VOLUME 32 | ISSUE 2 | AUGUST 2025

NCSM JOURNAL

OF MATHEMATICS EDUCATION LEADERSHIP

USEFUL, EASY, AND CONSEQUENTIAL: A PRACTICAL MEASUREMENT REPOSITORY TO ENHANCE THE WORK OF MATH EDUCATION INSTRUCTIONAL LEADERS AND TEACHERS

TEACHERS' DEVELOPMENT OF PROFESSIONAL VISION AND LEADERSHIP CONCEPTIONS IN AN ELEMENTARY MATHEMATICS SPECIALIST PROGRAM

LEARNING TO FACILITATE CONTENT-FOCUSED COACHING CYCLES: A COMPREHENSIVE FRAMEWORK TO SUPPORT COACHES' PROFESSIONAL GROWTH

ncsm Leadership in Mathematics Education –Leadership in Mathematics Education | www.mathedleadership.org

Table of Contents

USEFUL, EASY, AND CONSEQUENTIAL:

A PRACTICAL MEASUREMENT REPOSITORY TO ENHANCE THE WORK OF MATH EDUCATION INSTRUCTIONAL LEADERS AND TEACHERS

Kirk Walters
WestEd

Andrew Brannegan
WestEd

Angela Knotts
WestEd

Sola Takahashi
WestEd

ABSTRACT

To improve and inspire high-quality mathematics teaching and learning, teachers and instructional leaders need access to data that are meaningfully connected to practice. Although most schools and districts are inundated with data (e.g., annual state test scores, data from any number of interim assessment systems), these data are not always helpful in terms of making timely adjustments to instruction, teacher professional learning, and other crucial factors affecting student mathematics outcomes. In this paper, we discuss the potential of practical measurement to fill this gap and address tensions facing math leaders. Unlike most data-driven accountability measures, practical measures are easy for teachers and leaders to collect and interpret data, enabling teachers to adjust instruction in a timely manner. We provide a repository of practical measures leaders can add to their instructional tool belts, discuss how middle-grade mathematics instructional leaders have used the repository to promote continuous improvement, and outline considerations for leaders and coaches in using practical measures to support their ongoing work with math educators.

Introduction

Educators must leverage data and measurement to reflect critically on their work, inform next steps, and advance their practice. In a series of prior issues, editors of the *Journal for Research in Mathematics Education* argued the field of math education is unlikely to advance unless teachers and instructional leaders have access to data that are meaningfully connected to practice (Cai et al., 2018a, 2018b, 2018c; Cai, Morris, Hohensee, Hwang, Robison, Cirillo, Kramer, & Hiebert, 2020; Cai, Morris, Hohensee, Hwang, Robison, Cirillo, Kramer, Heibert, & Bakker, 2020a, 2020b). Without such data and a broader infrastructure to house and support the uptake of these types of measures, attempts to fuel instructional improvement at scale are likely to continue to miss the mark. Accountability-based data systems (e.g., benchmark or interim assessments) have not been timely or actionable enough for educators to make better instructional decisions. The editors also argued educators, and the field more broadly, deserve better. This argument aligned with NCSM's (2019) vision of the importance of providing structures and resources for instructional leaders and teachers to ensure students consistently have access to rigorous mathematics. The structures include mechanisms to (a) drive continual job-embedded professional learning (Essential Action 3 – EA3) and (b) collectively collect, analyze, and celebrate evidence of student learning (Essential Action 7 – EA7).

This paper describes a multiyear project that has begun to tackle this thorny measurement problem. The Math Practical Measurement Project, funded by the Bill and Melinda Gates Foundation, includes a measurement repository and associated use cases that are proximal to the classroom. The measures focus on the processes of teaching and learning and were designed to support continuous improvement. At a high level, practical measures are easy, useful, and consequential for educators. They were designed to enhance learning and continuous, practice-based improvement, as opposed to measures primarily used to enforce accountability (see Figure 1). The free online repository can be accessed at https://mpm.wested.org.

Figure 1
Practical Measurement

Practical Measurement is "the deliberate and routine gathering, analysis, and interpretation of information with the distinct purpose of enhancing the learning of system actors as they test changes and improve processes that are at the heart of their work" (Takahashi et al., 2022, p. 423). Measures are "practical" in that they can be collected, analyzed, and used in the daily work lives of practitioners.

They are also practical in that they reflect practice—they act as sensing mechanisms at the level at which work is carried out.

In what follows, we describe how practical measurement, which has been used extensively in fields outside education (often referred to as "process measurement" in fields such as health care), has begun to be used in education and why it is promising. Next, we explain how we, the research team, built the repository of practical measures, outlining the measures included and the organization of the repository, and provide suggestions for effective use of the practical measures. We conclude the paper with suggestions of continued opportunities for the field.

BACKGROUND

A Brief History of Practical Measures

Practical measurement grew out of quality improvement approaches in industry and health care (Provost & Murray, 2011; Solberg et al., 1997). These measures were used to break down organizational silos, prioritizing the work of frontline workers, who identified problems on the ground and were central to driving more systemic improvement efforts (Takahashi et al., 2022). A prominent example is Toyota, whose approach to continuous improvement helped the company become a highly respected global manufacturer (Morgan & Liker, 2020; Rother, 2009). Toyota's management principles focused on building quality throughout workplace systems, tasking everyone in the organization with being a quality control inspector. Under this system, problems identified by assembly line workers are flagged and solved collaboratively rather than waiting for the problem to emerge as a faulty component once the car has already been built.

Since 2015, researchers at the Carnegie Foundation for the Advancement of Teaching have argued practical measures and other aspects of continuous improvement from industries should be applied to educational contexts. In the auto industry, it is inefficient, and perhaps too late, to fix an interior engine bolt problem after the car has been built. In education, using test scores from the past academic year or even benchmark assessment data from the prior semester is similarly too late to deal with emergent issues in the instructional environment that necessitated improvement. Through networked improvement communities (NICs), communities of educators and researchers work together to solve well-specified problems through rapid inquiry cycles. The NIC model requires a measurement system that is aligned to the problems of practice and can be leveraged practically by educators in the network (Bryk et al., 2015).

A core principle of NICs is the belief that "[one] cannot improve at scale what [one] cannot measure" (Bryk et al., 2015, p. 111), meaning organizations seeking to create improvement (e.g., schools) must think carefully about the properties of measures that will allow them to learn in and through practice (Yeager et al., 2013). The measures that most centrally benefit improvement efforts are practical measures, measures taken directly from practice that are easy to use in working to change key processes (Bryk et al., 2015). Such measures are "practical" both in the sense that they are relatively easy to use and in that they are proximal signals of practice (A. Bryk, personal communication, October 10, 2014). Practical measures have five key features:

- They connect to aspects of systemic structures, processes, or norms believed to be critical to achieve an aim.
- They are meaningful to system actors who use the data.
- They are actionable.
- They are minimally burdensome to the system actors who use the data.
- They provide timely and regular information. (Bryk et al., 2015)

In short, practical measures are easy, useful, and consequential.

One of Carnegie's first NICs focused on improving outcomes for community college students enrolled in developmental mathematics, a course the NIC eventually redesigned (Bryk et al., 2015). For many institutions, most students are tracked into these non-credit-bearing development courses, yet very few succeed. Not succeeding in these courses creates a huge roadblock for students because college-level mathematics credit is required to transfer to 4-year universities or to pursue many occupational programs. This NIC, which later became the Carnegie Math Pathways (CMP) program, identified several key root causes of the program and, from that, established a working theory of improvement and associated change ideas to test and refine. One set of change ideas focused on development of productive student mindsets during their 1st week in developmental mathematics. For example, the NIC redesigned lessons with activities to build students' mindsets and created professional development sessions for teachers on the importance of developing a productive mindset. To know whether these and other change ideas were working, the NIC needed a coherent yet nimble measurement system. One practical measure they developed was a brief survey on student mindsets that teachers could implement easily in their classrooms. The NIC also collected other practical measures from teachers related to their professional development experiences and instruction. Such a system of measures helped the NIC carry out quick-cycle continuous improvement cycles as it worked to meet its larger aim of significantly improving outcomes for developmental mathematics students, which it accomplished (Hoang et al., 2017; Yamada & Bryk, 2016; Yamada et al., 2018).

Measurement for Improvement, Accountability, and Research

The emergence of practical measurement in education reflects the contrast between these measurement tools and practices and the tools and practices that predominate the U.S. educational system and are designed for accountability or research purposes (Solberg et al., 1997; Takahashi et al.,

[1] To hear more about how Toyota's approach compares with other manufacturers, including a General Motors plan in California, see This American Life's NUMMI episode (Langfitt & Glass, 2015).

2022). Accountability measures often have been used to judge summative performance or inform broader policy decisions retrospectively, whereas practical measures have been designed to stimulate rapid improvement and are not punitive in nature. Practical measures prioritize the need for school- and district-based practitioners (and practitioners in other types of educational organizations) to have regular and timely data that serve as quick measures to access feedback on how continuous improvement efforts are progressing before the end of a course or school year (Jackson et al., 2016). Accountability and research measures commonly used in the education system frequently have been time and labor intensive to collect, typically have been collected only after the end of some cycle (e.g., a school year) when those impacted can no longer benefit from the data, and have been tied to global measures of outcomes resulting from such a complex system that causes of those outcomes cannot be tied meaningfully to specific practices delivered at a specific time. Data collection techniques that keep teachers and learners engaged in the work of teaching and learning, in contrast, can be an important resource for improving teaching and learning (Yeager et al., 2013).

This is not to say that the relationship between practical measures and accountability measures should not be examined if the improvement goal is aligned to one of these measures. In fact, this relationship is one of the key analytic means to evaluate whether practical measures capture information that is consequential (Takahashi et al., 2022). In the Carnegie NIC example, the student survey measure of productive persistence was sensitive to change ideas being tested, but student answer patterns also predicted end-of-course performance (Bryk et al., 2015). However, importantly, the practical measures used to drive continuous improvement were not used to evaluate student or teacher performance for accountability purposes.

Researchers sometimes develop measures to test theories; for example, a researcher might hypothesize elementary teachers' increased understanding and use of the double number line representation might improve their ability to teach problems involving fractions with unlike denominators. The researcher might design teacher- and student-facing materials with the double number line, an observation instrument, and an assessment of teacher and student knowledge. The primary purpose of these measures would be for theory development, not rapid cycle improvement. The researcher would assume most of the burden related to data collection and analysis. Eventually, the findings would be shared with teachers, but not before the research project had been completed, which could take months or even years.

Practical measures offer tangible benefits to both classroom teachers and educational leaders. These measures:

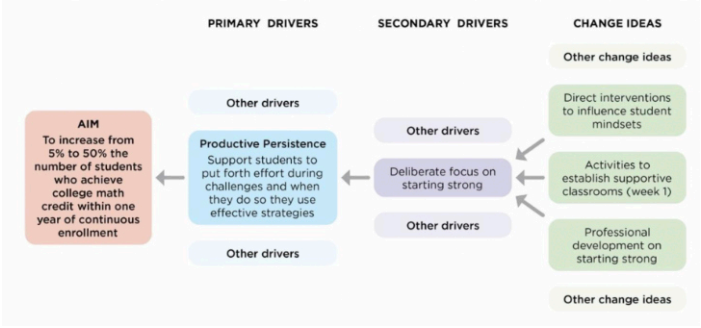
- shed light on how improvement efforts are going on a regular basis, not just infrequently or after the fact;
- bring discipline to the work of testing change ideas by providing data to confirm or refute teachers' and leaders' general sense of how things are going;
- focus attention on a particular challenge across a learning community (e.g., across math teachers at a particular grade level, across schools in a district);
- illuminate variation across a learning community and identify "bright spots" (e.g., schools or classrooms where a change idea seems to be effective) and instances where the same change seems to be less effective and additional support may be needed; and
- elevate the voices and experiences of people "closest to the problem," often students or teachers.

A Practical Example

Returning to the earlier example of developmental mathematics reform, the CMP program was designed to increase the proportion of students who achieve college math credit in 1 year of continuous enrollment. More specifically, the aim of this NIC was to increase the math credit achievement rate from 5% to 50%. As shown in Figure 2, the aim was part of a working theory of improvement called a driver diagram, which continuous improvement researchers use to organize their efforts to test and refine strategies to reach an aim. We present an excerpt of the driver diagram to illustrate connections between the aim, a working theory of improvement, change ideas, and practical measures. As shown, the NIC theorized productive persistence was a key driver in shifting student outcomes. They theorized that, by developing change ideas, including strategies to use as students began their math courses, students would develop healthy habits of mind and learn strategies to help them persevere when facing academic challenges.

Figure 2

Drivers and Changes Ideas of Productive Persistence (Excerpt)



Having a working theory of improvement is critical to solving complex problems; however, without practical measures, the theory might not inform concrete actions. For example, suppose a math department designed a 2-day professional development training focused on the importance of starting strong and included direct interventions to influence student mindsets and activities to create supportive classrooms. The department must determine if change ideas contribute to improvement. At a high level, department personnel would want to ensure the professional development was attended, received, and implemented well by faculty. This assessment could be accomplished through quick surveys and informal observations. Digging deeper, the department might develop surveys that measure students' attitudes toward learning and

[2] For a more detailed driver diagram and related discussion, see *Learning to Improve: How America's Schools Can Get Better at Getting Better* by A. S. Bryk, L. M. Gomez, A. Grunow, & P. G. LeMahieu (2015), pp. 75–79. Harvard Education Press.

their beliefs about persistence, which they could use to assess the effects of the change ideas. In fact, the CMP program found students' academic mindsets shifted in expected directions after 3 weeks of implementing associated change ideas (Bryk et al., 2015). The surveys proved to be timely, useful, and consequential for the NIC.

ABOUT THE MATH PRACTICAL MEASUREMENT PROJECT

The Middle Grades: A Crucial Time for Mathematics Success

In recent decades, math education researchers and policymakers have identified the middle grades (i.e., Grades 5–9) as critical years in the trajectory of math students, with Algebra I widely considered a gatekeeper to college- and career-ready math classes and beyond (Adelman, 1999; Finkelstein et al., 2012). Despite various types of reforms implemented over this timeframe, student achievement in mathematics has remained lackluster, with only about one third of U.S. eighth graders scoring at or above the proficient level on the most recent National Assessment of Education Progress report in mathematics (National Center for Education Statistics, 2023). Rates were even lower for Black, Latinx, and low-income students. U.S. 15-year-olds performed well below the international average on the most recent Program for International Student Assessment (PISA), a test of math problem-solving skills, but again, the results were even worse for underserved students (Organisation for Economic Co-operation and Development, 2018, 2022). The concerns were even greater in the COVID-19 global pandemic context, in which some researchers projected greater learning loss due to school closures compared to what had been seen due to "summer slide" in typical pre-COVID-19 conditions (Kufeld & Tarasawa, 2020). Given these disappointing outcomes and concerning trends, improving mathematics teaching and learning in middle grades has remained an important area of reform, including recent reforms that have drawn on continuous improvement approaches.

Practical Measures for Middle Grades Mathematics Improvement Networks

When teachers and instructional leaders have worked to improve math teaching and learning using a continuous improvement approach, they often have lacked access to a coherent set of high-quality measures to understand their progress and inform their next steps. Of course, teachers often use their own classroom data to inform instructional improvements in their day-to-day work (Rothkopf, 2009), but the means to gather information are typically informal, idiosyncratic, and not widely shared (Cai et al., 2020b). Math education researchers also have created measures that reach many classrooms, such as tools for elementary teachers to understand student approaches to solving arithmetic problems (e.g., Cognitively Guided Instruction; Carpenter et al., 1996, 1998), strategies to promote and assess student mathematical discourse (Smith & Stein, 2011), and various techniques and measures that fall under the broad category of formative assessment (Burton et al., 2018; Fennell et al., 2016). Yet, these measures are typically only accessible to practitioners who participate in the respective research projects or for purchase after the research concludes.

Reflecting this policy and reform context, the Bill and Melinda Gates Foundation launched the Networks for School Improvement (NSI) grant portfolio in 2018, which funded improvement networks focused on advancing middle and high school math and English language arts (ELA) outcomes for students in historically underserved communities. As the NSIs began to conduct their work, the foundation determined the overall project lacked a resource of existing practical measures (i.e., timely data about processes that are the focus of improvement and that can be collected in a minimally burdensome way with the least disruption to classroom learning) and a categorization of such measures in ways that are useful for educators and education leaders. The foundation partnered with WestEd to build the repository, strengthen the capacity of math-focused NSIs to incorporate measurement into their work, and develop two new practical measures aligned to focus areas of the NSIs (see Figure 3).

Figure 3
Goals of the Mathematics Practical Measurement Project (October 2020–June 2024)

1. to develop a repository of math practical measures;
2. to build the capacity of math-focused NSIs to effectively use measurement in their continuous improvement work; and
3. to develop two practical measures focused on improving middle grades mathematics for historically underserved populations.

Existing research supported grounding our approach to this problem in best practices of math teaching and learning, such as

- explicitly focusing on the development of conceptual understanding and providing students with problems that promote productive struggle (Hiebert & Grouws, 2007);
- giving students prompts to monitor problem-solving approaches and encouraging the use of multiple approaches (Woodward et al., 2012); and
- providing all students with regular opportunities to engage in mathematical discourse (National Governors Association for Best Practices, Council of Chief State School Officers, 2010).

In addition, the nature of this work lent itself particularly well to continuous improvement theories, such as the model for improvement (Langley et al., 2009); the Carnegie Foundation's model of NICs and their six principles of improvement (Bryk et al., 2015); the measurement work adapted, developed, and refined in the work of quality improvement in health care at the Institute for Healthcare Improvement; and others (e.g., Provost & Murray, 2022). We also sought to incorporate key lessons learned from a mature, math-focused NIC, The Better Math Teaching Network, which had been engaged in math instructional improvement work since 2015.

The vision for the project also was informed by interviews with six math-focused NSIs. Consistently, the NSI leaders and their data and analytic specialists discussed the difficulty of identifying and enacting timely and regular measurement for teachers and instructional leaders to inform continuous improvement in math instruction. They spoke of challenges in identifying measures aligned with the aims of the work and simultaneously provided a feedback loop for the changes being tested in the classroom. They discussed wanting to "make practice visible" to practitioners but not having the right data to do so. They also spoke about the challenges in getting teachers on board with measurement work, which often required additional time and effort (e.g., data collection) on top of teachers' existing responsibilities.

THE MATH PRACTICAL MEASUREMENT REPOSITORY[3]

To address the aforementioned challenges related to the field's lack of access to existing practical measures, we sought to develop an online repository to house measures that have been useful in practice; their associated instruments; tools of collection, analysis, and visualization; information about the validity and warrant of the measurement tools to inform continuous improvement; and, in many cases, written vignettes or illustrations of development and use. In particular, the project was designed to attend to measures and their uses particularly connected to increasing the success of Black, Latinx, and low-income students in middle-grade mathematics.

Developing the Math Practical Measurement Repository

To develop the repository, we performed a scan of the field adapting a 90-day cycle process to understand what measures researchers and educators were using for improvement in math. This process, adopted from its use in the health care quality improvement field, involved gathering information through a review of select literature and interviews with experts, broadly defined as experts of practice and scholars. Information was gathered and then synthesized and organized at 30- and 60-day points in time where external review and reflection enhanced the emerging product. We spoke with math education scholars and researchers, instructional leaders, and continuous improvement specialists with knowledge of data useful for instructional improvement. Interviewees included Karen Givvin (University of California, Los Angeles), Kara Jackson (University of Washington), Grace Kelemanik (Fostering Math Practices), Ann Edwards (WestEd), Skip Fennell (McDaniel College), and Christine Roberts (Tulare County Office of Education), among many others.

During our interviews, we asked the experts about the characteristics and use of their practical measures. We asked the following questions:

- How did the measure support an improvement effort?
- Who was using the measure, what did they learn, and how did they use what they learned?
- What contextual factors supported the effective use of the measure?

Practical measures included in the repository are useful in a continuous improvement effort, are relatively easy for practitioners to use and enact in their work, and are consequential in that they capture something that matters for the improvement of math teaching and learning. Although we encountered numerous measures that might be considered "practical," we did not include everything we encountered. Rather, we prioritized including measures that

- were clearly connected to math learning or best practices recognized by the field of math educators;
- had demonstrated evidence of support for instructional improvement, either at a single site or in a more comprehensive study;
- yielded actionable data;
- could be used across multiple contexts and settings;
- were easy to administer and produced data that were easy to analyze and interpret (e.g., could fit into educators' regular routines, required minimal training, were not overly time consuming); and
- were free or low cost.

Throughout the development and refinement of the repository, we received and incorporated feedback on the repository from our advisory committee and NSI leaders.

Measures in the Repository

The repository includes 18 measures, along with associated guidance. The modes of the measures include surveys (10), classroom observation tools (3), quick student reflections (2), artificial intelligence (AI)-powered apps (2), and a rubric (1). The content of the measures is organized by focus areas arrayed along the instructional triangle of teachers and students interacting with content, including socioemotional learning (student focused), teacher mindset (teacher focused), and processes of teaching and learning (the interaction of the three nodes) but also by guiding questions. These guiding questions provide measures for educators looking to understand certain topics, such as:

- What does mathematical discourse look like in our classrooms?
- How are students making sense of the rigorous math tasks they are given?
- What mindset and beliefs do students hold about themselves as learners in a math classroom?
- How are teachers experiencing feedback about their practice?

For each measure in the repository, there is guidance for users on how to use the measure in their improvement work; for most measures, there are vignettes that tell the story of educators using the measurement tool.

Three Snapshots of Measures in the Repository

Figure 3 provides descriptions of three questions faced by teachers and instructional leaders along with measurement tools in the repository that can be used to gain insights on

[3] The repository can be accessed at https://mpm.wested.org/

[4] The 90-day cycles were described in the handbook on the Carnegie Foundation for the Advancement of Teaching website: https://www.carnegiefoundation.org/improvement-products-and-services/articles/ninety-day-cycle-handbook/

these questions. Included in the measure descriptions is information about how they have been used in practice and the analytic infrastructure and social processes and routines that support measure use.

Figure 3
Repository Measures and Tools for Teachers and Instructional Leaders

Challenge faced by teachers and instructional leaders: **How are students participating in classroom discourse? Where are the opportunities for teachers to initiate more academically rigorous, student-led discussion?**
Measurement Tool
TeachFX: An AI-powered app focused on classroom discourse
Description
An app that measures the amount of teacher talk, student talk, student group talk, and silence occurring during a lesson
User
Classroom teachers and coaches
The measure in practice
As a 1st-year teacher and the only math instructor at a newly launched high school, Daniel knew they needed a way to focus their improvement efforts around classroom discourse and chose to use the TeachFX tool. For Daniel, the app's ease of use (just hit "record") and automatically generated data reports showing ratios of teacher-to-student talk time made it easy to integrate the tool into their routine. Importantly, Daniel emphasized that reports were a launching point to reflect more deeply on their practice: Visual displays of talk patterns allowed him to home in quickly on stretches of their class that were worth digging into. Daniel said: I really like how you can see where there are interesting blocks [of teacher, student, or group talk time] . . . from there, you can spend 10 minutes and get a lot of rich insights, without having to dive into the 60-minute-long recording and having to find those spots.
Analytic infrastructure that supports use of this measure
TeachFX is a full-service app that supports data collection, storage, and analysis. After downloading the TeachFX app to their device, the teacher opens the app before class begins and clicks "record." Afterward, the app emails the teacher a report of talk patterns for the class. The report shows the breakdown of talk and distribution of types of talk throughout the lesson. Teachers may work toward a goal of simply increasing the percentage of student talk over time, but teachers and coaches also can delve deeper into this breakdown to uncover talk patterns that suggest strong instructional practices.
Supporting the social routines of data sensemaking with this measure
• Reviewing talk pattern data can better support instructional improvements when paired with professional learning around high-level questioning; student discourse; and the relevant, authentic, and rigorous tasks that support high-quality student talk. • A school culture in which student voice is explicitly valued can set the groundwork for authentic and reflective engagement with TeachFX data. • Although individual teachers can look at data on their own, incorporating the tool into professional learning communities or using it as a tool to aid coaching conversations will support grade-level or schoolwide improvement. • TeachFX has seen success in schools that begin using the app with a small, enthusiastic cohort of teachers and then expand to the rest of the school as teachers become comfortable with the app. • School leaders should take care to communicate that teachers' individual data are private, and use of the app is voluntary. TeachFX data should never be used for evaluation purposes.

Figure 3 continued on next page…

Challenge faced by teachers and instructional leaders:

How are students experiencing learning through whole-class and small-group discussion? Where are there opportunities for teachers to enhance classroom discussion and foster access to rigorous mathematics?

Measurement Tool

Practical Measures, Routines and Representations (PMRR) whole-class discussion and small-group work in specific lesson student surveys

Description

Student survey about student experiences with the whole-class discussion in their classroom

User

Classroom teachers, instructional coaches, and district leaders

The measure in practice

To help understand if the instructional changes teachers were making around whole-class discussion were resulting in improvement, teacher–coach pairs embedded the whole class survey into their one-on-one coaching cycles. Teacher–coach pairs co-planned to set goals and select tasks around the whole-group discussion, administered the survey as part of classroom instruction, and used data from select items to inform their debrief discussions.

As an example, at the end of one teacher's coaching cycle, only about half of students responded "no" to the item, "Did you have trouble understanding other students' thinking in today's whole class discussion?" As the teacher and coach unpacked why this might be the case, the data point provided an opening for the coach–teacher pair to discuss one of the coach's observations: The teacher tended to rephrase students' thinking during whole-class discussion, and students tended to share without building on and making sense of their peers' thinking.

This measure also can be used by district leaders as a window into the quality of math instruction at scale. The results of these surveys across classrooms can provide leaders with insight into matching instructional coaches and teachers and targeting professional development experiences (Jackson et al., 2016). Critically, effective and meaningful use of these measures depends on their use for purposes of improvement, not for evaluation. The authors cautioned a collective inquiry stance is critical for use of these measures, which must be explicit and reinforced, especially when district leaders are involved.

Analytic infrastructure that supports use of this measure

Teachers administer the relevant survey either immediately following a whole-class discussion or small-group work or at the end of a lesson that incorporates a whole-class discussion or small-group work. Students can take the survey using paper and pencil or online via Google Forms (or another online survey platform) to allow for quicker representation of student responses.

To help a teacher assess whether a new instructional strategy has improved students' learning, the teacher–coach pair might compare responses from the end of the previous coaching cycle to responses at the end of the current coaching cycle. To best support teachers in making sense of why students responded the ways they did, survey data should be analyzed alongside student work, coach observations, and teacher reflections

Supporting social routines of data sensemaking with this measure

- Using surveys in the context of regular coaching or a professional learning community helps teachers make sense of their data and connect data to targeted instructional changes.
- To best make sense of student responses, survey data should be analyzed alongside other information, such as student work, coach observations, and teacher reflections.
- Positioning the surveys to elicit student feedback and voice can help users understand the survey as a tool for exploring practice rather than as an accountability or evaluation tool.
- When discussing survey results, users should bring an asset-based perspective and a willingness to reflect on their own practices to avoid data being used to reinforce existing perspectives. A context of ongoing professional learning can allow a coach or school leader to shape conversations about survey data and can prevent data from reinforcing problematic ways of characterizing students.
- If district leaders engage teachers around these data, this must be done in a context where it is abundantly clear that the goal is support and learning, not evaluation. Practical measures, including these, are not designed for evaluation purposes and should not be used as such.

Challenge faced by teachers and instructional leaders:

How do students feel about their identities as math learners? Where are there opportunities for teachers and school leaders to improve student attitudes toward rigorous mathematics content?

Measurement Tool

High-Tech High Mathematical Agency Improvement Community (MAIC): Student Agency Survey

Description

Student survey on perceptions of their agency, mathematical identity, and group work experiences in their math class

User

Classroom teachers, instructional coaches, and network leaders

The measure in practice

The MAIC, a network of K–12 schools in Southern California, wanted to understand how student-centered practices support mathematical agency and success. Students of MAIC teachers completed the student agency survey three times in the academic year. Teachers appreciated the way the survey data provided insights into student perceptions and needs. Furthermore, when data were disaggregated, teachers were able to focus on subgroups of students who were having adverse experiences or whose needs were not being met as well as others' needs. One teacher explained, "We really tried to visualize data according to very specific target groups of students, and to help us unearth our own biases and things we're bringing to teaching we didn't realize we had." This same teacher reviewed survey data alongside grades and test scores to see if "there were students who were not getting what they needed in [their] classroom." One teacher stressed that the culture of learning in the MAIC network was critical; they explained:

> Sometimes you see data, and it hurts . . . but because MAIC felt like a learning environment, I felt very comfortable sharing my data and wonderings about it with others. When you come at it with a lens of curiosity . . . that allows the conversation to move and be authentic.

MAIC network leaders also used these data to understand trends across teachers and schools involved in this work. For many of the concepts captured through this practical measure, network leaders saw improvements over time, but some concepts appeared more intractable. For example, math classroom status hierarchies among students have proven to be an ongoing area of challenge. Network leaders have used this information to focus on specific practices teachers can try in their classrooms.

Analytic infrastructure that supports use of this measure

The MAIC network developed a survey infrastructure that allowed teachers to administer surveys using a digital platform, connected individual student responses to the student information system, and generated reports and data visualizations for teachers. These data could be viewed by network leaders at an aggregate level to see overall trends. They also viewed data in a "small multiples" display that showed improvements at each site together, so they could know where there were "bright spots" and where teachers were struggling and needed more support.

Supporting social routines of data sensemaking with this measure

- Students should be informed as to why teachers are collecting the survey data. Without understanding the purpose of the surveys, especially for a survey meant to be given multiple times, students may resist or disengage from the survey leading to less accurate data and reducing students' sense of agency.
- The MAIC network provided important context and support for educators to analyze survey results and connect survey data to specific change ideas and instructional strategies.
- Teachers in the MAIC network used the network's data protocol to review their data reports and network-wide data together to identify trends, patterns, and longitudinal changes. Data review and discussion was coupled with a focus on practices and strategies around student math agency and growth mindset that were shared and modeled at MAIC convenings. Teachers could bring these strategies back to their classrooms and test them through a series of plan, do, study, act improvement cycles.

Reception of the Repository

Since the launch of the repository in 2021, it has been visited by over 5,000 unique users across 21 countries. Introduced to 16 leaders from the NSI in August 2021, all 16 said in a survey that it was very likely or likely the repository would be helpful to their network. Ten of the respondents said there were measures they may want to try using in their network, whereas the other six were unsure. At the following meeting with NSI leaders, 8 of 11 survey respondents reported they had looked at the repository since it launched, and 6 of 11 said they had shared it with someone else in their NSI.

USING PRACTICAL MEASURES EFFECTIVELY

Practical measurement is a tool that can be leveraged to powerful effect in a continuous improvement approach to systemic change. However, measures and data do not hold meaning in themselves; any measure only holds meaning in the social learning system in which it is used. Here, we highlight four key practices for math leaders looking to incorporate practical measures in their improvement efforts. These key practices derive from challenges faced by those who have used practical measurement in the field. We drew upon several sources to arrive at these four practices: the stories we heard through the 90-day cycle scanning process, including interviews with experts in the field; our communications with the NSI over the course of the Math Practical Measurement Project; and our own personal experiences attempting to leverage practical measurement in our support of K–12 school district improvement efforts.

Start With a Clear Theory of Improvement

Practical measures alone cannot improve math teaching and learning. To generate useful, actionable data, practical measures must be connected to a clear and specific theory that outlines how the team believes changes to the system will lead to improvements in service of an aim (Bennett & Provost, 2015; Bryk et al., 2015; Takahashi et al., 2022). A common pitfall in using practical measures is selecting or developing measures without first coming to consensus as a team regarding a clear and sufficiently detailed theory of improvement, including (a) what one is trying to achieve (i.e., aim statement); (b) what structures, processes, and norms of the organization must be improved to achieve that aim (i.e., drivers); and (c) what change ideas (e.g., classroom or professional learning activities and strategies) are planned to create the desired improvements via those drivers. Without going through the process of identifying drivers and making explicit how team members believe a change idea will lead to improvement, selecting and implementing practical measures is unlikely to produce interpretable, usable data regarding whether and how a change idea is helping an organization make progress toward achieving its aim.

The theory of improvement is the basis for identifying the right measures (Bryk et al., 2015; Takahashi et al., 2022). Measures can help to answer questions about whether changes have been enacted consistently across sites and over time, to what degree these changes are occurring with high quality, and whether the changes are leading to desired proximal and distal outcomes (i.e., goals). Educators and education leaders can use resulting data to gain insight into what is working well, for whom, and under what conditions. They can identify bright spots and areas of growth, focusing their attention where it is needed. A set of measures connected to a theory of improvement, also referred to as a "family of measures" (Provost & Murray, 2022), can illuminate the larger theory of improvement. In other words, educators and leaders can determine if the changes actually led to the hoped for outcomes and, if not, whether the theory needs to be revised.

Try to Avoid Developing a Measurement Tool From Scratch

Development of a practical measure can be a substantial undertaking (Jackson et al., 2016; Takahashi et al., 2022). These measures are meant to be practical to use, but that does not mean they are practical to develop. In fact, prioritizing usefulness of the measures and their ease of use requires work during the development phase that is typically not called for in development of research measures, such as involving potential users in the development process and studying how a measure is used in practice. Teams engaging in continuous improvement efforts should first consider measures already collected in their organization. School systems typically collect many kinds of data. Math leaders should consider if anything can be leveraged from extant data. These data may not be housed in core data storage systems in a school district or other education organization. In one project in which one of the authors was involved, data about students' attempts to revise their own work were gleaned from assignment-completion data teachers kept, which existed outside the larger learning management system where core district data were stored.

If relevant data do not exist in the organization, the next question the team might ask is if there are valid measures other practitioners in the field have used successfully (Walston & Conley, 2022), such as those that can be found in the math practical measurement repository. These types of measures might also be found in related improvement research focused on similar aims, drivers, or change ideas. Once potential measures have been identified, they (or elements of them) can be mapped to the team's theory of improvement. This mapping process allows the team to prioritize items for certain change ideas or drivers, identify which items may need to be modified to better match the team's work, and identify gaps where new measures are needed.

Minimize Measurement Burden

A crucial design feature of any practical measure is that it is minimally burdensome and that it ideally takes little to no time to collect the data, especially for staff engaged in the core work that is the focus of the improvement effort (e.g., teachers who are already stretched thin with responsibilities; Takahashi et al., 2022). This is particularly important considering practical measures should be collected on multiple occasions to see change over time. Burdensome measures will be abandoned over time if they are not

sustainable. Measurement burden can be minimized in the following ways:

- Focus on "just enough" data. What are the minimal amount of data that could give educators a signal their work is heading in the right direction or might need to be modified (or abandoned)?
- Use already-collected data rather than engaging in new data collection (as discussed previously).
- Reduce the data collection task (e.g., using a 3-item survey to check in with students rather than a full survey).
- Leverage technology and automation to ease data collection and analysis. A tool like TeachFX requires a tap on a phone to collect data. A tool like the PERTS Elevate survey automatically processes data so graphs summarizing survey results can be accessed in days. More examples like this have continued to emerge.
- Embed measurement in existing workflows. For example, if teachers are already entering data about student learning into their gradebook, this process could be leveraged to capture information valuable to the improvement effort at hand. If instructional coaches are already taking observation notes from classroom visits, these kinds of data may be leveraged.
- Make measurement meaningful to people who collect, give, and use the data. There is burden associated with the amount of time or resources data collection and analysis may take, but there is also the perception of burden affected by how meaningful individuals find the measures. If key organizational members (and here we include students and families) understand why data are being collected, know how those data are being used, and see data used to benefit teachers and students, they are more likely to continue to participate in data giving and data collecting. Data can be shared with teachers and students so they can partake in sensemaking and collaborate in development of next steps and actions.

A clear, well-specified data collection and analysis plan focused on just enough data and involving core interest holders is key to creating routines that ensure measurement activities occur in a sustainable way for ongoing learning.

Design Data Discussions That Enable Critical Reflections and Thoughtful Next Steps

The "rubber meets the road" concept reflects the work of practical measurement when educators collectively engage with data to make sense of it and use learning to reflect on practice and generate next steps. The meaning that comes from practical measurement happens in dedicated routines of social sensemaking. We highlight five factors that contribute to these sensemaking spaces:

- Structures and routines for data sensemaking conversations (Coburn & Turner, 2011). These routines include identifying who is meeting with whom, how often, for how long, and with what purpose. These collaborative spaces are where data sensemaking can find a natural home, but they require aligning schedules, protecting time, and prioritizing these conversations. This approach may involve creation of new routines or embedding data sensemaking in existing structures of collaboration.
- Identifying participants. Walston and Conley (2022) recommended teams keep in mind that the primary determining factor for participating in data sensemaking meetings is whether the person is affected directly by the condition the group is attempting to change and has the power to implement or enable any decisions resulting from data inquiry (e.g., teachers, students, parents). We would add, when teams are prioritizing equity work, it is particularly important to be mindful of interest holders who are traditionally "not at the table" or underserved by the system. For example, educators and leaders may examine how students and families can be included in these spaces to make sense of and improve the system intended to benefit them (Takahashi & Norman, 2025).
- Norms of openness, transparency, and innovation (Takahashi & Norman, 2025). Fruitful data conversations require participants to talk with transparency about what is not going well, perhaps even more than discussing what is going well (although this also can be productive). This practice, in turn, requires a culture of safety and trust where individuals can discuss their own areas of growth and feel supported in taking risks by trying new practices.
- Data discussion protocols and meeting agendas (Little & Curry, 2009; Takahashi & Norman, 2025). The data conversation can be structured and designed to enable participation of all voices, to stay grounded in data before introducing hypotheses, and to lead toward collectively shared learning connected to next steps. A protocol used by a skilled facilitator can deepen the learning that can be had from practical measures.
- Visual representations of data. Most people tend to process and understand information better when it is presented visually, so it is critical to provide clear and organized representations of data designed to inform actionable discussion around identified research questions. Data visualizations are most effective when they reflect data accurately, use clear labels and uncluttered design, and take advantage of the graph type best suited to the data and research questions (Evergreen, 2017). Walston and Conley (2022) recommended organizing data by the questions they are intended to address. For example, if one wants to know how often a teacher implemented a particular change idea (e.g., conducting a number talk, orchestrating a whole-group summary discussion), one might look at data from a teacher's self-report log, a student survey, an observer's checklist, or some combination of all three.

OPPORTUNITIES FOR THE FIELD

This project illustrated the relevance and importance of practical measures to the work of instruction leaders and educators working in continuous improvement networks, yet there is much more to learn about practical measurement in this context and, more broadly, in the field of math education. The following subsections outline three concrete opportunities that would extend what is known about

practical measurement and potentially provide greater support to the work of instructional leaders.

Employ Practical Measures in Common Professional Learning Structures

Most districts and schools are not part of large, externally funded improvement networks, yet districts and schools have resources for professional learning and, at a high level, work toward continuous improvement. Instructional leaders might integrate practical measures in regular professional learning community meetings or coaching. For example, if a fourth-grade math team identifies comparing fractions as an area for improvement after reviewing formative assessment data, they could spend time as a group identifying other sources of data that could be used to address the issue. Perhaps the team has not yet gotten information from students about why they struggle with this topic, and they could introduce a short survey. Perhaps the issue is related to instruction: Teachers may not be using representations or different types of representations with students, and they could collect data on how different types of instructional routines resonate with students. This type of inquiry could be applied to instructional coaching cycles and be the basis of empirical discussions about instructional improvement.

Apply Practical Measurement to Implementation of High-Quality Instructional Materials

Independent organizations, like EdReports, analyze the quality of curricular materials in terms of rigor, usability, and coherence, providing district leaders with additional data to inform their selection of instructional materials. Curricula are rated higher in quality when they include more mathematically rigorous content and integrate the mathematical practices into their instructional routines. Such materials provide opportunities for practical measures to support implementation and continuous improvement. For example, a program emphasizing student discourse through one or more instructional routines could benefit from a practical measure that captures the quantity of discourse or students' perceptions of the routine. The math practical measures repository has measures that could be used for this purpose. Instructional leaders could apply practical measures to understand better how teachers use their planning time and improve the support they provide. These sorts of efforts could provide more nuanced information about implementation of high-quality instructional materials rather than the all-too typical "the program was too difficult for teachers to implement" storyline.

Leverage Technology to Improve the Efficiency of Instructional Improvement Work

Technology in education has continued to advance at a rapid pace, and some technological tools have the potential to make the work of instructional leaders and educators more effective. Although this message is not new, and the mixed effects of technology as part of instruction are well documented (Campuzano et al., 2009; Higgins et al., 2019), some tools are promising. For example, TeachFX could make small group or 1:1 coaching sessions more productive because of the classroom data it automatically generates. This tool is also less invasive than video observations and might expand the number of teachers willing to share artifacts from their own classrooms. Survey data collection tools have become better and easier to use over time, including the types of displays that can be automatically generated to guide the work of instructional leaders and educators. Countless other examples exist of technological tools that could support the work of instructional leaders. The question is why certain tools are helpful and how those tools can be used in authentic educational settings.

Educators and leaders must attend to ongoing tension between authentic data interpretation and organizational pressures. Because the data cannot speak for itself (Coburn & Turner, 2011), a key role of the instructional leader is to facilitate active sensemaking among educators as they evolve into concrete, authentic actions for continuous improvement. The more opportunities teachers have to practice improvement with their coaches, the stronger the uptake (Biag & Sherer, 2021). Still, maintaining an inquiry stance with educators can be challenging in the face of intense accountability pressures from school and district leaders. It is important for leaders to acknowledge this tension and find opportunities for synergy. For example, as teachers use data from practical measures to shift instruction to deepen student engagement, leaders can underscore the connection between engagement and performance on the next formative or benchmark assessment. Such advocacy allows educators to thrive from ongoing, job-embedded professional development (EA3) and identify a broader evidence base that celebrates student learning (EA7).

We urge leaders to consider these field-building opportunities and document and share lessons learned with other instructional leaders and educators. It is only through creating a larger professional community, informed by evidence from practitioners, that the field will move forward. Indeed, such work is the engine for the useful, dynamic knowledge base for teaching that Hiebert et al. (2002) envisioned 2 decades ago, and development of that engine is long overdue.

REFERENCES

Adelman, C. (1999). Answers in the tool box: *Academic intensity, attendance patterns, and bachelor's degree attainment* (ED431363). ERIC. https://files.eric.ed.gov/fulltext/ED431363.pdf

Bennett, B., & Provost, L. (2015). What's your theory? *Quality Progress,* 48(7), 36–43.

Biag, M., & Sherer, D. (2021). Getting better at getting better: Improvement dispositions in education. *Teachers College Record,* 113(4), 1–42. https://doi.org/10.1177/016146812112300402

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better.* Harvard Education Press.

Burton, M., Silver, E. A., Mills, V. L., Audrict, W., Strutchens, M. E., & Petit, M. (2018). Formative assessment and mathematics teaching: Leveraging powerful linkages in the US context. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe* (pp. 193–205). Springer.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., & Hiebert, J. (2020). Working across contexts: Scaling up or replicating with variations? *Journal for Research in Mathematics Education,* 51(3), 258–267. https://doi.org/10.5951/jresemtheduc-2020-0007

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., Hiebert, J., & Bakker, A. (2020a). Addressing the problem of always starting over: Identifying, valuing, and sharing professional knowledge for teaching. *Journal for Research in Mathematics Education,* 51(2), 130–139. https://doi.org/10.5951/jresematheduc-2020-0015

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., Hiebert, J., & Bakker, A. (2020b). Maximizing the quality of learning opportunities for every student. *Journal for Research in Mathematics Education,* 51(1), 12–25. https://doi.org/10.5951/jresematheduc.2019.0005

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018a). Building and structuring knowledge that could actually improve instructional practice. *Journal for Research In Mathematics Education,* 49(3), 238–246. https://doi.org/10.5951/jresematheduc.49.3.0238

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018b). Reconceptualizing the roles of researchers and teachers to bring research closer to teaching. *Journal for Research in Mathematics Education,* 49(5), 514–520. https://doi.org/10.5951/jresematheduc.49.5.0514

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018c). Using data to understand and improve students' learning: Empowering teachers and researchers through building and using a knowledge base. *Journal for Research in Mathematics Education,* 49(4), 362–372. https://doi.org/10.5951/jresematheduc.49.4.0362

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts—Executive summary (*NCEE 2009-4042). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/sites/default/files/migrated/nces_pubs/ncee/pubs/20094041/pdf/20094042.pdf

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal,* 97(1), 3–20. https://doi.org/10.1086/461846

Carpenter, T. P., Franke, M. L., Jacobs, V., & Fennema, E. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education,* 29(1), 3–20. https://doi.org/10.2307/749715

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective,* 9(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Evergreen, S. D. H. (2017). *Effective data visualization: The right chart for the right data.* SAGE Publications.

Fennell, S., Kobett, B. M., & Wray, J. A. (2016). *The formative 5: Everyday assessment techniques for every math classroom.* Corwin Press.

Finkelstein, N., Fong, A., Tiffany-Morales, J., Shields, P., & Huang, M. (2012). *College bound in middle school & high school? How math course sequences matter.* Center for the Future of Teaching and Learning at WestEd. https://www.wested.org/resource/college-bound-in-middle-school-and-high-school-how-math-course-sequences-matter/

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher,* 31(5), 3–15. https://doi.org/10.3102/0013189X031005003

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). National Council of Teachers of Mathematics.

Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research,* 57(2), 283–319. https://doi.org/10.1177/0735633117748416

Hoang, H., Huang, M., Sulcer, B., & Yesilyurt, S. (2017). *Carnegie Math Pathways 2015–2016 impact report: A five-year review* (ED582438). ERIC. https://files.eric.ed.gov/fulltext/ED582438.pdf

Jackson, K., Henrick, E., Cobb, P., Kochmanski, N., & Nieman, H. (2016). *Practical measures to improve the quality of small-group and whole-class discussion* [White paper]. http://core.miserver.it.umich.edu/omeka-s/s/ire/item/5088

Kufeld, M., & Tarasawa, B. (2020). *The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement.* NWEA. https://www.nwea.org/research/publication/the-covid-19-slide-what-summer-learning-loss-can-tell-us-about-the-potential-impact-of-school-closures-on-student-academic-achievement/

Langfitt, F., & Glass, I. (Hosts). (2015, July 17.) NUMMI [Audio podcast episode]. In *This American Life.* WBEZ Chicago and PRX. https://www.thisamericanlife.org/561/nummi-2015

Langley, G. L., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009) *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). Jossey-Bass.

Little, J. W., & Curry, M. W. (2009). Structuring talk about teaching and learning: The use of evidence in protocol-based conversation. In L. M. Earl & H. Timperley (Eds.), *Professional learning conversations: Challenges in using evidence for improvement* (pp. 29–42). Springer Netherlands.

Morgan, J., & Liker, J. K. (2020). *The Toyota product development system: Integrating people, process, and technology.* Productivity Press.

National Center for Education Statistics. (2023). *Mathematics performance. Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/pdf/2023/cnc_508.pdf

National Council of Supervisors of Mathematics. (2019). N*CSM essential actions: Instructional leadership in mathematics education.* https://www.mathedleadership.org/ncsm-essential-actions-series/

National Governors Association for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards.* Washington, DC: Author.

Organisation for Economic Co-operation and Development. (2018). *PISA 2018 U.S. results.* https://nces.ed.gov/surveys/pisa/pisa2018/pdf/PISA2018_compiled.pdf

Organisation for Economic Co-operation and Development. (2022). *PISA 2022 U.S. results.* https://www.oecd.org/publication/pisa-2022-results/

Provost, L. P., & Murray, S. K. (2011). *The health care data guide: Learning from data for improvement.* John Wiley & Sons.

Rother, M. (2009). *Toyota kata.* McGraw-Hill Professional Publishing.

Rothkopf, A. J. (2009). Courageous conversations: Achieving the dream and the importance of student success. *Change: The Magazine of Higher Learning,* 41(1), 24–41. https://doi.org/10.3200/CHNG.41.1.24-41

Smith, M. S., & Stein, M. K. (2011). *Five practices for orchestrating productive mathematics discussions.* National Council of Teachers of Mathematics.

Solberg, L. I., Mosser, G., & McDonald, S. (1997). The three faces of performance measurement: Improvement, accountability, and research. *The Joint Commission Journal on Quality Improvement,* 23(3), 135–147. https://doi.org/10.1016/S1070-3241(16)30305-4

Takahashi, S., Jackson, K., Norman, J., Ing, M., & Krumm, A. E. (2022). Measurement for improvement. In D. Peurach, J. Russell, L. Cohen-Vogel, & W. R. Penuel (Eds.), *The foundational handbook on improvement research in education* (pp. 423–432). Rowman & Littlefield.

Takahashi, S., & Norman, J., (2025). Practical measurement for improvement: From conceptual foundations to practical expression. In P. LeMahieu & P. Cobb (Eds.), *Measuring to improve: Practical measurement in the service of continuous improvement in education* (pp. 423–432). Harvard Education Press.

Walston, J., & Conley, M. (2022). *Practical measurement for continuous improvement in the classroom: A toolkit for educators* (ED623759). ERIC. https://files.eric.ed.gov/fulltext/ED623759.pdf

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in Grades 4 through 8. IES practice guide* (NCEE 2012-4055). What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/docs/practiceguide/mps_pg_043012.pdf

Yamada, H., Bohannon, A. X., Grunow, A., & Thorn, C. A. (2018). Assessing the effectiveness of Quantway®: A multilevel model with propensity score matching. *Community College Review,* 46(3), 257–287. https://doi.org/10.1177/0091552118771754

Yamada, H., & Bryk, A. S. (2016). Assessing the first two years' effectiveness of Statway®: A multilevel model with propensity score matching. *Community College Review,* 44(3), 179–204. https://doi.org/10.1177/0091552116643162

Yeager, D. S., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement.* Carnegie Foundation for the Advancement of Teaching. http://www.carnegiefoundation.org/wpcontent/uploads/2014/09/practical_measurement_yeager-bryk1.pdf