In doing so, MICMI provides timely and actionable information to educators on the intersecting identity created when race, gender, and disciplinary belonging converge. With this in mind, LE now analyzes not only stakeholders' survey data but also their internal team's own assessment of conditions as well as several conditions-related artifacts. This combination of evidence helps paint a fuller picture that supports leaders in understanding areas of convergence and divergence. This process supports stakeholders to build consensus about their current state with respect to the enabling conditions as well as to identify priorities for improvement. Triangulating survey responses against PL organizations' insights and relevant artifacts is a way to mitigate issues of bias. TL also engages in a comprehensive, mixed methods assets and needs assessment on partners' enabling conditions and similarly triangulates across different data sources, including teacher and school leader surveys, teacher and school leader interviews and focus groups, artifact analysis, and student work analysis in order to more objectively determine partner needs and growth areas.

OPPORTUNITIES TO INNOVATE

Properties of the Measures (i.e., the "how")

Content shifts alone won't solve these measurement problems; there is also **significant room for improvement in the types of standardized tools** that are available to organizations.

First, the field needs more measures that are **short, concise, and mutually beneficial**. Schools and school systems generate and use vast amounts of data, requiring considerable time and resources to collect and analyze. According to Nelson et al. (2013), standardized testing and lesson preparation occupy as much as 18% of instructional time. Therefore, any additional data collection can feel burdensome if teachers and students already feel over-surveyed and assessed. Researcher-designed measurement tools often sacrifice efficiency for nuance, comprehensiveness, and reliability, making it difficult for practitioners to utilize them. To combat this, the Lastinger Center for Learning is in the process of developing a new professional learning system product called the Math Matrix that includes the Algebra Screening and Progress Monitoring (ASPM) procedural and conceptual measures. These measures were designed to be completed within 5-10 minutes, administered quite frequently across a school year, easily scored by classroom teachers, and to provide useful, and timely, information to both teachers and researchers about student learning progress. Moreover, technical reports from the ASPM project have provided a strong validity argument for their use for these purposes. Instruments like this could help create stronger partnerships between PL providers and schools/districts as they support the needs of both partners to be short, concise, and mutually beneficial.

Measures that are lightweight, flexible, and easy to implement are therefore needed as they lend themselves to easier disaggregation and tracking change over time. For example, measures that are validated at the construct level, or are abbreviated, such as Transcend Education's Leaps Student Voice Survey, allow teachers to pull certain constructs of interest from a larger survey. Specifically, the Leaps survey is highly customizable and allows administrators to choose from 11 different survey scales. Validated educator surveys would be of great help to the PL field because, despite these affordances (being short and concise), schools and districts are often still challenged in finding ways to collect such data due to conflicting needs and goals of their own. Thus, a mutually beneficial tool to support their ongoing needs as well as researchers' evaluation purposes can help build bridges to effective data collection.

CASE STUDY

Student Achievement Partners

Prioritizing Equitable and Culturally Responsive Approaches to Measurement

STUDENT ACHIEVEMENT PARTNERS

Note to our readers: SAP is not a "PL provider" like our other working group organizations; they are a systems provider. SAP works at the systems level with districts, agencies, other nonprofits, etc., which impacts their approach to evaluation. RPPL invited SAP to join our working group to highlight their work incorporating equitable and culturally responsive evaluation methods into how they measure the impact of their work and how they operationalize the shifts in their organization's grounding vision/mission related to equity when defining and measuring impact.

Historically, Student Achievement Partners (SAP) has been deeply committed to ensuring that all students, no matter who they are or where they live, are supported to access and successfully engage with grade-level literacy and mathematics content in the classroom. This commitment has resulted in the creation of resources like the Instructional Practice Guide (IPG) and Instructional Materials Evaluation Tool (IMET), which have been used nationwide by educators, systems leaders, and other nonprofits for over a decade.

SAP has now built on this foundation to help educators design instruction that leverages the assets and honors the brilliance of students historically underserved and marginalized by our education system. This next phase of work is being defined by their e^2 Instructional Practice Framework™, which redefines high-quality instruction as being on grade level, joyful, culturally responsive-sustaining, and linguistically sustaining. The e^2 Framework is as much an internal roadmap for their work as it is an external resource for educators and was the culmination of years of their team's internal reflection and learning. During that period, it also became evident to SAP that refocusing their work's content alone was insufficient. They needed to redesign how they engaged in project work with partners and how they measured the impact of their work in the education field.

For the latter, SAP turned to methodological and measurement approaches that aligned with their mission and ensured they were engaging educators and communities in the same ways they envisioned educators engaging with students—leveraging their assets and honoring their communities. Following their pre-established approach to project work, SAP interviewed students, caregivers, and educators to understand how they defined success broadly and specifically in the context of SAP's work. SAP also interviewed their staff to understand the impact they wanted to have on their project work, how that reflected the organizational statement of impact and progress measures, and what challenges and opportunities existed at the time for measuring the impact of their work. SAP then sought to learn from approaches to evaluation that aligned with their general equity-focused orientation to project work—namely, the Culturally Responsive Evaluation and the Equitable Evaluation Framework—and spent time orienting all their staff to this way of designing evaluation plans, regardless of their role.

→ READ FULL CASE STUDY

Better practical and equitable measures would not only provide value to organizations such as ours, but would also take seriously the need to **honor students' and teachers' voices in the process** by returning their data back to them in useful and timely ways. We can show teachers and students how their voices matter by providing results to them, summarizing what we heard in their responses, and sharing plans to act on the findings. Such measurement tools should treat participants not just as a source of data but rather as key stakeholders who should be included in meaning-making and action-planning phases of evaluation efforts.

For study participants and our organizations, the **timeliness** of measures is key to providing real value. Teachers need practical tools that will support them to make short-term or immediate adjustments in their classrooms. Measures that minimize the time between data collection and getting results back into teachers' hands will enable them to be more responsive to the needs of their students. When it comes to data on students' engagement and their classroom experiences, teachers often have an entirely different class of students by the time data from one school year are analyzed and shared. At that point, it is much too late for them to gather additional information from their students and adjust their instructional practices. Recognizing the importance of timely data, LE designed its Teaching For Equity student survey collection and reporting system to allow teachers to get their results back within one week. This quick turnaround empowers teachers to share with their students what they heard and create a plan of action for changing their practice as a result of the data.

Finally, we would also like to see **measurement libraries that include different sets of tools for different ages of students**, enabling developmentally-appropriate data collection. Ferguson (2000) shows that student surveys are predictive of student achievement despite demonstrating significant variation among classrooms. However, too often, measurement tools such as student surveys consist of only one version written at a high school reading level. In reality, other grade bands require different assessment strategies. Therefore, measures should target, account for, and address variation in diverse classrooms. As an example, LE designed a process for gathering data on students' perceptions of instruction and classroom experiences directly from the students themselves. Students in grades 3-5 and 6-12 independently respond to one of two versions of a self-report survey designed using grade-level appropriate terminology and concepts. Younger students in grades K-2 participate by completing an individual interview designed to be developmentally-appropriate and proctored by one of their teacher's colleagues. Further, these differentiated surveys and the interview protocol have been translated into Spanish to make them accessible to multilingual students. For student surveys, organizations also need developmentally-appropriate benchmarks for growth and improvement that align to students' age, grade level, race, ethnicity, etc. Given some trends PL organizations have seen in student mindsets and social-emotional development, we need more information about how much change over time is typical. Although some validated student surveys like Panorama or Tripod have versions for lower grade bands (grades 3-5), some PL organizations need more resources and capacity to administer them since they need to be read to younger students. This issue is magnified if schools do not yet have the routines in place to gather and analyze this data on a regular basis.

Conclusion

Throughout our working group's exploration of measurement challenges that PL organizations face, we **outlined the challenges that PL organizations face, described the types of measurement properties and focus areas needed, and provided considerations for individuals and organizations associated with teacher PL.**

Through this discussion, two themes emerge across our organizations' needs:

1. **First, there is increasing demand for measurement tools that better situate practitioners to directly address equity in their work.** Often, the primary focus of PL impact research and evaluation is students' academic outcomes. While important, there is growing demand to see how student and teachers' experiences, CRSE, and SEL are related to academic performance.

2. **Second, there is tension between standardized measures and context-specific measures.** PL organizations recognize that school and classroom contexts matter in measurement work, but highly specific measures complicate their ability to secure data-sharing agreements, collect data, make comparisons, and scale PL services.

The first theme surfaces as attention to students' identities and individual classroom needs increases in the field, while the second is derived from the first. Equitably monitoring student progress for continuous improvement is top of mind for practitioners and PL organizations alike. However, as discussed, context-specific measures can lead to inconsistencies across classroom, school, and district partnerships, making them difficult to validate. Although equitable measures enable responsiveness and attending to the needs of different school communities, inconsistent measures trouble efforts to scale and sustain PL. Additionally, the use of different measurement tools by practitioners and researchers further complicates this challenge. The disconnect between the tools practitioners use and the instruments researchers develop only exacerbates the tension between standardization and context-specific measures.

Where We Go from Here

One of RPPL's goals over the next several years is to continue to build the evidence base in the field by creating a standardized set of shared measurement instruments across PL organizations and a data infrastructure that houses equitable and practical tools to broaden access among researchers and PL organizations.

We also see ways that each group mentioned in this white paper can contribute individually to the broader goal of stronger measurement to support more equitable and effective professional learning:

1. PL ORGANIZATIONS

PL organizations can collaborate regularly to share effective measurement strategies and seek opportunities to align around shared measurement in service of a broader cross-organizational learning agenda (i.e., through RPPL). Building stronger collaborations with other PL organizations will help them understand how others have tried to address their current measurement challenges and provide ideas for what measures they use where, how, and why. In addition, PL organizations should consider how they can strengthen the capacity of their district partners to align their measurement approaches to educational outcomes they seek to change collectively and the constructs and tools they employ to assess headway toward those changes.

2. RESEARCHERS

While organizations surfaced the practical challenges of measurement, it is also essential to acknowledge that the evidence base for measurement in this space is scarce. Researchers should leverage and mine existing data sources to see, test, and improve underutilized measures already in play. In addition, continued research to better understand the technical and substantive challenges of measuring teacher PL is needed. As noted, however, such research must consider the challenges practitioners face and the contexts in which they operate. To do this, researchers, PL organizations, school and district leaders, and teachers must develop deeper partnerships to ensure alignment when developing evidence-based tools. Researchers should also broaden their measurement efforts to include additional important sources of info (e.g., interviews, focus groups, artifact reviews, etc.). As we illustrate in this report, many of these tools already exist in the field, but better use and dissemination on the part of researchers is necessary.

3. SCHOOL PARTNERS

Sustaining and scaling PL requires partnerships with school districts to understand how measurement tools can and should meet their needs. When schools understand and state the value of using practical and equitable measures, especially in helping educators improve their practice and classroom environments in service of equitable student outcomes, it supports research-practice partnership (RPPs) efforts and sparks further engagement of researchers and funders in the work. We recommend school partners engage PL organizations around data strategy with the expectations that:

- They will need to build a data pipeline between their district and their partner organizations,
- There will be some new data they are not currently collecting that will need to be added to understand the impact of, and direct, the partnership, and
- There is a value-add of the PL organization leveraging common data from its other partnerships as a point of comparison and benchmarking.

4. FUNDERS

PL organizations require funding support to build the data capacity of their organizations to develop and validate measures that meet the needs we have identified. Funding can support and incentivize researchers, organizations, and districts to design, pilot, train on, and collect data to help ongoing contextualization and construct validation of measures. A clear example of a funder's role in championing specific measures in the PL field is the increased use of SAPs' IPGs. PL organizations' uptake of the IPG can be primarily attributed to funders supporting SAP in rolling it out and training PL organizations to help calibrate, pilot, and facilitate the relationships between the tool's developers and IPG users. Organizations need the resources to learn and understand tools to encourage wider use, and ideally, tools that are meant for that purpose and have been designed in that way.

5. POLICYMAKERS

Allocate federal, state, and local funds that specifically support work to address measurement challenges and attempt to build measures in response to these challenges. For example, the Institute of Education Sciences (IES) has funded many measurement studies, but measurement has shifted dramatically over the past 30 years and will continue to do so as the field's measurement priorities shift.

In addition to the recommendations, our working group developed a set of reflection questions for our stakeholder groups to consider as they continue to think deeply about the future of measurement in teacher PL.

PL Organizations: How can you collaborate with other PL organizations to address measurement challenges and reduce capacity constraints? How can we collectively determine the most critical constructs to measure to understand changes in teachers' instructional practices and students' experiences?

Researchers: What supports and partnerships might allow you to build greater alignment with stakeholders in the development of evidence-based tools? What kinds of trade-offs are required to develop more lightweight and modular measures around some of the constructs described in this paper?

School Partners: How might you lead your district and school teams to adjust the ways in which you measure interim and impact outcomes in more practical and equitable ways? What role could families and communities play?

Funders: How can you support and incentivize researchers, organizations, and districts to design, train, pilot, and validate measures?

Policymakers: What progress have we made in the field of measurement? Where can we improve? Where do you think you can politically make the most movement and/or gain the most traction and buy-in? How can policy facilitate the flow of data between states, districts, schools, and PL organizations?

→ Given our organizations' commitment to improving teaching and learning and supporting engaging and affirming learning environments for *all* teachers and students, it is critical that we, as a field, can meaningfully understand the impact of our PL services on teacher and student experiences and educational outcomes. To do this, we must have practical and equitable measurement frameworks and tools we can leverage to support this work. This paper serves as a call to action for *all* actors in our field to collectively shift how we think about, develop, deploy, and fund measurement efforts in teacher PL. It is a call to build on and deepen our current knowledge base to move us closer to our ultimate goal—a world where every child, in every classroom, learns rigorous content and thrives in an equitable learning environment that supports their academic, social, and economic advancement.

Citations

Achievement Network. (2023). *Empowering student and educator voice.* Medium.

Achieve the Core. (n.d.). *Common Core Knowledge and Practice Survey*. https://achievethecore.org/page/1104/common-core-knowledge-and-practice-survey

Annenberg, Brown University. (n.d.). *EdInstruments*. *https://edinstruments.com/*

Carnegie Foundation for the Advancement of Teaching. (n.d.). *Practical Measurement for Improvement Resource Library.* https://www.carnegiefoundation.org/pmi-home/resources/

Chicago Public Schools. (n.d.). *Cultivate Survey.* https://www.cps.edu/strategic-initiatives/cultivate-survey/

Equitable Evaluation Initiative. (n.d.). *The Equitable Evaluation Framework.* https://www.equitableeval.org/framework

Ferguson, R.F. (2000). Can student surveys measure teaching quality? *Phi Delta Kappan*, *94*(3), 24-28.

Guskey, T. R. (2016). Gauge impact with 5 levels of data. *Journal of Staff Development*, 37(1), 32-37. https://tguskey.com/wp-content/uploads/Professional-Learning-1-Gauge-Impactwith-Five-Levels-of-Data.pdf

Gutiérrez, R. (2009). Framing equity: Helping students "play the game" and "change the game." *Teaching for Excellence and Equity in Mathematics*, *1*(1), 4-8.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In Newcomer, K. E., Hatry, H. P. & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 281-317). Jossey-Bass.

Iowa State University. (n.d.). *Algebra Screening and Progress Monitoring.* https://education.iastate.edu/soe-outreach/algebra-screening-and-progress-monitoring/

Joseph, N. M. & Anderson, E. (2023). *Project: Measuring Inclusive Constructs of Mathematics Identity (MICMI).* Advanced Education Research and Development Fund.

Lastinger Center for Learning, University of Florida. (2023). *Math Matrix.* https://lastinger.center.ufl.edu/lab/math-matrix/

Lothians Equal Access Programme for Schools. (n.d.). *LEAPS Student Survey.* https://leapssurvey.org/survey/student

Nelson, H. (2013). *Testing more, teaching less: What America's obsession with student testing costs in money and lost instructional time.* American Federation of Teachers.

Panorama Education (n.d.). *Panorama Student Survey*. https://www.panoramaed.com/panorama-student-survey

PMRR. (n.d.). *Practical Measures, Routines, and Representations.* https://www.pmr2.org/

RAND Corporation. (2023). *American Mathematics Educator Study Survey, Spring 2023.* Rand Corporation. https://www.rand.org/education-and-labor/projects/aep/surveys/items/american-mathematics-educator-study-survey-spring-2023.html

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, *23*(7), 1086-1101.

Student Achievement Partners. (n.d.). *Instructional Practice Guide.* https://achievethecore.org/page/1119/instructional-practice-guide

Texas Education Agency. (2023). *Texas Teacher Evaluation and Support System (T-TESS).* https://teachfortexas.org/

The Danielson Group. (n.d.). *The Framework for Teaching.* https://danielsongroup.org/the-framework-for-teaching/

The Project for Education Research That Scales. (n.d.). *Elevate: Elevate student voice and learning.* https://www.perts.net/elevate

Tripod. (n.d.). *Tripod surveys.* https://www.tripoded.com/

We All Count. (n.d.). *The Data Equity Framework.* https://weallcount.com/the-data-process/

WestEd. (n.d.). *Math Practical Measurement*. https://mpm.wested.org/

Wilson, J. (2022). Initial steps in developing classroom observation rubrics designed around instructional practices that support equity and access in classrooms with potential for "Success". *Teachers College Record*, *124*(11), 179-217.

Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement.* Carnegie Foundation for the Advancement of Teaching.

rpplpartnership.org

@rpplpartnership

@rpplpartnership

team@rpplpartnership.org