3

Developing and Using Practical Measures to Inform Instructional Improvement in Mathematics at Scale

Kara Jackson
Paul Cobb
Marsha Ing
June Ahn
Thomas M. Smith
Nicholas Kochmanski
Starlie Chinen
Hannah Nieman

A persistent issue in the United States concerns understanding how to improve the quality of mathematics teaching, and thus student learning, on a large scale (e.g., across classrooms, across schools), and specifically doing so in ways that redress long-standing inequities in students' opportunities to learn mathematics.[1] In this chapter, we draw upon our work on the Practical Measures, Routines, and Representations (PMRR) project

(2015–2023) to focus on the contributions that practical measures can make to instructional improvement efforts that aim to address this issue.[2] PMRR consisted of three research-practice partnerships, in each of which a team of university-based researchers worked closely with a large school district to support the improvement of middle-grade mathematics teaching and learning.

In the following pages, we first clarify a research-based vision of ambitious and equitable mathematics teaching that supports students' attainment of rigorous learning goals, and we discuss the types of support that teachers need if they are to develop instructional practices consistent with this vision. We then provide a short overview of the PMRR project before describing the process that we followed to develop a set of practical measures designed to assess key aspects of the mathematics classroom learning environment. We next discuss how our district partners have used the measures in their instructional improvement efforts, and we explain why it is essential that the measures are integrated into supports for teachers' learning, such as one-on-one coaching and collaborative meetings. Against this background, we then discuss three distinct ways in which the practical measures can contribute to ongoing instructional improvement efforts: (1) determining whether instructional changes are improvements, (2) enhancing the effectiveness of various supports for teachers' learning, and (3) enhancing the coherence of instructional improvement efforts. Finally, we will share how we have investigated the validity of these measures.

AMBITIOUS AND EQUITABLE INSTRUCTIONAL PRACTICES

Decades of research suggest the importance of students developing both procedural fluency and conceptual understanding of key mathematical ideas, and also of engaging in authentic problem solving, in which they analyze novel situations to figure out which strategies to use. In addition, research indicates the importance of students learning to communicate effectively about their mathematical reasoning, and of seeing themselves and their peers as people who "do" mathematics.[3] These rigorous goals for students' mathematics learning, represented in documents like the Common Core State Standards for Mathematics, guide most if not all the official standards in states, districts, and schools.[4]

A substantial body of research in mathematics education and the learning sciences indicates how classrooms should be organized to enable students to attain rigorous goals. We refer to the resulting vision of instruction as *ambitious and equitable*. This vision is ambitious both in the sense that it specifies what instruction *should* entail and yet is currently the exception rather than the rule in most US classrooms; and in the sense that it aims at deep, enduring, and personally meaningful understandings of mathematics. It is equitable in the sense that it deliberately aims to support a diverse range of students to substantially participate in and learn from all phases of classroom lessons.

A crucial aspect of ambitious and equitable instruction concerns the nature of the *tasks* that are used as the basis for instruction. In many US mathematics classrooms, and especially in classrooms serving students of color, students for whom English is not their first language, and/or students living in poverty, it is common for instruction to be organized around tasks of low cognitive demand that students can solve by using a procedure that the teacher has explicitly demonstrated.[5] However, if students are to develop deep and enduring understandings of central mathematical ideas, it is critical that all students have regular opportunities to solve tasks of high cognitive demand that closely approximate what it means to "do" meaningful mathematics from a disciplinary perspective. Tasks of this type require students to analyze novel problem situations to figure out what strategies to use and why.[6]

Research also provides guidance on the critical features of specific *phases of lessons* that are organized around cognitively demanding tasks. One crucial phase of a lesson is the introduction, or *launch*, of a cognitively demanding task.[7] In a high-quality launch, teachers ensure that students develop understandings of any linguistic and cultural suppositions that may be unfamiliar in a given task. Just as important, teachers also support students in understanding the key mathematical ideas required to understand what a problem is asking of them, but without suggesting specific procedures for solving it. This phase of a lesson is especially important from an equity perspective. Unless the launch is of high quality, students who are unfamiliar with any linguistic or cultural suppositions of the task, or who have difficulty picturing what is happening mathematically in the task, will likely struggle to begin working productively on it and might well become discouraged, thus limiting their learning opportunities in the remaining phases of the lesson.

Following a launch, students engage in either *individual or small group work* to solve cognitively demanding tasks. If students are working in small groups, it is important that norms and routines have been established such that they see value in sharing ideas with one another, especially tentative and exploratory thoughts, and in asking one another to explain their reasoning so that the group can consider different solution paths.[8] During this phase of lessons, it is important when students work both individually and in small groups that teachers monitor how they are attempting to solve tasks to plan for a productive whole-class discussion.[9]

In the final phase of a lesson, teachers orchestrate a *whole-class discussion*, in which they intentionally select certain students' strategies to be shared, and in what order, to advance students' understandings of the key mathematical ideas.[10] Crucially, in a whole-class discussion that advances students' learning, students do more than share their solution strategies. A key aspect of the teacher's role is to press and support students to explain not merely how they attempted to solve tasks, but also *why* they used a particular approach, thereby making their previously implicit interpretations of tasks visible. In addition, the teacher presses and supports the listening students to make sense of and assess those explanations and to make connections between different solution strategies, thereby highlighting the key mathematical ideas under consideration.[11]

SUPPORTING TEACHERS' DEVELOPMENT OF AMBITIOUS AND EQUITABLE TEACHING PRACTICES

This vision of ambitious and equitable mathematics instruction represents a significant shift from current practices in most US classrooms.[12] For example, achieving this vision requires a fundamental change in the role of the mathematics teacher from showing students how to use procedures to solve familiar types of problems to providing opportunities for students to derive or even invent procedures that they understand conceptually to solve novel tasks. It also requires teachers to shift from acting as the sole arbiter of what's correct and incorrect, to supporting students in evaluating options and deciding what makes sense based on mathematical arguments.

A substantial body of research in professional development and teacher education indicates that making these kinds of changes in the *how* of teaching involves significant teacher learning.[13] This research also makes it clear that most teachers will require two types of support if they are to develop ambitious and equitable instructional practices. One type of support concerns instructional materials that aim at central mathematical ideas and that target both conceptual understanding and procedural fluency. A substantial body of evidence indicates the importance of including cognitively demanding tasks and of organizing these tasks into instructional sequences that are coherent, in that what students learn in one phase of a sequence provides a basis for their learning in subsequent phases.[14]

A second type of support concerns sustained opportunities to work closely with accomplished colleagues to develop their pedagogical content knowledge and instructional practices.[15] *Pedagogical content knowledge* refers to the content knowledge that is specific to the work of teaching mathematics.[16] For example, teachers need to understand not only the meaning of fractions themselves, but also how students tend to make sense of fractions and how to represent key ideas of fractions to support students' development of a deep and enduring understanding of them. Teachers need intentional support to develop this kind of specialized knowledge and to investigate and try new instructional practices, including planning for and orchestrating whole-class discussions that advance students' understandings of key mathematical ideas. Depending on resources in a given school or district, supports for teacher learning might come in the form of one-on-one coaching, teacher collaborative meetings, a series of professional development sessions, or some combination of these. It is important that the types of support are aligned and constitute a coherent system for teacher learning, such that, for example, what teachers are working on with a coach is consistent with and builds on what they are working on in grade-level teacher collaborative meetings.[17]

Over the last fifteen years, we have engaged in long-term partnerships with several large, urban districts to support their efforts to improve the quality of mathematics teaching and learning districtwide.[18] A primary goal of this work has been to support the districts in establishing and improving a coherent system of high-quality supports for teachers' learning.[19] It is in the context of these partnerships that we have come to see that practical measures can make essential contributions to such instructional improvement efforts. To be clear, we do not see practical measures as a stand-alone silver bullet. However, as we will describe in this chapter, we have found in our partnership work that practical measures can play an important role in enhancing supports for teachers' learning.

OVERVIEW OF THE PMRR PROJECT

The PMRR project has two overriding goals. The first is to develop a system of practical measures that are useful in creating and improving supports for teachers' learning. To date, we have developed a set of *classroom measures* that assess students' perceptions of key aspects of the classroom learning environment that prior research findings indicate make a difference for students' learning. They include measures of the launch of cognitively demanding tasks, small-group work, and whole-class discussions. In addition, we have developed a rubric designed to assess the level of cognitive demand of the instructional tasks that teachers use as the basis for their instruction. Alongside these measures, we developed *routines* for administering them and for analyzing the resulting data, and *representations* that facilitate educators' analysis and productive use of the resulting data.

The second goal of the project is to investigate the contributions that the practical measures (and their associated routines and representations) make to our partner districts' instructional improvement efforts. Our partner districts have used the classroom practical measures in initiatives that support the ongoing improvement of one-on-one mathematics coaching, teacher collaborative meetings, and the writing of curriculum guides. Our decision to investigate the use of the classroom measures in three districts was intentional, as our goal was to develop measures that will be useful across schools and districts that are aiming to improve the quality of mathematics teaching and learning.

DEVELOPING PRACTICAL MEASURES OF STUDENTS' PERCEPTIONS OF THE CLASSROOM LEARNING ENVIRONMENT

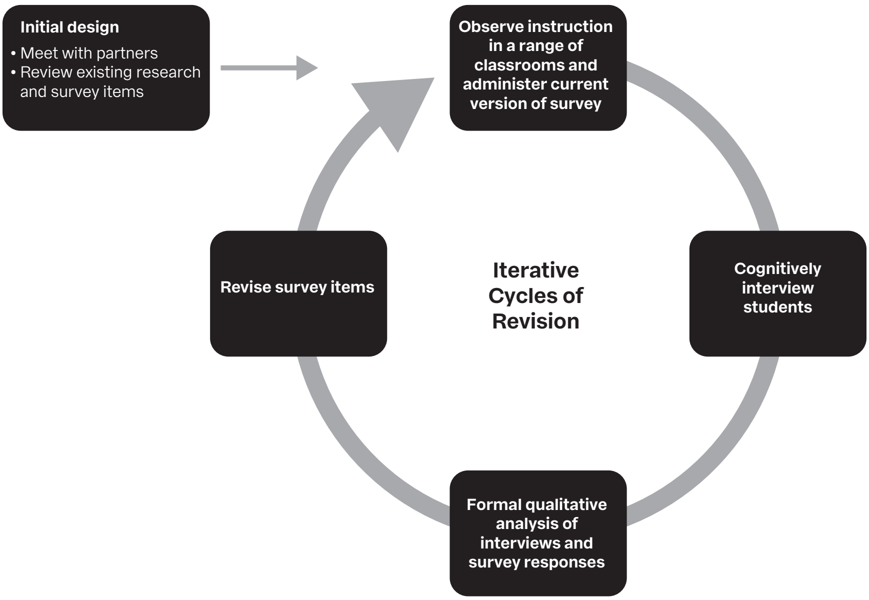
In this section, we describe the process that we undertook to develop practical measures of students' perceptions of key aspects of the classroom learning environment. Often, there is an assumption that because practical measures are, by intention and design, easy to administer and the resulting data are relatively easy to analyze, they should also be easy to develop. This has not been our experience. Generating conventional measures of key aspects of a classroom learning environment is already a difficult task, and the requirement that it should be relatively easy for practitioners to collect and analyze data while minimizing disruptions to their work are additional design conditions that only increase the challenge.

As we describe here, the process of developing the practical measures of students' perceptions of the classroom learning environment—measures of launches, small-group work, and whole-class discussions—was quite protracted. In an effort to ensure that the administration of the measures would be minimally burdensome to users, each of the three measures takes the form of short student surveys that take one to three minutes to complete.

Initial Design of a Set of Items

Figure 3.1 illustrates the process that we followed to develop measures relevant to the launch, small-group work, and whole-class discussion. For each of these measures, the first step was to review the existing research literature to identify key aspects of ambitious mathematics instruction that were the focus of each of the three surveys developed. We then developed

Figure 3.1 Development process for the practical measures of students' perceptions of the classroom learning environment



an initial set of student survey items to assess each of these aspects. Importantly, district mathematics specialists and coaches (and teachers, when possible) were involved in the development process, including in the design of an initial set of items. Their insights into what would be both feasible and compelling to a range of users were invaluable.

Iterative Cycles of Design, Analysis, and Revision

Once an initial set of items had been developed, the second step in developing each of the three student surveys was to ensure that the survey items assessed what they were designed to assess. In each revision cycle, a team of five or six researchers and district colleagues from one of the partnerships tested the current version of a survey in at least four classrooms, which varied in terms of the quality of instruction. The team members, all of whom had expertise in mathematics teaching and supporting teachers' learning, observed a mathematics lesson in each classroom and assessed the quality of instruction in terms of student learning opportunities specific to each survey item. At the end of each lesson, one of the team members then administered the survey to the students.

The team then compared students' responses, aggregated for each classroom, with their own assessment of the quality of instruction in each classroom to determine whether the survey was sensitive to those critical attributes and therefore discriminated appropriately between the classrooms in terms of key aspects of instruction that prior research has linked to student learning. In addition, immediately after observing a lesson, each team member conducted an audio-recorded cognitive interview with a student in which they asked the student to explain their response choices and probed the student's interpretations of the items.[20] The researchers then analyzed the cognitive interviews to identify mismatches between the team's assessment of the quality of instruction and students' interview responses. Finally, the researchers proposed revisions to the survey based on these two analyses that included eliminating, adding, and modifying items. We provide an illustration of the revision of one of the items in the following section. On average, we found that five rapid testing and revision cycles were required to develop each of the three surveys, which focus on different phases of ambitious lessons.

Each of the surveys was developed in English. When possible, we intentionally included students for whom English was not their first language in the trials to guide revisions to the items. Once the surveys were finalized, we had them translated into twelve languages, which reflects the diversity of languages spoken by students in our partner districts.

The Resulting Sample Measure: Whole-Class Discussion Survey

Given space limitations, we focus on one of the final set of measures: the whole-class discussion survey. Our review of the relevant research literature resulted in the identification of five aspects of whole-class discussions that affect students' learning. These aspects and the corresponding survey items are shown in table 3.1.

One aspect concerns the cognitive demand of the task as *implemented*. As mathematics education researchers Peg Smith and Mary Kay Stein observed, whole-class discussions that support students' attainment of rigorous learning goals "are unlikely to occur if the task on which students are working requires limited thinking and reasoning."[21] The research base also indicates that even when teachers use cognitively demanding tasks as written in curriculum materials (e.g., tasks that can be solved in several ways), it is very common for teachers to lower the demand of the tasks implemented by suggesting a procedure that students should use to solve the tasks.[22] Our iterative trials resulted in two survey items that provided consistent information about the cognitive demand of the task as it was implemented in the classroom and experienced by students: "What did you need to do in order to be successful in your math class today?" and "Was there only one right way to solve the problem(s) today?" (See table 3.1 for the response options for these and other items.)

The second aspect of productive whole-class discussions that we identified concerns what students are held accountable for. The intent of mathematics discussions is often to evaluate whether students' answers are correct rather than to bring significant mathematical issues to the fore by capitalizing on the range of ways in which students have attempted to solve tasks.[23] Focusing on answers alone does not help students make sense of or deepen their understanding of central mathematical ideas, and thus does not advance their thinking. Further, an exclusive focus on the correctness of answers does not enable the teacher to learn about students' current mathematical reasoning. We found that asking students, "What was the purpose of today's whole-class discussion?" was especially generative in determining whether students were primarily held accountable in discussions for making sense of mathematical ideas or producing correct answers.

In productive discussions, teachers attempt to achieve their mathematical agendas by building on students' current thinking and reasoning.[24] The third and fourth aspects concern the extent to which discussions focus on students' ideas and students have opportunities to listen to, reason about, and make sense of other students' ideas. We found that a simple item that asked, "Who talked the most in today's discussion?" with response options "Students" or "The teacher" provided an admittedly crude, but also illuminating assessment of the extent to which students

Table 3.1 Overview of the whole-class discussion measure

Aspects of the classroom learning environment that impact students' learning opportunities in whole-class discussions	**Items (Students select the one best response to each question.)**
Cognitive demand of the task as implemented	• What did you need to do in order to be successful in your math class today? ☐ *Solve problems using the steps the teacher showed me* ☐ *Listen to and make sense of other students' reasoning* • Was there only one right way to solve the problem(s) today? ☐ *Yes* ☐ *No*
Student accountability during discussion *Note: When analyzing students' responses, we collapse the first, second, and fifth options as "producing correct answers" and the third and fourth options as "sensemaking."*	• What was the purpose of today's whole-class discussion? ☐ *Share how we solved problems using the steps our teacher showed us* ☐ *Learn the way the teacher showed us to solve the problem* ☐ *Learn different ways that work to solve a problem from other students* ☐ *Share a mathematical idea we came up with on our own* ☐ *Check to see if our answers are correct*
Extent to which discussions focus on students' ideas	• Who talked the most in today's whole-class discussion? ☐ *Students* ☐ *The teacher*
Opportunities to make sense of others' ideas	• Did listening to other students in today's whole-class discussion help make your thinking better? ☐ *Yes* ☐ *No* • Did you have trouble understanding other students' thinking in today's whole-class discussion? ☐ *Yes* ☐ *No*
Extent to which students want to share their ideas and feel their ideas are valued	• Were you comfortable sharing your thinking in today's whole-class discussion? ☐ *Yes* ☐ *No* • Would it have been OK to share thinking you were unsure about in class today? ☐ *Yes* ☐ *No* • Did you feel like your teacher really thought about your mathematical ideas in class today? ☐ *Yes* ☐ *No* ☐ *I did not share today* • Did you feel like other students really thought about your mathematical ideas in class today? ☐ *Yes* ☐ *No* ☐ *I did not share today*

shared their thinking in discussions. However, while students sharing their thinking is a necessary aspect of mathematically productive discussions, research indicates that sharing, in and of itself, does not help students to deepen their mathematical understanding.[25] It is also critical that teachers press and support students to explain and justify their reasoning in ways that other students can understand.[26] Two survey items proved useful for assessing the extent to which listening students were supported to make sense of the ideas being shared: "Did listening to other students in today's whole-class discussion help make your thinking better?" and "Did you have trouble understanding other students' thinking in today's whole-class discussion?"

Of course, the extent to which students share their thinking with one another depends on the classroom norms that have been established, as well as whether students feel that their ideas are valued.[27] This fifth aspect of productive whole-class discussions is especially important from an equity perspective. It is well documented that teachers often view students of color, students for whom English is not their first language, and/or students living in poverty as being less capable of participating in rigorous mathematical activity, and these students, in turn, perceive that their ideas are not valued in classroom discussions.[28] We have found that four survey items surface students' perspectives on this aspect of the classroom culture. One item assesses students' comfort in sharing: "Were you comfortable sharing your thinking in today's whole-class discussion?" A second item assesses whether students view the classroom as a setting in which they can take intellectual risks: "Would it have been OK to share thinking you were unsure about in class today?" Two items assess whether students perceive that the teacher and other students valued their mathematical ideas: "Did you feel like your teacher really thought about your mathematical ideas in class today?" and "Did you feel like other students really thought about your mathematical ideas in class today?"

To illustrate the iterative nature of the development process described here and shown in figure 3.1, we briefly describe the history of this last item. In its first iteration, students were asked, "Was your thinking *respected* by other students in class today?" However, in cognitive interviews, four of seven students indicated that they interpreted *respected* as "listened to," suggesting that although the question indicated to these students that listening students were attentive, it did not indicate to them that the listening students *valued* their thinking. In the second iteration, we asked more directly, "Did other students *value* your thinking in class today?" However, six of the fourteen students interviewed indicated they would have checked "No" as the response option if other students had *disagreed* with their ideas, even if their ideas were seriously considered. This was problematic because of the important role that disagreements play in productive mathematical discussions. In the third iteration, we trialed "Did anyone ignore your mathematical ideas in class today?," but six of twelve students reported in cognitive interviews that *ignore* connoted disagreement or personal conflict. In the fourth iteration, we trialed the current item, "Did you feel like other students really thought about your mathematical ideas in class today?" All twelve students whom we cognitively interviewed indicated that *really thought about* communicated valuing and said that the modifier *really* was important in this regard. We then tested this version of the item in another set of contrasting classrooms and found that it continued to communicate as intended and was sensitive to differences in whether students appeared to value each other's ideas, as indicated by the ways that students responded to one another's ideas. Across all the trials of this item, our team conducted cognitive interviews with forty-five students from nine classrooms in three districts.

Each of the three practical measures of students' perceptions of key aspects of the classroom learning environment is intended to assess student learning opportunities in a specific lesson, and therefore predict what students will likely learn in that lesson. They are intentionally proximal, in that they focus on a current lesson, and thus specific instances of practice. For example, students' responses to the whole-class discussion measure predict whether their participation in a specific whole-class discussion supports them in deepening and/or elaborating their understanding of key mathematical ideas. The resulting data provide feedback on how effectively the teacher orchestrated the whole-class discussion in this lesson. For example, imagine that several students indicated that they had trouble understanding other students' thinking in the whole-class discussion. This information can then prompt the teacher to consider why students might have had trouble, and what they could have done differently to help more students to make sense of others' thinking. In the context of instructional improvement work, it is a considerable strength of practical measures that the resulting data are directly actionable.

In addition to the three student surveys that assess the launch, small-group work, and whole-class discussions, we developed a rubric to assess the cognitive demand of the instructional tasks as written, which teachers can use in planning an upcoming lesson. It was adapted from an existing research measure to make it easier for teachers and professional learning leaders to understand and use.[29] The rubric supports users in distinguishing between three types of tasks that differ in terms of what students are asked to do: "using procedures," where students have to use a procedure to solve a familiar type of task; "making sense of procedures," where students have to use a procedure to solve a task and either demonstrate why or figure out why the procedure works; and "problem solving," where students have to analyze a novel task in order to figure out a feasible solution strategy.

INVESTIGATING THE USE OF PRACTICAL MEASURES OF STUDENTS' PERCEPTIONS OF THE CLASSROOM LEARNING ENVIRONMENT

As we have illustrated, the three practical measures of students' perceptions of key aspects of the classroom learning environment target fundamental changes in how most teachers of mathematics interact with students on a daily basis. These types of changes require teachers to develop new forms of practice—ones that often challenge deeply held assumptions about what it means to teach, learn, and understand mathematics. For example, learning to facilitate a genuine discussion in response to students' current reasoning entails a profound shift in how most teachers envision and enact their role in the classroom. Research indicates that developing fundamentally new forms of practice, like facilitating authentic discussions, requires sustained professional learning opportunities in which teachers both investigate their current practices and try new forms of practice with someone more accomplished in the intended form of practice (e.g., a coach).[30] In contrast, research indicates that it is reasonable to expect that teachers can make minor adjustments to their practices after relatively limited professional development. As an example, teachers can learn to use a new pacing guide effectively from one or two stand-alone professional development sessions.[31] The key issue is the depth of the change in practice being called for, and thus the nature of the learning involved for teachers to develop the intended form of practice.

Given that the measures that we have developed target ambitious forms of practice that entail significant changes in most teachers' current practices, we contend that it is essential to integrate the practical measures into supports for teachers' learning. As an example, large-scale studies of mathematics teaching indicate that, in most mathematics classrooms, the typical purpose of whole-class discussions is to evaluate whether students' answers are correct.[32] Orchestrating whole-class discussions in which teachers elicit and build on students' reasoning, while also supporting other students to make sense of their peers' explanations, therefore represents a substantial change in practice for most teachers. Using data from practical measures productively involves interpreting students' survey responses and then deciding on specific changes to practices that might improve students' learning opportunities. Students' responses to the *whole-class discussion* measure may support teachers to rethink the purpose of whole-class discussions. However, even if teachers identify the importance of supporting students to make sense of their peers' solution strategies as an instructional improvement goal, the literature on teacher learning indicates that it is unreasonable to expect that most teachers will be able to determine how to change their practices to attain this goal on their own.[33] It is for these reasons that we contend that administering the practical measures and analyzing students' responses should be integrated into existing supports for teachers' learning, such as one-on-one coaching and teacher collaborative meetings.

In what follows, we describe how two of our partner districts have used the classroom measures and illustrate two distinct ways in which the measures can contribute to instructional improvement efforts when they are integrated into supports for teacher's learning: (1) determining whether an instructional change is an improvement; and (2) enhancing the effectiveness of the supports for teachers' learning. In addition, we discuss a third possible contribution of the measures—namely, that the use of the measures can enhance the coherence of instructional improvement efforts at scale. The two districts in which we ground the discussion, District E and District I, both aimed to support teachers' development of ambitious and equitable instructional practices. However, there were differences in the districts' theories of improvement and in the available resources and structures for supporting teachers' learning. In District E, the whole-class discussion measure was integrated into one-on-one coaching cycles, whereas District I used the measure in a Curriculum Guide Writing initiative to improve the quality of the written mathematics curriculum. Both cases illustrate how the measure can be used to determine whether an instructional change is an improvement. In addition, the District E case illustrates how the integration of the measure into a key support for teachers' learning (i.e., coaching) can enhance the effectiveness of the support.

Determining Whether an Instructional Change Is an Improvement

As Takahashi and Norman describe in chapter 2 of this book, practical measures enable users to get timely feedback regarding whether a change that they deliberately made to their practice is an improvement. There are two ways in which we have found that our practical measures can support practitioners in determining whether an instructional change is, indeed, an improvement: whether a change *in teaching* results in improvements in students' learning opportunities, and whether a change in a *support for teachers' learning* results in improvements in instruction.

Case 1: Using Practical Measure Data to Determine If an Instructional Change Is an Improvement

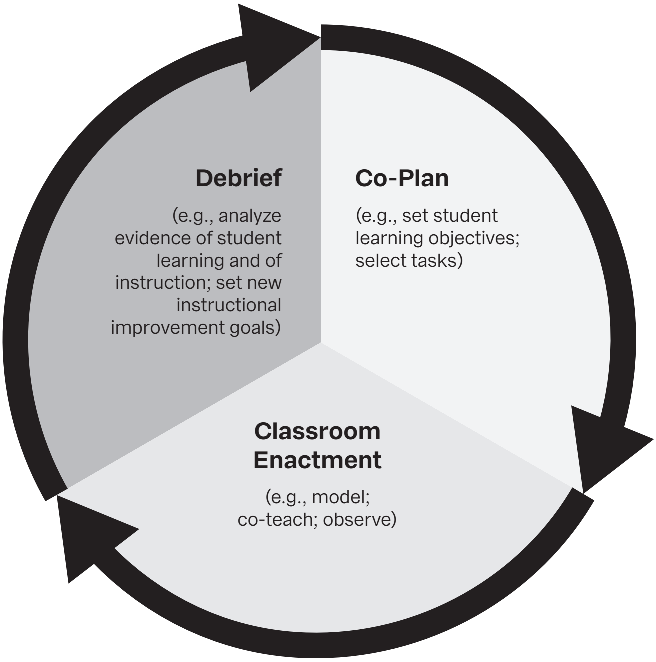
The overall goal of the partnership with District E was to establish a cadre of accomplished mathematics coaches who could work with teachers one on one to support their development of ambitious and equitable instructional practices. To this end, the researchers and district mathematics specialists collaborated to design and lead a sequence of eight professional development sessions for each of two years for fifteen middle-grade mathematics coaches, each of whom served one school. The goal of the professional development was to support the coaches in enacting coaching cycles effectively, where each cycle consists of the three phases shown in figure 3.2:

- A *coplanning phase*, in which the teacher and coach coplan an upcoming lesson, with particular attention to the aspect of instruction that the teacher is attempting to improve.
- A *lesson enactment phase*, in which the coach either models particular aspects of high-quality instruction while the teacher observes, the teacher and coach coteach the lesson, or the teacher teaches the lesson while the coach observes.
- A *debrief phase*, in which the teacher and coach analyze the impact of instruction on students' learning during the lesson.

Readers familiar with improvement science as it has been applied to education will note the strong parallel between the three phases of a coaching cycle and the four phases of a Plan, Do, Study, Act (PDSA) cycle, with the concluding Study and Act phases of the cycle being collapsed into the final debrief phase of a coaching cycle.

The integration of the practical measures into a coaching cycle involves a coach-teacher pair first identifying the student survey that is aligned with the aspect of instruction that the teacher is attempting to improve (e.g., they select the whole-class discussion survey if the teacher was attempting to improve aspects of classroom discussions). They then administer that survey at the end of the lesson enactment phase and analyze the students' responses during the debrief phase along with other

Figure 3.2 One-on-one coaching cycle



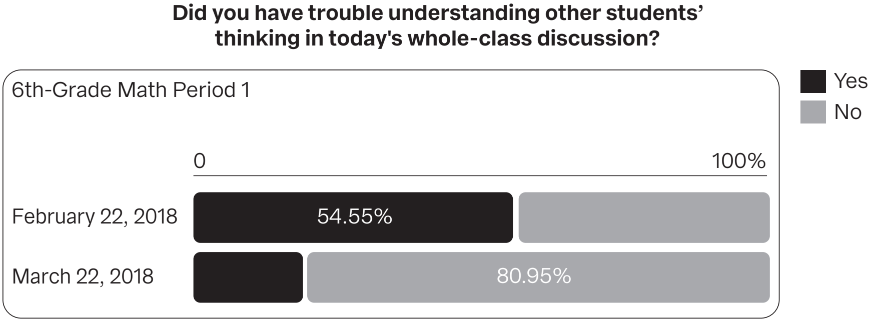
sources of evidence collected to document students' learning, including student work. Using the practical measures data in this way enables them to determine whether a change that the teacher has made to instruction has improved students' learning opportunities. If a coach and teacher judge an instructional change to be an improvement, they might draw on their analysis of the lesson to set a new instructional improvement goal that would then orient their work during the next coaching cycle. However, if the instructional change did not result in the intended improvement in students' learning opportunities, the teacher and coach might propose an alternative instructional change that could have the desired impact.

As an illustration, one coach-teacher pair had administered the whole-class discussion survey in the prior coaching cycle and found it problematic that more than half of the students had responded "Yes" to an item that asked, "Did you have trouble understanding other students' thinking in today's whole-class discussion?" (see figure 3.3; data from February 22, 2018). Crucially, the coach and teacher attempted to understand why the students responded in this way, with the coach observing that the teacher had frequently rephrased students' explanations during the discussion. This led the teacher to suggest that if she encouraged the listening students to rephrase other students' explanations, this might help them understand those explanations:

Coach: I notice when a student will share, you would rephrase what they were sharing. I wonder if—
Teacher: —having another student rephrase?
Coach: Exactly.

In subsequent mathematics lessons, the teacher worked to support her students in restating others' explanations in their own words rather than restating them herself. When the coach and teacher administered the whole-class discussion survey one month later in the next coaching cycle, they found that less than a quarter of the students indicated that they had trouble understanding other explanations (see figure 3.3; data from March 22, 2018). They interpreted these data as indicating that the instructional change that the teacher had made was indeed an improvement, with her commenting, "That's great. Put that on my résumé."

Figure 3.3 Students' responses to whole-class discussion survey item, at two time points



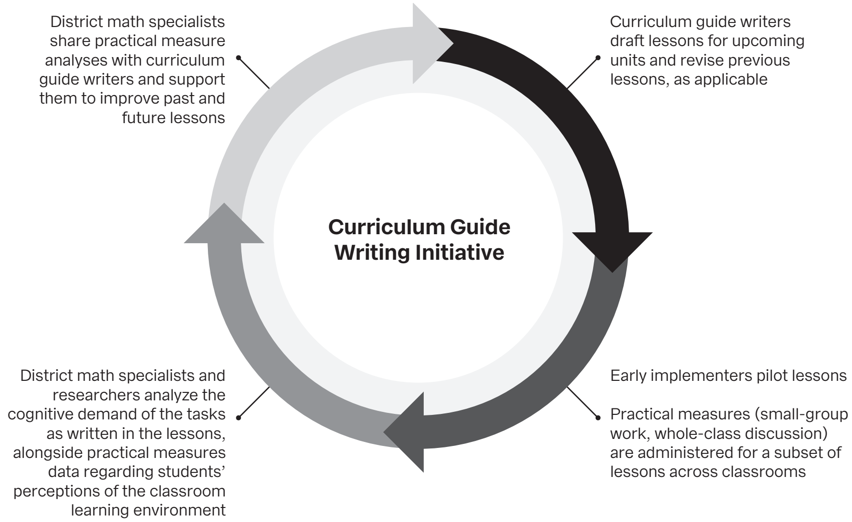
This illustration is representative in that the coaches and their partner teachers increasingly used the practical measures to determine whether a change in teaching was an improvement as the professional development progressed.[34] Integrating the practical measures into coaching cycles enabled them to gauge whether they were making progress in improving students' learning opportunities, and to adjust their work together accordingly. Furthermore, supporting the development of a cohort of coaches who used the classroom measures in this manner as they each worked one on one with several teachers would contribute directly to improvements in the quality of instruction and students' learning at the district level.

Case 2: Using Practical Measure Data to Determine If a Change in a Support for Teachers' Learning Is an Improvement

We now turn to a case of using these practical measures to assess whether a change in a support for teachers' learning results in improvement in instruction. To illustrate this case, we draw on how District I mathematics specialists integrated the classroom practical measures in a Curriculum Guide Writing Initiative.[35] Prior to this initiative, middle school mathematics teachers in District I were expected to use a freely available online curriculum, but they were given little guidance regarding which aspects of the lessons to prioritize and how. District mathematics specialists were concerned that most teachers were selecting tasks of low cognitive demand from the online curriculum, thereby limiting students' opportunities to develop conceptual understanding and procedural fluency. District mathematics specialists also expressed equity concerns; without district guidance, it was likely that the quality of instruction varied widely across classrooms, and very likely in ways that furthered inequities in students' learning opportunities.

In response, the district undertook a Curriculum Guide Writing Initiative, which is summarized in figure 3.4. District mathematics specialists recruited accomplished middle-school mathematics teachers to serve as curriculum guide writers. Writers worked with the district mathematics specialists for a day every four to six weeks to draft new units and to revise previously developed units. Another group of teachers was recruited to

Figure 3.4 Curriculum Guide Writing Initiative

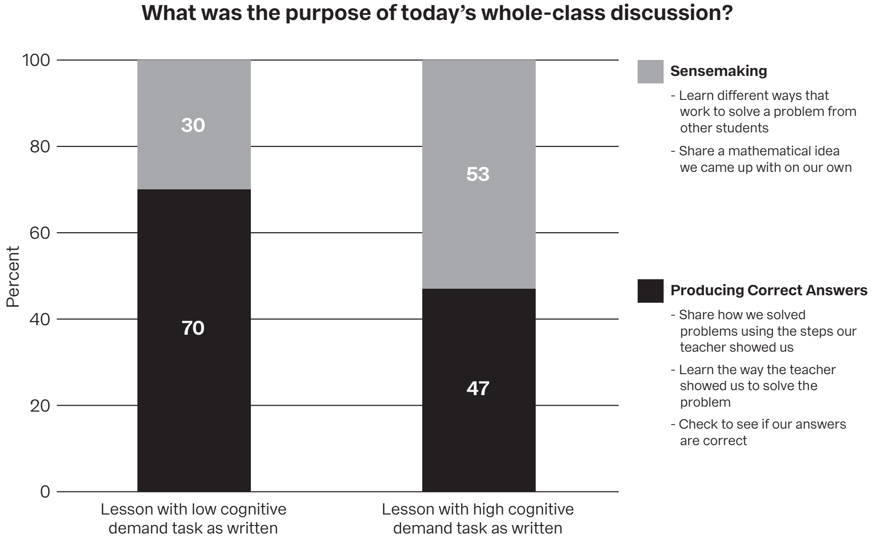


be "early implementers." They were expected to pilot the newly drafted units and to provide feedback to the writers about the materials. In addition, early implementers administered either the small-group work or the whole-class discussion practical measures for a subset of lessons within an instructional unit.[36]

District mathematics specialists and researchers then met to review an assessment of the cognitive demand of the tasks as written in each of the lessons, based on the rubric described earlier, and analyze students' survey responses, aggregated across classrooms for specific lessons in an instructional unit, to gain insight into both the cognitive demand of the tasks as implemented and into students' learning opportunities in small-group work and whole-class discussions in those lessons. In addition, they investigated the relationship between the cognitive demand of a task as written and students' learning opportunities in the lesson as implemented, by looking at the assessment of the cognitive demand of a task as written alongside survey data for specific lessons.

To illustrate how practical measure data was used in this process, figure 3.5 shows students' responses to one whole-class discussion survey

Figure 3.5 Sample practical measure data shared with district mathematics specialists and curriculum guide writers

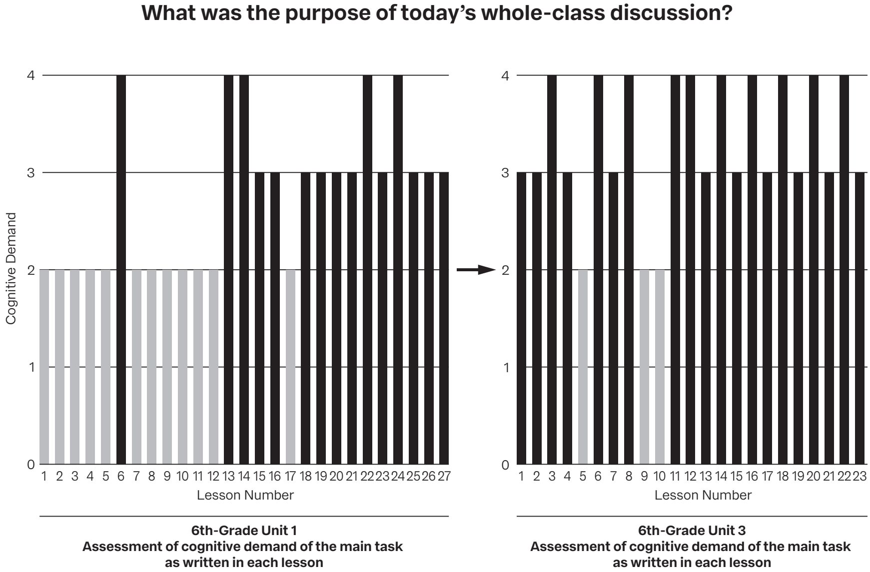


item, "What was the purpose of today's whole-class discussion?" for two different lessons within a unit, aggregated across three seventh-grade classrooms. The bar chart on the left corresponds to a seventh-grade lesson in which the task as written in the curriculum writing guide was judged to be of low cognitive demand, whereas the bar chart on the right corresponds to a seventh-grade lesson in which the task as written was judged to be of high cognitive demand. In the lesson organized around a task of low cognitive demand as written, less than one third of the sixty-two seventh graders' responses suggest that the purpose of the whole-class discussion was to engage in sensemaking. However, in the lesson organized around a task of high cognitive demand as written, more than half of the sixty seventh graders responded that the purpose of the whole-class discussion was to engage in sensemaking. Analyses like this illustrated that the cognitive demand of the instructional tasks as written influenced students' learning opportunities in the whole-class discussion phase of a given lesson. Of course, there were variations in individual teachers' implementation of tasks of high cognitive demand, and thus in students' survey responses at the classroom level. Overall, however, the pattern illustrated in figure 3.5 was evident across several lessons and grade levels.

Informed by the results of analyses such as that shown in figure 3.5, the district mathematics specialists met with the writers and used findings like these to support the writers in understanding how students' learning opportunities in whole-class discussions were influenced by the cognitive demand of tasks as written. The district mathematics specialists pressed and supported the writers to increase the cognitive demand of the tasks as written in the lessons. In addition, on the basis of student survey results indicating that there were some variations in how teachers implemented tasks of high cognitive demand, the district mathematics specialists also pressed the writers to provide more explicit guidance for teachers with regard to planning for small-group work and whole-class discussions (e.g., identification of key ideas to focus on in discussions). The writers then drafted new lessons and revised the previous lessons. In total, the district mathematics specialists, curriculum guide writers, and researchers engaged in the cycle (shown in figure 3.4) four times.

A subsequent analysis of the cognitive demand of the tasks as written in lessons indicated that the cognitive demand of the written lessons improved in all three grade levels during the initiative. For example, figure 3.6 shows the findings of an analysis of the cognitive demand of the tasks as written in sixth-grade lessons and units. The *y*-axis of each of the bar charts indicates the level of cognitive demand of the main task as written in a lesson. A score of 1 or 2 indicates that the cognitive demand of the main task as written was low, and a score of 3 or 4 indicates that the cognitive demand of the main task as written was high. As figure 3.6 shows, the cognitive demand of the tasks as written in the first half of Sixth-Grade Unit 1 (lessons 1–12) was generally low. The curriculum guide writers received feedback after they had written the first half of Unit 1 for each grade level. With support and pressure from the district mathematics specialists, the writers increased the cognitive demand of the tasks as written in later lessons in this unit (lessons 13–27) and in

Figure 3.6 Improvement in cognitive demand of the main tasks of lessons as written, by unit



subsequent units, as shown by the cognitive demand of the tasks as written in Unit 3.

More generally, this curriculum guide writing example serves to highlight how these classroom practical measures can enable practitioners to assess whether a change in a support for teachers' learning leads to improvements in instruction. In addition, it illustrates how educators, like district mathematics specialists, who often do not work directly with teachers in their classrooms, may use the practical measure data to gain insight into instructional practice across classrooms, grade levels, and schools, in relation to a district-level professional development initiative.

Enhancing the Effectiveness of Support for Teachers' Learning

The primary rationale for developing practical measures proposed in the literature is that they can provide practitioners with a means of determining whether a change is an improvement, and whether a change effort is producing the intended improvements across classrooms or schools.[37] As we have illustrated, the practical measures served this purpose effectively when they were integrated into a coaching initiative and into a curriculum writing initiative. However, we also found that coaches' use of the practical measures can further enhance their effectiveness in supporting teachers' learning by enabling them to facilitate debrief conversations that are more productive than would have been the case otherwise.

As background, debrief conversations, in which a coach and teacher analyze the enactment of a lesson that they planned together, typically focus almost exclusively on the teacher's instruction. However, the findings of two recent studies indicate that in productive debriefs, the coach and teacher first analyze students' reasoning to agree on whether they attained the learning goals for the lesson, and only then analyze the instruction, not as an end in itself but to understand why the students learned what they actually learned.[38] A major goal of the professional development for coaches that we codesigned with our district colleagues, therefore, was to support coaches' facilitation of debrief conversations in which they and their partner teachers connected student learning goals, students' reasoning, and instruction in this manner. A follow-up analysis indicates that the professional development was reasonably successful, as most of the coaches did begin to support the teachers with whom they worked in making these connections.[39] This analysis also revealed that once this development had occurred, the coaches' use of the practical measures could enhance their facilitation of debriefing conversations in three significant ways beyond determining whether an instructional change was an improvement.

The first of these enhancements was that the use of the data from practical measures supported coaches and teachers in resolving disagreements that arose as they analyzed the focal lesson. In the absence of these measures, they could collect students' work to document their learning during the lesson but typically have no tangible evidence of the quality of instruction. They therefore have to depend on their recollections, which often differ and are sometimes in conflict, and also on any notes that the coaches made. It was therefore noteworthy that the coaches with whom we worked spontaneously spoke of students' survey responses as providing what they characterized as a third, independent, and much-valued perspective on the lesson in addition to their own and their partner teacher's perspectives. Coaches viewed this as a significant advantage, as they no longer had to pit their recollections against those of the teacher but could instead take the students' perspective as a primary point of reference. Thus, they considered it advantageous to use the practical measures precisely because the measures enabled them to bring the students' voice to the fore.

The second enhancement of debriefing conversations was that coaches' and teachers' use of the practical measures could support them in developing more powerful explanations of how instruction influenced students' learning than otherwise would have been the case. This was primarily because the measures oriented them to focus on those aspects of the classroom learning environment that prior research indicates matter for students' learning but that might have otherwise been unnoticed by them. Thus, in the previous example, which illustrates how coaches and teachers can use the practical measures to determine whether an instructional change is an improvement, the extent to which the students understood each other's explanations might well have been invisible to the coach and teacher in question had they not administered the whole-class discussion survey. Most of the other items on the whole-class discussion survey (see table 3.1) and most of the items on the other two student surveys also focus on aspects of the classroom learning environment that are not directly observable and thus might well be invisible to teachers and to less experienced coaches unless they administered one of the surveys. The practical measures, therefore, served a signaling function for both teachers and coaches by emphasizing the importance of those aspects of the classroom learning environment on which the survey items focused. This observation suggests that the practical measures, therefore, might be an important support for coaches' as well as teachers' learning, although further research will be needed to investigate this conjecture.

The third enhancement of debriefing conversations was that the use of the measures could support coaches and their partner teachers in identifying what we term *productive instructional improvement goals*. Our criteria for productive goals are that they are a feasible next step in individual teachers' learning, given their current knowledge and practices, and attaining the goals is likely to result in immediate improvements in students' learning.[40] This third enhancement builds directly on the second because when coaches and their partner teachers analyze students' survey responses to explain their actual learning in a lesson, they identify not only instructional strengths, but also instructional weaknesses that delimit students' learning opportunities (e.g., students have trouble understanding others' explanations). In suggesting an instructional change that might address an identified weakness, coaches and teachers reframe the weakness as an improvement goal (e.g., enable students to understand others' explanations) and propose testable conjectures about how they might attain those goals (i.e., potential change ideas). Crucially, improvement goals formulated in this way are productive because attaining them is likely to result in immediate improvements in students' learning.

It is important to emphasize that practical measures are not silver bullets. The use of practical measures will not inevitably enhance coaches' effectiveness beyond enabling them to determine whether a change is an improvement. Instead, the extent to which the use of the practical measure enhances coaches' effectiveness depends on how coaches facilitate teachers' analyses of and use of practical measures data. More specifically, it is essential that coaches press their partner teachers to relate students' survey responses to their instruction.[41] For example, if a majority of the students indicate that they have trouble understanding other students' explanations, the coach might ask the teacher, "Why do you think most of the students chose that response?" In asking this question, the coach supports the teacher to consider their crucial role in shaping the classroom learning environment as the students experienced it, and thus to connect aspects of their instruction (e.g., the teacher rephrases students' explanations) to students' learning opportunities (e.g., the extent to which students understand others' explanations). It is only when coaches support teachers in making this connection that they complete the links between the teacher's instruction and students' learning, as shown in figure 3.7.

Figure 3.7 Practical measures data as a bridge between students' learning and the teacher's instruction

In doing so, they unlock the full potential of the practical measures by making the connection between the partner teacher's instruction and students' learning visible, and thus open to scrutiny.

We strongly suspect that the three ways in which the use of the practical measures can enhance coaches' effectiveness also apply to two other types of support for teachers' learning: professional development sessions and teacher collaborative meetings. As we have noted, high-quality coaching supports teachers to connect student learning goals, students' reasoning, and instruction. Discussions in teacher collaborative meetings and professional development sessions that have the potential to support substantial teacher learning also support teachers in making these connections.[42] It is therefore plausible to suggest that integrating the practical measures into these other types of support for teachers' learning might be similarly beneficial, although additional research is needed to investigate whether this actually is the case.

Enhancing the Coherence of Instructional Improvement Efforts

Thus far, we have discussed two types of contributions that practical measures of the classroom learning environment can make to instructional improvement efforts. Here, we draw on our partner districts' experiences in using the measures to suggest a third contribution. Specifically, we anticipate that the use of the measures by practitioners at different levels of the system (teachers, coaches, and district mathematics specialists) can enhance the coherence of instructional improvement efforts at scale by orienting them all to focus on improving the same key aspects of classroom learning environments, aspects that prior research findings indicate matter for students' learning. Research on instructional improvement at scale has consistently demonstrated the importance of what Fred Newmann and colleagues referred to as "instructional program coherence" in creating supportive conditions for teachers' work and advancing student learning.[43] Instructional program coherence is characterized by a set of coordinated, long-term instructional improvement strategies (e.g., curriculum materials, coaching) that span different role groups (e.g., teachers, coaches, school leaders) and that are oriented by a common vision of high-quality instruction. In schools and systems marked by strong instructional coherence, educators at various levels of the system coordinate their efforts, so, for example, what school leaders hold teachers accountable for is aligned with the instructional materials that teachers are using and with the support provided for them to improve their instruction. Our observations indicate that the use of the practical measures might enhance the coherence of improvement efforts, even though different groups of practitioners pursue improvement goals that are specific to their role within the system (teachers in supporting their students' learning, coaches in supporting teachers in a school or across several schools, and district mathematics specialists in supporting the development of capacity for instructional improvement across schools). Here, we clarify the rationale for this conjecture by drawing on the two cases introduced thus far in this chapter.

In the coaching initiative described here, the coaches integrated the use of the student surveys into the one-on-one coaching cycles that they conducted with teachers. The district mathematics specialists, who supported the coaches, might then review students' responses aggregated across classrooms at regular points in time to assess whether the professional development that they are providing to coaches is having the intended impact on classroom instruction, and also to identify additional areas of instructional weakness. Their review of the resulting data might then lead the specialists to revise their plans for upcoming professional development sessions. In this scenario, the specialists would use the same measures to assess whether changes in the professional development were improvements as the coaches and their partner teachers would use to assess whether instructional changes were improvements. The specialists' work with coaches and the coaches' work with teachers would be tightly coordinated and oriented to pursue complementary improvement goals that focus on the same key aspects of classroom learning environments that research indicates matter for students' learning.

Our second illustration focuses on the Curriculum Guide Writing Initiative described earlier. Once the guides had been developed, a next step was for district mathematics specialists to ensure that all teachers were provided with support to improve their enactment of the cognitively demanding tasks in their classrooms. These supports included regularly scheduled grade-level teacher collaborative meetings that were facilitated by coaches. The coaches might integrate the student surveys into this work, with the participating teachers administering one or more of the surveys in specific lessons that they had planned together. The coaches could then review these data to target the support that they provided in collaborative meetings regarding planning upcoming lessons. And district mathematics specialists could review students' responses aggregated by grade level for specific lessons to inform their support of the coaches' facilitation of the collaborative meetings. In addition, district mathematics specialists could analyze the resulting data to identify patterns in students' responses regarding particular lessons, with the intent of then continuing to revise the lessons in the curriculum guide.

In both these illustrations, the measures themselves (e.g., the items) and the resulting data might serve as *boundary objects* that support coherent work by practitioners at different levels of a system.[44] Although there are differences in the purposes for which users at different levels of the system would analyze students' responses, the use of the common measures can serve to coordinate and align the focus of the various types of support provided to teachers and coaches, and with respect to a common vision of instruction.

Routines of Use and Representations of Data

Before turning to considerations of validity, we highlight two important issues that need to be addressed if practical measures are to contribute to instructional improvement efforts. The first issue concerns the design and maintenance of routines for data collection, analysis, and use that can be integrated into users' current practices relatively seamlessly; and the second concerns the development of visual representations of the resulting data that enable users to address the questions they want to ask of the data.

A defining characteristic of practical measures is that they can be easily integrated into practitioners' current practices. In the case of our classroom measures, we found it productive to work closely with colleagues in each district to design and implement routines for administering the surveys and collecting the students' responses. These routines varied across districts, depending on their technological resources, with the survey being delivered electronically in some districts and through paper and pencil in others.

It is also important to design routines for analyzing the data and for sharing the findings in ways that support the ongoing improvement efforts. As we have indicated, we consider it essential that these analysis routines are integrated into support for teachers' learning.[45] For example, we have illustrated how the analysis of students' responses to the whole-class discussion survey was integrated into the debrief phase of one-on-one coaching cycles. The resulting data analysis routine involved coaches and their partner teachers first analyzing students' work to assess what the students had learned in the focal lesson, and only then analyzing students' survey responses to understand why students had learned what they had actually learned. In the final step of this routine, the results of this latter analysis informed coaches' and teachers' identification of instructional improvement goals that oriented all phases of the next coaching cycle.

In the case of the Curriculum Guide Writing Initiative, district mathematics specialists shared the findings of their analyses of practical measures data with the writers in their regular work sessions. The data analysis routine involved district mathematics specialists and curriculum guide writers (by grade-level team) analyzing both students' responses to the whole-class discussion surveys for specific lessons and the cognitive demand of the tasks as written in those lessons. Based on the findings of this analysis, the specialists pressed the curriculum guide writers to consider how they might increase the cognitive demand of tasks as written in an instructional unit and how they might insert questions in particular lessons to help teachers focus whole-class discussions on key mathematical ideas.

Both examples illustrate the importance of integrating routines for collecting and analyzing practical measure data into ongoing instructional improvement routines, rather than treating the use of practical measure as a separate activity. In fact, in both the coaching and curriculum guide writing examples, the instructional improvement efforts were designed before leaders began using the practical measures. Furthermore, in both cases, district mathematics specialists delayed the introduction of the practical measures until their use could enhance the ongoing instructional improvement efforts. For example, in the coaching initiative, the practical measures were not introduced until the fourth of eight professional development sessions, which focused on facilitating debrief conversations, the phase of the coaching cycle in which teachers and coaches analyze practical measures data. Similarly, in the Curriculum Guide Writing Initiative, district mathematics specialists did not introduce the practical measure data to the writers until after the first sets of lessons had been drafted and the writers were eager to receive feedback on the lessons. Based on our experiences, we cannot underscore strongly enough that considering when and how to introduce practical measurement in an instructional improvement initiative is a critical step in supporting productive use of the resulting data.

The design of visual representations of the resulting data merit as much care as the design of the routines of use themselves. Intentionally designed representations can enable practitioners to conduct authentic analyses in which they examine data to address questions that are consequential for their instructional improvement work without statistical sophistication being the price of entry.[46] Members of the research team with expertise in data visualization partnered with teachers, coaches, and district mathematics specialists in each of the partner districts to codesign an online platform on which the students' survey responses can be viewed and analyzed, and to develop visualizations that would help them make informed decisions that are relevant to their improvement initiatives and purposes for analyzing students' survey responses.[47]

To design productive representations, it was important to understand both the specific theory of improvement in a given effort (e.g., coaching, curriculum guide writing) and the intended users' current practices, as this enabled the research team to anticipate the types of questions that different groups of users would likely want to address by analyzing the data. For example, in the case of the coaching initiative, it was important that coaches could compare students' responses to particular survey items over time to determine whether a change in teacher's instruction was an improvement at the individual classroom level (see figure 3.3). The data representation shown in figure 3.3 reflects many design decisions jointly made between district partners and the research and design team. For instance, we decided on stacked bar chart representations rather than other options, such as a line graph that plotted the frequency for a preferred answer (in the case of figure 3.3, the preferred answer would be the proportion of students who answered "No"). This decision was informed by our codesign sessions with partners, in which they indicated that it would be problematic if the representations focused primarily on the desired student responses as this might lead to gaming those responses. Instead, our partners suggested that coaches and teachers would be more likely to discuss and attempt to explain students' responses if those responses were represented as stacked bar charts.

In the Curriculum Guide Writing Initiative, district mathematics specialists and curriculum guide writers wanted to investigate the relationship between the cognitive demand of instructional tasks as written and students' learning opportunities. They therefore needed representations of students' survey responses for specific lessons that were aggregated across classrooms by grade level, as well as representations that enabled them to compare students' responses to particular survey questions for lessons in which the cognitive demand of the tasks as written differed (see figure 3.5). These two examples illustrate how seemingly small decisions about data representations (e.g., stacked bar charts rather than a line graph) matter, and the value of soliciting input from intended users to inform the design of data visualizations that lead to productive conversations and sensemaking.[48]

CONSIDERATIONS OF VALIDITY

In the context of our work, we have conceptualized the validity of the classroom measures as comprising two aspects. The first, *validity-for-use* of a practical measure, concerns whether a practical measure actually assesses what it was designed to assess and is thus appropriate for its intended uses. For example, the validity-for-use of the classroom measures discussed here concerns whether they actually measure those key aspects of the classroom learning environment that are of interest and can be used to determine whether an instructional change is an improvement. The second, the *validity-in-use* of a practical measure, concerns whether specific *instances of using* a measure are appropriate, and thus whether practitioners, given appropriate skill and processes, can act on the resulting data with confidence. As proposed by several measurement scholars, including Edward Haertel, Michael Kane, and Pamela Moss, determining whether a particular instance of use is valid needs to take account of the purpose for which the measure was used, the context in which it was used, and the user's current knowledge, perspectives, and practices.[49]

Validity-for-Use

The *validity-for-use* of a practical measure is about whether it assesses what it claims to assess.[50] For example, the purpose of the whole-class discussion measure is to assess key aspects of the classroom learning environment that prior research links to student learning. Evidence for validity-for-use of the whole-class discussion measure includes that students' responses predict whether their participation in a specific whole-class discussion supported them in deepening their understanding of key mathematical ideas. The cycles of development and revision described previously (see figure 3.1) were essential to establishing the evidence of the validity-for-use of the whole-class discussion measure and of the other two survey measures. As we have described, we drew on existing research that links key aspects of the classroom learning environment to students' learning to determine what the measure should assess. Having identified these critical foci, it was imperative that the items communicate to students, that the intended users (e.g., teachers, coaches, district mathematics specialists) view the aspects of the classroom learning environment on which the items focus as being relevant to their instructional improvement efforts, and that students' responses to the items, aggregated at the classroom level, reflected differences in the quality of discussions, and thus students' opportunities to learn mathematics. Repeated development cycles in which we tested and revised items were therefore essential to establish the validity-for-use of the measure. We have also found it essential to continue to assess the validity-for-use of the measures in contexts beyond the initial development sites by comparing researchers' observations of classroom instruction at new sites with students' survey responses aggregated at the classroom level to check that they reflect students' learning opportunities.

Validity-in-Use

From a validity-in-use perspective, a practical measure can be used appropriately for some purposes but not others, in some contexts but not others, and by some users but not others.[51] This perspective is not unlike measurement and assessment scholars' efforts to investigate the consequences of using measures as an important aspect of validity.[52] Based on our investigations of the use of the classroom measures in our partner districts, we have identified appropriate (and inappropriate) purposes for using the measures, as well as key aspects of contexts and of users' knowledge and perspectives that influence whether they are likely to act reasonably in response to the resulting data.

Appropriate Purposes and Uses

As described by Takahashi and Norman in chapter 2 of this book, the purpose of practical measures is to support educators' engagement in *improvement*; practical measures are never intended to be used for accountability purposes. In this regard, the practical measures of the classroom learning environment are designed to be used in efforts to improve the quality of mathematics instruction, and thus students' learning. It would be completely inappropriate to use these practical measures to evaluate the quality of instruction, teachers, or schools for accountability purposes. If the measures are used erroneously for accountability purposes, teachers may well feel pressure to suggest to students that they respond to survey items in specific ways, thereby compromising the validity of the data for any purpose. In both the coaching and the curriculum guide writing initiatives, the students' survey responses were used to assess users' current practices and to set improvement goals. The explicit establishment of an improvement frame was essential and nontrivial in both cases, especially given the prevalence of using data for accountability purposes in many districts, including our partner districts. While practitioners might use practical measures for purposes beyond those originally envisioned, those who design practical measures have a responsibility to describe the intended purposes.[53]

Key Aspects of School and District Contexts

A second critical element of validity-in-use concerns aspects of school and district context that enable a measure to be used for the purposes of improvement. Failure to attend to context will likely result in practical measures being used in an ineffective or possibly even a harmful manner. For example, we recommend using our classroom measures only in schools and districts that are implementing improvement initiatives that include the provision of supports for teachers' development of ambitious and equitable instructional practices. Relevant aspects of school and district contexts therefore include that there is a shared understanding of what counts as high-quality mathematics teaching (and thus what counts as improvement in teachers' practice), and teachers are provided with ongoing support to improve the quality of their instruction. As we have indicated, while some teachers may find students' survey responses useful on their own, it is essential for most teachers that the analysis of the resulting data is integrated into ongoing coaching, teacher collaborative meetings, or other professional learning initiatives. It is also important that the school or district has the capacity to respond to practical measures data by improving the quality of supports for teachers. The classroom measures will be most helpful, therefore, when there are district- and/or school-level leaders who have expertise in supporting mathematics teachers' learning and the time to plan and facilitate professional learning. In addition, it is critical that trusting relationships are established in these professional learning settings, such that teachers feel comfortable treating the review of students' responses as an opportunity to raise genuine questions about instruction, as opposed to an evaluation of their current instructional practices. We have also found it is essential that students feel their perceptions are valued. Although they are rare, we have witnessed cases in which students did not perceive that their teachers were genuinely interested in their perspectives and therefore did not provide truthful responses on the survey.

Key Aspects of Users' Perspectives

Another element of the validity-in-use of the practical measures concerns teachers' current perspectives on their students. An investigation of teachers' interpretations of their students' survey responses found that their interpretations are strongly influenced by their perspectives on their students' current mathematical capabilities.[54] In particular, teachers who contend that some or all of their students are incapable of participating in and learning from ambitious mathematics instruction typically explain students' survey responses in terms of limitations in the students rather than limitations in their current instructional practices, and thus do not view students' responses as providing trustworthy or relevant feedback on their instruction. Unfortunately, the findings of a prior study indicate that in the absence of support, a significant majority of mathematics teachers in districts serving large numbers of students of color, students for whom English is not their first language, and/or students living in poverty will likely interpret students' practical measures responses in this way.[55]

Based on these findings, we suggest that the measures are best used by teachers who view their students as capable of engaging in the intended forms of practice, are pursuing genuine questions about teaching, and see value in learning about students' perspectives. However, coaches have reported that examining students' responses with teachers in the debrief phase of coaching cycles has provided them with insight into teachers' perspectives on their students' capabilities—perspectives that they doubt the teachers would have discussed as explicitly had they not examined the students' responses together. The coaches also reported that these insights prompted them to consider how they might support teachers to develop more productive views of their students' capabilities. Consistent with the coaches' insights, we contend that it is critical for professional development leaders to prioritize supporting teachers' development of more productive views of their students' mathematical capabilities, especially in districts serving large numbers of students from historically marginalized communities.

CONCLUSION

Supporting sustained improvements in the quality of mathematics teaching and learning at some level of scale necessarily requires the establishment of high-quality support for teachers' learning.[56] At present, educators at various levels of the system have little means of gaining insight into what is happening between teachers and students in mathematics classrooms, what is happening between teachers and facilitators in professional learning contexts, and whether changes introduced in either classrooms or professional learning contexts lead to the intended improvements. System leaders' use of practical measures can enable them to gain insights that can, in turn, inform their decisions about how to improve the support for teachers' learning, and thus teaching and students' learning.

In this chapter, we have illustrated how the classroom measures that we have developed provide actionable feedback about high-leverage aspects of mathematics instruction. We have illustrated three ways in which they can contribute to instructional improvement efforts. First, the measures enable educators to determine if an instructional change is an improvement. Second, when the measures are integrated into supports for teachers' learning (e.g., a coaching cycle), they can enhance the quality of that support. Third, we have conjectured that use of the measures by practitioners at different levels of the system who are involved in an instructional improvement effort can enhance the coherence of that improvement effort.

We have argued that these practical measures, which can be used to assess teachers' progress in developing ambitious and equitable instructional practices, should be integrated into supports for teachers' learning. However, we also see value in other types of practical measures that focus on whether an action occurred rather than on the quality of the action. For example, it could be useful for coaches and district mathematics specialists to have a practical measure of the proportion of teachers who select and use cognitively demanding tasks as a basis for their instruction, as this information would help them understand the extent to which this aspect of ambitious instruction has become routine and widespread. It could also be valuable to have practical measures that focus on time allocation, for example, the amount of time that teachers devote to the various phases of lessons (e.g., launching cognitively demanding tasks, small-group or individual work, whole-class discussion). We anticipate that, in some cases, teachers might be able to interpret and act productively on data from these types of measures without substantial professional learning support.

In this chapter, we have focused on practical measures of key aspects of the classroom learning environment. However, our long-term goal is to develop a *system* of practical measures, routines, and representations that, in addition to the classroom measures, include practical measures of *teachers' perceptions of aspects of the professional development learning environments* that prior research has linked to teachers' learning. We have developed a measure specific to teacher collaborative meetings, and are currently refining measures specific to one-on-one coaching cycles. We are hopeful that, as with the classroom measures, the professional development measures will allow users (e.g., coaches, professional development facilitators, and district mathematics specialists) to determine whether a change in their facilitation practice is an improvement, thereby contributing to the improvement of the support for teachers' learning. We also hypothesize that most coaches and facilitators will likely need support in interpreting and responding productively to the data. Further, we anticipate that the professional development measures will be especially useful to those people who are charged with supporting the learning of coaches and professional development facilitators, especially district mathematics specialists. For example, a repeated assessment of the quality of coaching cycles or of teacher collaborative meetings would allow district mathematics specialists to make informed decisions about how to target their support for coaches' and facilitators' learning, and might also enhance the support that they provide beyond determining whether a change in coaches' or facilitators' practices is an improvement.

As a final observation, we anticipate that a system of practical measures that assesses key aspects of both classroom learning environments and professional development learning environments will enable district leaders and researchers to distinguish between an inadequate theory of instructional improvement and the inadequate implementation of a theory. A robust theory of instructional improvement specifies clear goals for students' learning and a vision of high-quality instruction, and it also proposes specific forms of support for teachers' learning, together with a rationale for why it is reasonable to expect that those supports will enable teachers to improve their instruction, and thus their students' learning. On those occasions when system leaders detect a lack of improvement in students' learning, it is typically difficult for them to determine whether this disappointing outcome stems from inadequacies in their theory of improvement or from challenges in implementing the theory. However, a system of practical measures of the type that we envision would enable them to assess whether supports for teachers' learning were implemented adequately (measures of key aspects of professional development learning environments), and if they were, whether they enabled teachers to make the intended improvements in their instructional practices (measures of key aspects of classroom learning environments). As part of this analysis, systems leaders might also investigate variations in implementation across classrooms, grade levels, and schools, and might revise or elaborate their existing theory of improvement based on their findings.

We conclude this chapter by highlighting how use of the measures of the classroom learning environment can raise issues of equity that otherwise might go unnoticed. First, the three surveys assess the quality of instruction by soliciting students' perceptions of key aspects of the classroom learning environment, thereby giving students a voice in decisions about how their teachers might change their instructional practices to better support their learning. Second, all three student surveys assess aspects of instruction that are directly relevant to the equitable distribution of students' learning opportunities. For example, the decision to assess the quality of teachers' task introductions or launches was motivated by a prior investigation finding that in the majority of 1,700 video-recorded mathematics lessons, many of the students would likely not be able to begin working productively on the instructional tasks, particularly students from traditionally underserved groups.[57] Third, as we have discussed in this chapter, the use of the measures can reveal important issues regarding teachers' perspectives of their students' current capabilities, both in terms of students' responses to specific items (e.g., the item that assesses whether students perceive that the teacher valued their mathematical ideas), and in terms of how teachers engage in the analysis of the resulting data. This goes right to the heart of one of the most persistent and difficult equity issues in education today. Finally, the classroom measures enable system leaders to identify what Anthony Bryk and colleagues term "unwanted variation" in teachers' instructional practices. As Bryk and colleagues observe, reducing such variation is essential in redressing inequities in the performance of educational systems.[58]

ACKNOWLEDGMENTS

The work described and the writing of this manuscript was supported by the Spencer Foundation's Research-Practice Partnership Program and the National Science Foundation (Nos. DRL-1620851, DRL-1620863, DRL-1621238, DRL-1911492). The opinions expressed do not necessarily reflect the views of either foundation. We are grateful to our many collaborators, which include researchers, teachers, coaches, and district mathematics specialists. See a full list of our collaborators at https://www.pmr2.org/team.