Practical Measures to Improve the Quality of Small-Group and Whole-Class Discussions

Kara Jackson[1], Erin Henrick[2], Paul Cobb[2], Nicholas Kochmanski[2], and Hannah Nieman[1]
[1]University of Washington
[2]Vanderbilt University

Please cite this paper as follows:
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]. Retrieved (date) from University of Washington: http://www.education.uw.edu/pmr/files/2016/09/White-Paper.pdf

Introduction

This White Paper is intended to serve as a resource for educational researchers and practitioners interested in using practical measures and routines to support improvement efforts in schools. In particular, it describes two practical measures aimed at supporting efforts to improve the quality of small-group and whole-class discussions in middle-grades mathematics. The measures were developed in collaboration with partners in Jefferson County Public Schools (Louisville, KY) and San Francisco Unified School District (San Francisco, CA), and a team of researchers.[1]

The White Paper is meant to accompany the measures we have developed. One of our primary goals is to develop tools and routines that can inform improvement efforts in mathematics instruction. Therefore, we have made our initial measures publicly available at www.education.uw.edu/pmr. Our hope is that users will try out and adapt the measures to fit with their specific improvement initiatives. In exchange for using the measures, we ask that all users share with us adaptations, rationales for making adaptations, and information on the results of the adaptations. We intend to then make this information publicly available on the website.

The paper is organized as follows. First, we describe what we mean by the term practical measure. Second, we describe the process by which we developed the measures. Finally, we put forth our perspective regarding how the measures might be used (and not used).

What Are Practical Measures?

Our work on the development of practical measures is inspired by the Carnegie Foundation for the Advancement of Teaching, and in particular Tony Bryk and colleagues' (Bryk, Gomez, Grunow, & LeMahieu, 2015) advocacy for the use of improvement science to address persistent problems of education at some scale. Improvement science consists of a set of principles, tools, and methodologies aimed at supporting educators to use scientific inquiry to

[1] The measures described were developed in collaboration with (in alphabetical order) June Ahn, Harold Asturias, Elizabeth Hull-Barnes, Hilda Borko, Phil Daro, Suzanne Donovan, Kanna Edison, Stephanie Fields, Louis Gomez, Angela Harris, Michael Jarry-Shore, Ho Nguyen, James Ryan, Thomas Smith, Sola Takahashi, Emma Treviño, Karen Tran, Tim Truitt, and Laura Wentworth. This work has been supported by the Spencer Foundation and the National Science Foundation (The Research + Practice Collaboratory, Grant No. DRL-1238253; and Grant No.DRL-1119122). The opinions expressed do not necessarily reflect the views of either Foundation.

develop solutions to practical problems that impact their own educational settings. Such solutions can then be adapted to support improvement efforts in other contexts.

A leading principle of improvement science is that "we cannot improve at scale what we cannot measure" (Bryk et al., 2015). As such, a key tool of improvement science concerns what Carnegie has called "practical measures," or "measures for improvement" (Yeager, Bryk, Muhich, Hausman, & Morales, 2013). Practical measures are designed to provide practitioners with frequent, rapid feedback that enables them to assess and adjust their practices during the process of implementation. Practical measures therefore serve a much different purpose than other, more familiar types of measures – research measures and accountability measures. Highly refined *research measures* typically are designed to inform theory and cannot usually be used to inform districts' instructional improvement efforts, in large part because they tend to be demanding to administer, thus limiting the frequency with which they can be used (Bryk et al, 2015). *Accountability measures* are typically used to identify "exemplary or problematic individuals (e.g., districts, schools, teachers) in order to take some specific action, such as extending a reward or imposing some sanction" (Yeager et al., 2013, p. 9). They are not well suited to informing improvement because the data are often collected at the end of some cycle (e.g., after a 6-week unit of instruction, after a year's worth of instruction). In addition, such measures are usually "global measures of outcomes" that make it difficult to pinpoint where to take action (p. 9).

Characteristics of practical measures include the following (adapted from Bryk et al., 2015; Yeager et al., 2013):

- The focus of the measure is specific to an improvement goal.
- The measure uses language that is relevant and meaningful to practitioners.
- Data collection and analysis are undemanding and can be easily embedded in practitioner routines, thus making it feasible to use the measure on a monthly, weekly, or even daily basis.
- The measure is sensitive to change.
- The data produced by the measure are relevant to practitioners and have implications for action.

In addition to the characteristics listed above, we are investigating the extent to which practical measures not only provide an *assessment* of improvement with respect to a specific goal, but can also act as a *lever* for improvement. That is, we see great value in measures that have the potential to direct practitioners to important aspects of their practice that may have been previously invisible.

Focus of Practical Measures: Quality of Small-Group and Whole-Class Discussions

As noted above, practical measures are specific to the improvement goals of a particular initiative. In our case, our partner districts were aiming to improve the quality of discussion in middle-grades mathematics classrooms. They were doing so in light of the established research base that suggests that opportunities to engage in rich, conceptually-oriented discussion matters greatly for the development of students' mathematical understandings and mathematical dispositions (e.g., Franke, Kazemi, & Battey, 2007). In addition, this emphasis on the quality of student discussion is apparent in the Common Core Practice Standards in Mathematics, which emphasize the need for students to engage in disciplinary practices such as justifying solutions, constructing viable arguments, and critiquing the reasoning of others.

Given that our partner districts were emphasizing both small-group and whole-class discussions, and that there are differences in how each can advance student learning, we decided to develop two distinct practical measures.

Intended Users

Our district partners identified three role groups for whom they thought the resulting data would inform their practice – **teachers**, **instructional coaches**, and **district math leaders**. A primary intent of the measures is that teachers will find them useful to both track and inform improvements in their practice on an ongoing basis. That said, an assumption of the districts' efforts was that administering the measures would not, by themselves, lead to improvement in practice. Rather, the measures would need to be embedded in ongoing professional learning.[2]

The measures are also designed for instructional coaches and others who support teachers' learning. Our partner districts intended that quick assessments of teachers' practice would inform coaching work with individual teachers (e.g., a teacher might administer a measure as part of a coaching observation, and then reflect to set improvement goals). In addition, many of the coaches in our partner districts support groups of teachers, either through professional learning communities or pull-out professional development. Coaches suggested that the resulting data could inform what they would choose to work on in these group settings.

Math leaders in our partner districts (e.g., secondary mathematics directors) are also eager to use the resulting data to track improvements at some scale across their systems. They clarified that they do not have access to data that is specific to the quality of instruction, and that such data would help inform their instructional leadership decisions (e.g., assigning coaches, determining the focus of district-based professional development).

Development of the Measures

A key decision point in the development of practical measures concerns the form the measure will take. Some key factors to consider when deciding on the form of a practical measure are as follows. First, it is essential that the data can be collected easily and analyzed quickly, and without disrupting teachers' and others' work practices. In our case, we wanted to create measures of small-group and whole-class discussions that a teacher or coach could easily administer and analyze. Second, it is important to ensure that only limited training is needed to administer and analyze the data. A third factor is that the measure can be easily used across a system (e.g., in our case, across middle-grades classrooms). And, last but certainly not least, it is important to design a measure that the intended users see as useful and valid.

In light of these considerations, we determined in discussion with district mathematics leaders that one- to three-minute student surveys would be appropriate instruments. Student surveys offer greater insight into the classroom environment than teacher perception surveys, as students are more likely than teachers to accurately report what occurs in the classroom (Kane & Staiger, 2012). In addition, we anticipated that focusing on student voice would prove compelling to teachers. As a measurement instrument, student surveys require minimal training to administer and can be analyzed quickly. It is therefore no surprise that Carnegie has found that three-minute student surveys can be generative (Yeager et al., 2013).

Figure 1 illustrates the process we undertook to design and refine the surveys.

---

[2] In our current work, we are investigating the potential of embedding the measures in ongoing professional learning for teachers, including the potential of different routines for supporting teachers to interpret and act on the basis of the resulting data.

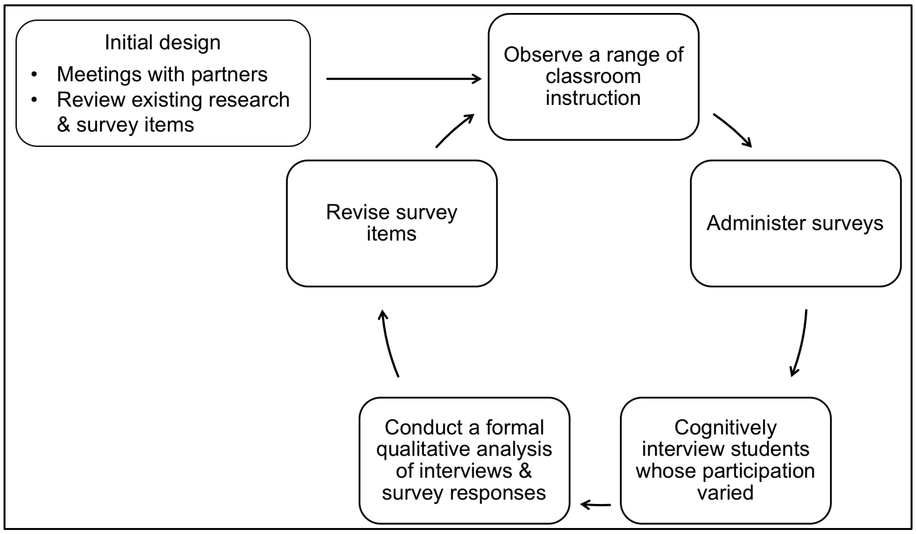

Figure 1. The Process of Designing, Trialing, and Revising the Surveys.

Initial Design of Survey Items

To generate the initial design of the surveys, we met with our partners to decide on the form of the measures (described above) and to generate initial survey items. The generation of survey items was informed by 1) a literature review of current findings related to the identified improvement focus (in our case, small-group and whole-class discussions) and 2) a review of currently available measures related to this improvement focus.

Literature review. Our literature review surfaced a number of key distinctions that informed the focus of the measures. As indicated above, the goals of classroom discussion include a) supporting students to develop conceptual understanding of key mathematical ideas, which entails understanding not just how to perform procedures but why procedures work and why they make sense in particular situations; b) supporting students' engagement in disciplinary practices of mathematics, such as argumentation; and c) supporting students to see themselves as having mathematical authority.

The literature clarified that small-group and whole-class discussions differ in their purposes. In a small-group setting, discussion is often meant to support students in grappling with important mathematical ideas, prior to deciding on a particular solution path to follow (Wood & Yackel, 1990). Small-group discussions also provide teachers with an opportunity to assess how their students are thinking, and to plan for the upcoming whole-class discussion (Stein, Engle, Smith, & Hughes, 2008). Whole-class discussions often follow after students have had time to work on tasks, and serve the purpose of advancing students' understanding of central mathematical ideas. Productive discussions are ones in which students not only share their thinking, but are also pressed to make connections between various solutions and the key mathematical ideas targeted by the lesson (Stein et al., 2008). Whole-class discussions often take the form of "show-and-tell" in which a student or students describe how they solved a problem, with minimal opportunity for the listening students to make sense of what has been shared and/or to make connections between the shared solution and key mathematical ideas (Ball, 2001). As a consequence, show-and-tell forms of discussions are not likely to advance the listening students' learning.

As we have indicated, the extent to which a small-group or whole-class discussion advances students' learning depends crucially on the role of the teacher. One key aspect of this role is to guide the establishment of norms regarding the purpose of discussion, as well as how students should interact with one another to ensure that they are indeed listening to and engaging with their peers' ideas (Horn, 2012; Kazemi & Stipek, 2001). This is equally important in small-group settings as it is in whole-group settings. However, the norms established in the two types of discussions will likely vary given their different purposes. For example, in a small-group setting, it is important to establish that the purpose of discussion is to make sense of problem-solving situations and to share initial thoughts on viable solutions. Students might share solution strategies, but only after members have had an opportunity to develop some common ground regarding their interpretations of the problem. It is critical to establish that "mistakes" should be viewed as opportunities to learn, not something to avoid or to be embarrassed by. It is also critical to establish norms of participation, such that all students have opportunities to share their thinking, and that no one student dominates discussion at the exclusion of others.

In a whole-class setting, it is important to establish that the purpose of the discussion is to make sense of students' solutions, as opposed to focusing primarily on the "right answer." Making sense of students' solutions entails asking for clarification regarding why students performed particular steps. In addition, it is important to establish that the purpose of whole-class discussion is to support students in making mathematical connections across solutions and representations. This entails establishing norms concerning asking questions regarding how solutions might relate to one another. In addition, it is often valuable pedagogically to discuss solution strategies that are incorrect or underdeveloped. As such, just as with small-group discussions, it is critical to establish norms regarding the value of mistakes in mathematics learning.

Another role of the teacher entails pressing students to engage in the forms of practice described above, and to provide conceptually-oriented explanations of their thinking, especially in whole-class discussions (Kazemi & Stipek, 2001). As it is unlikely that students (at least initially) will spontaneously provide such explanations, it is essential that the teacher supports and presses students to explain and justify their reasoning in ways that other students can understand (Cobb, 1998; Thompson, Philipp, Thompson, & Boyd, 1994). It is unlikely that a teacher can play this role in small-group discussions, given time factors. However, as noted above, it is important that the teacher establish norms that students should press one another to clarify their thinking, and to justify their proposed solution paths.

Thus far, we have described qualities of discussion that advance all students' learning. There is also evidence that the quality of discussion depends greatly on the nature of the tasks on which the students are working (e.g., Stein & Lane, 1996). If students are asked to complete purely procedural tasks, it is highly unlikely that discussion will focus on the concepts underlying procedures. On the other hand, if students are asked and supported to solve more challenging tasks (e.g., tasks with multiple solution paths, tasks that asks students to explain their reasoning, tasks that asks students to represent an idea in multiple ways), there is much greater potential for discussion that supports conceptual understanding and disciplinary reasoning.

Review of available measures. Unfortunately, we did not locate any student surveys designed to assess the quality of mathematics classroom discussion or the level of challenge of tasks. However, several different student perception surveys, including those developed in the Measures of Effective Teaching study (Bill & Melinda Gates Foundation, 2012) informed the structure of our student survey questions.

We generated a number of possible items specific to the various distinctions of small-group and whole-class discussions in mathematical classrooms described above. When writing the survey items, we used language that we anticipated would communicate to upper elementary and middle-grades students. We then carried out the cycle shown in Figure 1 and described below to trial and revise the survey items. Our primary goals were to ensure that 1) students' understandings of the items were consistent with those that we intended; and 2) students' responses to the items matched with our observations of the quality of discussion in the classroom. We carried out this cycle five times to develop these particular measures.

Observe a Variety of Classroom Settings

We worked closely with our districts partners to trial the items in middle-grades and upper-elementary classrooms that reflected variation in the quality of whole-class discussions. As a rule of thumb, in any cycle, we aimed to trial the surveys in two to four classrooms, half of which were characterized by procedurally-oriented activity, and half of which were characterized by conceptually-oriented activity. This was important because the surveys needed to distinguish between classrooms in which the quality of small-group and whole- class discussions varied if they were to have any value. Having selected classrooms and approached teachers for their permission, we then assembled a team of five to six observers that included district math leaders, school based math coaches, and researchers. We then observed instruction and sat with different groups of students, taking note of the tasks and the kind of discussions that ensued in the small group of students nearest to each of us and in the whole class.

Administer Surveys

At the end of class, one of our team members administered the surveys to the students (both the small group and whole class if both formats were included in the lesson). Students typically took one to three minutes to complete each survey, as we intended.

Cognitively Interview Students

Each team member identified a student who they had observed closely to participate in an audio-recorded cognitive interview (Desimone & LeFloch, 2004) that typically took 15 minutes. Each interviewer talked through the student's survey responses, item by item, asking the student to explain his or her response choices; this allowed us to compare our own observation of classroom discussions with the student's interpretation of the discussions. We also probed on the students' interpretations of individual items; this allowed us to surface cases in which language did not communicate as intended.

Analyze Interviews in Light of Observations

We then analyzed the cognitive interviews in light of our classroom observations. Our goal in this analysis was to identify mismatches between the quality of the observed small-group and whole-class discussions, and students' responses. In addition, we developed possible explanations for mismatches, and proposed revisions on the basis of those explanations. Analysis took place in two phases. The first phase was informal. Immediately after the cognitive interviews, our team met to debrief what we heard from the interviewed students . We flagged cases in which the language in items was confusing, and cases in which students' responses were at odds with what we had observed. The second phase was formal. Two of the authors, Kochmanski and Nieman, listened to all of the recordings and tracked problematic wording.

They also compiled survey responses across the classrooms so that we could identify whether the surveys were indeed distinguishing the quality of discussion across classrooms. As an example, in earlier iterations of the whole-class discussion surveys, we included items that asked whether students provided explanations that described only how they solved problems, or whether they provided explanations that described both how and why they solved a problem in a particular way. We found that the distinction between "how" and "why" appeared to resonate with students in classrooms in which the observed lesson was organized around a rigorous task, and the rigor of the task was maintained across the lesson. However, this was not the case for students in classrooms in which the observed lesson was organized around a task in which the goal was to apply a given procedure to solve a problem, or in which a rigorous task was proceduralized over the course of the lesson. In these classrooms, we observed that students, at best, explained how they solved a problem; however interviewed students repeatedly told us that their peers' explanations included a focus on how and why. The why, from the students' perspective, was clear and implicit – it was "to get the right answer."

Revise the Surveys

On the basis of our analysis, we eliminated, added, and/or refined survey items. For example, the analysis described above led us to write items that asked students to assess the *purpose*, or function, of discussions as opposed to the form they might take. Following this, we were in a position to trial the revised versions of the surveys. To access the current version of the surveys, along with annotated versions that describe the warrant for each item, see www.education.uw.edu/pmr.

Using Practical Measures for Improvement – Not Accountability

As we have emphasized, practical measures – including the ones we have described herein – are designed to measure and inform *improvement*, not for accountability purposes. For example, these practical measures of classroom discussion described in this paper are specifically designed to support teachers in improving their instructional practices, and instructional coaches and district mathematics leaders in supporting them to do so.

In our view, it would be counterproductive if school and district administrators attempted to force teachers to administer practical measures. These measures, and the resulting data, should not be used by school or district administration as a form of teacher evaluation. In such an instance, we anticipate administering the practical measures would become a compliance activity and teachers might, for example, teach the students how to respond to the items directly, which will result in the measures losing their utility. Careful thought needs to be given as to how to introduce the surveys to teachers so that they do not view them as accountability measures. As a rule of thumb, it is important that teachers see value in the measure; attention must therefore be given to supporting teachers to want to inquire into the nature of small-group and whole-class discussions in their classrooms prior to implementing the surveys. (That said, we have also found in initial uses that the surveys themselves support teachers to want to inquire further into their practice.)

References

Ball, D. L. (2001). Teaching, with respect to mathematics and students. In T. Wood, B. S. Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 11-22). Mahwah, NJ: Erlbaum.

Bill & Melinda Gates Foundation. (2012). Asking students about teaching: Student perception surveys and their implementation. Seattle, WA: Bill & Melinda Gates Foundation.

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*. Cambridge, MA: Harvard Education Press.

Cobb, P. (1998). Theorizing about mathematical conversations and learning from practice. *For the Learning of Mathematics, 18*(1), 46-48.

Desimone, L. M., & LeFloch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis, 26*(1), 1-22.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225-256). Greenwich, CT: Information Age Publishers.

Horn, I. S. (2012). *Strength in numbers: Collaborative learning in secondary mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Kane, T., & Staiger, D. O. (2012). Gathering feedback on teaching: Combining high-quality observations with student surveys and achievement gains. Seattle, WA: Bill & Melinda Gates Foundation.

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *Elementary School Journal, 102*(1), 59-80.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313-340.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50-80.

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. (1994). Calculational and conceptual orientations in teaching mathematics. In A. Coxford (Ed.), *1994 Yearbook of the National Council of Teachers of Mathematics* (pp. 79-92). Reston, VA: National Council of Teachers of Mathematics.

Wood, T., & Yackel, E. (1990). The development of collaborative dialogue within small group interactions. In L. Steffe & T. Wood (Eds.), *Transforming children's mathematics education: International perspectives* (pp. 244-252). Hillsdale, NJ: Lawrence Erlbaum.

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