Conceptualizing Teachers' Capacity for Learning Trajectory-Oriented Formative Assessment in Mathematics

Caroline B. Ebby and Philip M. Sirinides

TASK (Teachers' Assessment of Student Knowledge) is an online tool designed to measure teacher's capacity for learning trajectory-oriented formative assessment in mathematics, specifically focusing on their ability to analyze student work and make instructional decisions based on that work. Formative assessment has proved to be one of the most powerful current educational practices in terms of improving student learning (Black & Wiliam, 1998; Kluger & DeNisi, 1996). A meta-analysis of more than 250 studies on formative assessment indicates substantial evidence linking formative assessment with higher student achievement, with typical effect sizes ranging from an impressive 0.4–0.7 (Black & Wiliam, 1998). Yet numerous studies have concluded that teachers struggle to make effective use of student learning data (Datnow, Park, & Wohlstetter, 2007; Heritage, Kim, Vendlinski, & Herman, 2009; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Young, 2006).

While the term formative assessment is often used erroneously in educational contexts to refer to assessment instruments themselves, it is more accurately defined as a process whereby an assessment provides feedback to both the learner and the teacher and this feedback causes an adjustment in instruction (Bennett, 2014; Black & Wiliam, 1998; Shepard, 2008). Formative assessment is therefore fundamentally an interpretive process. Effective formative assessment—assessing student understanding relative to a standard or goal, providing feedback to the student in the form of instructional guidance, and continually working to diminish the gap between the student's performance and the instructional goal—requires that teachers are able to understand and analyze student thinking to develop an instructional response that will move the learner forward. TASK is an open-ended measure situated in the context of looking at student-generated work that can be used to measure these specific aspects of teacher knowledge and also explore the nature of that knowledge.

C.B. Ebby (✉) • P.M. Sirinides
Consortium for Policy Research in Education, Graduate School of Education,
University of Pennsylvania, 3440 Market St, Suite 560, Philadelphia, PA 19104, USA
e-mail: cbe@gse.upenn.edu

J.A. Middleton et al. (eds.), *Large-Scale Studies in Mathematics Education*,
Research in Mathematics Education, DOI 10.1007/978-3-319-07716-1_8

In this chapter, we begin by articulating the conceptual framework behind learning trajectory-oriented formative assessment and describing the instrument, scoring rubrics, and ongoing development of TASK. We then present the results of a large-scale field test of TASK, both in terms of the overall results and additional studies of the properties of the instrument. We also draw on the results of this field test to investigate the relationships between various dimensions of teachers' ability to analyze student work in mathematics and their instructional decision making.

Conceptual Framework: Learning Trajectory-Oriented Formative Assessment

At the foundation of formative assessment is a clear understanding of the gap between the learner's current state of understanding and the learning goal or standard. A well-designed assessment should illuminate the learner's current state so that the gap is evident. The assessment becomes formative only when (1) the information provides useful feedback to the learner, (2) the information provides useful feedback to the teacher, and (3) the teacher is able to provide an instructional response that will help the learner move closer to the goal. This is an iterative process, the cycle repeating until the gap is closed and new learning goals are established (Bennett, 2014; Black & Wiliam, 1998; Heritage, 2008).

Learning progressions, or "successively more sophisticated ways of thinking about a topic" (National Research Council, 2007, p. 219), have recently become prominent in mathematics educational research as well as in conceptualizations of assessment and instruction (Clements & Sarama, 2004; Confrey, 2008; Daro, Mosher, & Corcoran, 2011; Sztajn, Confrey, Wilson, & Edgington, 2012). *Learning trajectories*, as they are most often called in mathematics education, can provide a guiding framework as teachers assess where students are in the trajectory of learning those concepts and skills and then use that information to design and enact instructional responses that support students' movement along that trajectory towards the learning goal (Heritage, 2008). Learning trajectories can be described as a path through the complex terrain of a particular mathematical topic (Battista, 2011; Daro et al., 2011). While this path is not necessarily linear, knowledge of the key stages or levels that characterize this path can help teachers both determine where students are and what experiences are likely to help them move forward. In other words, knowledge of learning trajectories can enhance the formative assessment process.

In conceptualizing the knowledge that teachers need to implement effective formative assessment in the classroom, we draw upon a conception of teaching as a complex activity that is dependent on distinct but interconnected bodies of knowledge (Ball, Thames, & Phelps, 2008; Putnam & Borko, 2000; Shulman, 1987). Arguing that teachers draw on knowledge that is distinct from either knowledge of subject matter, Shulman defines *pedagogical content knowledge* (PCK) as "the ways of representing and formulating the subject matter that make it comprehensible to others" (p. 9) and frames it as the intersection between content and pedagogy. Building on this work to study the work that teachers do when teaching mathematics in the classroom setting, Ball and colleagues have further defined and delineated mathematical knowledge for teaching (MKT) by breaking down the domain of content knowledge into *common content knowledge*, *specialized content knowledge,* and *horizon content knowledge* and pedagogical content knowledge into *knowledge of content and students*, *knowledge of content and teaching*, and *knowledge of content and curriculum* (Ball et al., 2008).

More recently, Sztajn et al. (2012) bring together research on learning trajectories with research on teaching to propose the construct of *learning trajectory-based instruction* as "teaching that uses student learning trajectories as the basis for instruction (p. 147)." In addition to presenting a learning trajectory interpretation of the six MKT categories, they define a learning trajectory interpretation of formative assessment as the case where teachers are "guided by the logic of the learner" rather than only by disciplinary goals when eliciting student thinking and providing feedback to students. In developing the TASK instrument and analyzing the results of the field test, we draw on these frameworks to explore how teachers actually make sense of evidence of student thinking for their instruction.

The TASK Instrument

We designed the TASK instrument to capture and explore teacher knowledge in relation to learning trajectories in several core mathematical content areas. Open-ended prompts were designed to elicit the information teachers glean from student work and the instructional response they develop based on that evidence. While the MKT is an established measure of "mathematics knowledge for teaching," these multiple choice measures have not been as useful in capturing teacher reasoning or more subtle manifestations of teacher conceptual change (Goldsmith & Seago, 2007). Hill, Ball, and Schilling (2008) describe the challenges of using multiple choice measures to assess "knowledge of content and students," or teachers' knowledge of mathematical thinking and learning, including the fact that performance can be influenced by test-taking skills or mathematical content knowledge. They conclude that open-ended items may be a more effective way to assess the kind of reasoning skills about student thinking that are called for in classroom-based instructional practice. We have developed, field tested, and validated TASK to provide a contextualized measure of teachers' ability to (a) analyze students' mathematical thinking within a grade-specific content area in relation to research-based learning trajectories, and (b) formulate effective instructional responses.

We began the development of TASK by first determining what kinds of knowledge are brought to bear in the process of formative assessment. Formative assessment involves a critical shift from *scoring* student work to *interpreting evidence* of student thinking and considering that evidence in light of research on the development of understanding of mathematical content. With this in mind, we initially posited that the following six domains of knowledge are relevant for *learning trajectory-oriented formative assessment*:

1. *Content Knowledge*—At the most basic level, teachers need to be able to understand and correctly solve math problems that assess the content they are teaching.
2. *Concept Knowledge*—To assess student understanding, teachers must be able to identify and articulate the concept and related sub-concepts that a particular mathematics problem or item is assessing.
3. *Mathematical Validity*—Once a teacher administers an assessment to a student, he/she must be able to understand the logic or mathematical validity of the strategy that the student uses to solve the problem.[1]
4. *Analysis of Student Thinking* (*AST*)—To build on student thinking, teachers need to be able to go beyond determining whether or not a response is correct or incorrect to identify the underlying conceptual understanding or misconceptions that are present in student work.
5. *Learning Trajectory Orientation* (*LTO*)—After analyzing the strategy a student uses to solve a math problem, teachers need to be able to position that strategy along a learning trajectory for the respective math content. Thus, teachers must have a sense of what the developmental progress looks like for the particular math concept and where to place students along that continuum and be able to use this as a framework to interpret and respond to student thinking.
6. *Instructional Decision Making* (*IDM*)—Finally, teachers must choose an appropriate instructional response and be able to describe why that instructional intervention is designed to move students from their current level of understanding along the developmental trajectory towards greater understanding.

To further explore these domains, we constructed a performance assessment that requires teachers to draw upon and articulate these types of knowledge in the context of classroom practice. Specifically, we situated TASK in the activity of looking at and responding to a carefully designed set of typical student responses to a mathematics problem in a particular content area. The student responses characterize different levels of sophistication of student thinking as well as common misconceptions that are supported by mathematics education research. Through an online instrument, teachers are presented with the student work and then led through a series of questions designed to measure these six key domains of knowledge related to the specific mathematical concept that is being assessed.

Seven TASK instruments have been developed in the following mathematics content areas: (1) *addition*, for teachers in grades K-1; (2) *subtraction*, for teachers in grades 2–3; (3) *multiplicative reasoning*, for teachers in grades 3–5; (4) *fractions*, for teachers in grades 3–5; (5) *proportional reasoning* for teachers in grades 6–8; (6) *algebraic reasoning* for teachers in grades 7–12; and (7) *geometric reasoning*, for

[1]As Ball et al. (2008) point out, determining whether a students' thinking is mathematically sound requires a kind of knowledge that a person with strong knowledge of mathematics content who is not a teacher may not necessarily possess. It is therefore distinct from common content knowledge.

Each carton holds 24 oranges. Kate's carton is 1/3 full. Paul's carton is 2/4 full. If they put all their oranges together, would Kate and Paul fill one whole carton?

Solve the problem. Show your work.

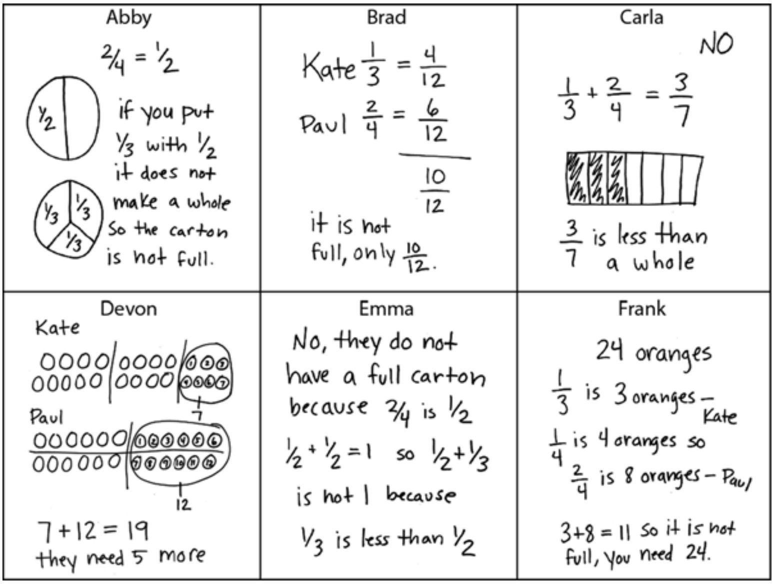

Fig. 1 Problem and designed student responses from the grades 3–5 fractions TASK

teachers in grades 9–12. These content areas represent core or fundamental mathematical ideas at the different grade level bands, and the TASKs are designed around key concepts in those domains (e.g., part/whole, equivalency, and magnitude for fractions). While the content areas are different across grade levels, all TASKs follow a consistent structure in both the prompts and the fact that the student work reflects key stages in the development of student thinking in the content area. The K-8 TASKs focus around six samples of student solutions; however, for algebra and geometry, since the problems have a higher level of complexity and longer student responses, there are only four samples of student work for the teacher to interpret.

An example of the different levels of sophistication of students' thinking and common strategies and misconceptions that are embedded in the student responses is presented in the fractions TASK for grades 3–5 in Fig. 1. The problem involves reasoning about whether two fractional quantities combine to make a whole. As shown in Fig. 1, Abby, Carla, and Devon's work reflect the use of visual models to make sense of parts and wholes, while Brad and Emma's work demonstrate more abstract reasoning about equivalence and addition. Carla, Devon, and Frank's work are less developed and contain misconceptions about partitioning, part/whole understanding, and the meaning of fractions. In this way, the student work represents some of the important landmarks that have been identified in current research on children's learning of fractions as well as an overall progression from concrete to more abstract understanding of fractional quantities (Confrey, 2008; Lamon, 2012; Steffe & Olive, 2010). Thus, TASK is designed to provide a realistic context from which to elicit information about what teachers pay attention to when they examine student strategies that they are likely to come across in their own classrooms.

Similarly for the other content areas, student work was constructed to represent key stages in the development of addition, subtraction, multiplication, proportional reasoning, and algebraic thinking, with student responses reflecting strategies of different levels of sophistication as well as strategies reflecting both procedural and conceptual errors.

The prompts are shown in relation to each dimension of knowledge and method of scoring in Table 1. Three of the response types are forced-choice or short answer and can be scored automatically while the rest are constructed responses scored by trained raters with a rubric or a combination of a coding scheme and rubric. The rubrics, described in the next section, are based on a four point ordinal scale to characterize the teachers' orientation towards the interpretation of the student work on a continuum that ranges from general to procedural to conceptual to developmental.

Table 1 TASK prompts and scoring

Domain of teacher knowledge	Prompt	Scoring	Scale
Content knowledge	Examine the math problem and state the correct answer	Automated	Correct/incorrect
Concept knowledge	Explain what a student at that grade-level needs to know and/or understand to solve the problem	Scored and coded by rater	Rubric score (1–4)
Mathematical validity	Examine the solutions of 4–6 typical students and determine if their solution processes are mathematically valid	Automated	Percent correct
Analysis of student thinking	Comment on four students' solution process in terms of what the work suggests about the student's understanding of the mathematics	Scored and coded by rater	Rubric score (1–4)
Learning trajectory orientation	Rank each student's solution of the level of sophistication of the mathematical thinking that is represented	Automated	Rubric score (1–4)
	Explain the rationale for the rankings given to each student	Scored by rater	Rubric score (1–4)
Instructional decision making	Suggest instructional next steps and explain the rationale for those next steps for two student solutions	Scored by rater	Rubric score (1–4)

As described above, the six samples of student work were constructed to represent both correct and incorrect solution strategies, common conceptual errors, as well as a range of sophistication of strategies. To prevent the instrument from becoming too time-consuming, respondents were asked to comment on a subset of four solution strategies, but then to rank and explain their ranking for all six. The four solutions represent beginning, transitional, and advanced strategies (with correct answers) as well as one solution that reflected a correct strategy with a conceptual error and incorrect answer. Likewise, respondents were only asked to describe instructional responses for two of the solutions: (a) a correct, but less sophisticated response to the problem and (b) a response with a conceptual weakness.

TASK was developed as an online instrument where teachers are sent an email link to complete the survey. Respondents move through several screens where the student work is shown as it is in Fig. 1 along with the respective prompts. Responses for mathematical validity and ranking are entered by clicking on radio buttons (see Fig. 2 below), while the open-ended responses for concept knowledge, analysis of student thinking, ranking-rationale, and instructional decision making are entered into text boxes. Respondents also have the option to expand their view of the student work by hovering the mouse over the image. A benefit of the online administration is that the system can target reminders to non-respondents to achieve a high response rate.

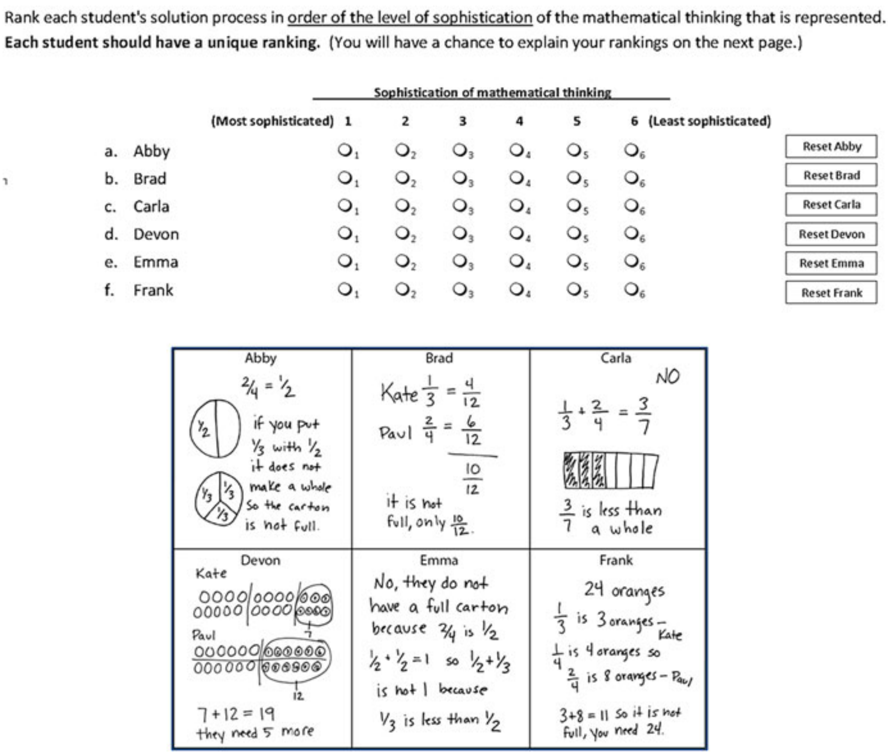

Fig. 2 Ranking screen of the online TASK

Scoring Rubric

The rubrics that raters used to score specific prompts about student work were based on a four-point ordinal scale to capture the overall orientation toward teaching or student understanding. We developed this rubric from the pilot data (described in the next section) through both an inductive and deductive process. First, a team of researchers read the entire set of responses to generate initial categories and codes to capture what teachers were referencing in their responses to each question. These codes were then grouped into larger categories, drawing on existing research in mathematics education to guide the analysis in terms of the degree to which the response reflected elements of a learning trajectory orientation. The distinction between procedures, or what students did, and concepts, or what students understood, became salient across all domains. The shift from procedural to more conceptual views of mathematics has long been promoted in mathematics reform literature (e.g., Hiebert, 1986; National Council of Teachers of Mathematics, 1988; National Research Council, 2001), and since learning trajectories by nature focus on conceptual development, a conceptual orientation toward student work was rated as higher than one that was only procedural. More recently, research on learning trajectories has promoted a developmental view, where students' conceptual knowledge develops in relation to instruction along a predictable path toward more complex and sophisticated thinking (Battista, 2011). Therefore, for a response to be at the highest level of the rubric, we determined that a teacher's focus on conceptual understanding must have evidence of drawing upon a developmental framework. We then had four ordinal categories (general, procedural, conceptual, and learning trajectory) that applied to each question on the TASK. The general rubric shown in Table 2 describes each of the TASK rubric categories. These categories are seen as cumulative where each level builds on the one before it; therefore, a conceptual response might also contain some procedural focus. Four domains were scored with more specific and detailed versions of this rubric: Concept Knowledge, Analysis of Student Thinking, Learning Trajectory Orientation, and Instructional Decision Making (Ebby, Sirinides, Supovitz, & Oettinger, 2013).

For Concept Knowledge and Analysis of Student Thinking, raters were asked to utilize a coding scheme organized in the form of a checklist, with descriptors under the main categories: general/superficial, procedural, conceptual, and learning trajectory. After the raters assigned the relevant codes, they used those results to help determine a rubric score. This technique also allows for tabulation of the specific concepts and procedures that are referenced by teachers which can be used to decompose patterns of teacher responses within each of the rubric categories.

The sample teacher responses shown below in Table 2 are taken from the pilot administration of the grades 6–8 proportions TASK. Teachers are describing a piece of student work where the strategy reflected a conceptual error stemming from additive thinking and led to an incorrect response. The sample teacher responses reflect different levels of analysis of that evidence: at the lowest level, the teacher evaluates the strategy but misses the nature of the conceptual error completely. At the procedural level, the teacher describes what the student did to get the incorrect answer, but

Table 2 TASK rubric levels and descriptions

Score	Category	Description	Sample response
4	Learning trajectory	Response draws on developmental learning trajectory to explain student understanding or develop an instructional response	Devon shows that he has some basic understanding of multiplicative reasoning when it comes to doubling both quantities of the rate. However, he then goes to additive reasoning to get to $20. He is not distinguishing the difference between multiplying and adding/subtracting in relation to proportionality
3	Conceptual	Response focuses on underlying concepts, strategy development, or construction of mathematical meaning	Devon has just a beginning understanding of a proportion as demonstrated by doubling both 12 and 15. He knew he needed to get to $20, but he didn't know how to use the proportion so he subtracted
2	Procedural	Response focuses on a particular strategy or procedure without reference to student conceptual understanding	Devon did not figure out the cost per can. He subtracted $10 so he also subtracted 10 cans but 1 can is not equivalent to $1
1	General	Response is general or superficially related to student work in terms of the mathematics content	Devon had a good strategy but did not perform the operations correctly

does not relate this to underlying conceptual understanding. At the conceptual level, the teacher recognized the proportional understanding in part of the student's solution strategy. The learning trajectory response is further distinguished by interpreting the students' use of doubling in terms of multiplicative reasoning and the conceptual error in terms of additive reasoning, an important distinction in established learning trajectories for proportional reasoning (e.g., Confrey, 2008; Lamon, 2012).

Ongoing TASK Development

TASK began with a pilot administration in the fall of 2011 with a convenience sample of 60 teachers and at least 10 responses at each grade band. The pilot data were used for two purposes. The first purpose was to begin development of the detailed scoring rubrics for each domain of the instrument and the second was to advance the design of the instrument. Both the actual responses and participant feedback contributed to our modifications of the instrument. Based on what we learned from this feedback, the instruments were substantially modified and scoring rubrics were developed.

In the spring of 2012, we administered TASKs in 6 content areas to a sample of about 1,800 teachers in grades K-10 from 5 public school districts in 5 states. Recruitment for this validation study used a stratified random sample of teachers by grade/subject; however, participation was voluntary. The five districts vary in terms of size, student demographics, and programs of math instruction. Overall, about 1,400 teachers completed the TASK,[2] for a 74 % response rate. Fifteen raters, including researchers with math content expertise and experienced mathematics teachers and coaches, were trained to code the open-ended responses for references to procedures and concepts and then make an overall judgment about a teachers' written response in relation to the four point rubric. TASKs were only scored by raters after they had established a reliability of at least 75 % direct agreement on all of the rubric scores with other reliable raters. Drawing from the results of the analysis of this field test data, much of which is described in the sections that follow, we have refined the TASK instrument to focus on the three most salient and robust domains: Analysis of Student Thinking (AST), Learning Trajectory Orientation (LTO), and Instructional Decision Making (IDM) while further streamlining the coding and analysis process.[3] In addition, we are developing multiple forms for repeated administrations as well as new TASKs in additional content areas.

Large-Scale Field Trial Results

Descriptive Statistics

Analysis of the field test data resulted in descriptive statistics for each of the domains on each TASK using unit weighting scoring as the average of scores within domain (Ebby et al., 2013). Across the domains examined on the TASK the majority of teacher responses were procedural, focusing on what the student did to solve the problem, rather than underlying conceptual understanding or sophistication of reasoning. Table 3 shows the breakdown of rubric scores for the domains of AST, LTO, and IDM. While these results are briefly summarized below, a more complete analysis of the descriptive results can be found in our interactive report (Supovitz, Ebby, & Sirinides, 2014).

Analysis of student thinking (**AST**). In this domain, the vast majority of teacher responses were procedural, focusing on what the student did to solve the problem rather than commenting on underlying conceptual understanding. Fewer than one fifth of the teachers surveyed, across all grade levels, interpreted the student solutions in terms of underlying conceptual understanding. The highest level of procedural responses were found in grades K-1 addition (93 %), while the highest level of conceptual and learning trajectory responses (19 %) were found in grades 3–5 fractions. Particularly striking is the fact that all of the responses for proportions in grades 6–8 were either general or procedural, with 21 % of the teachers providing

[2] Thousand two hundred and sixty-one fully completed TASKs in five content areas were analyzed from this field test. Responses to the geometry TASK have not yet been analyzed.

[3] For example, the latest version of the TASK for multiplicative reasoning includes some multiple choice questions to augment the open ended prompt for Instructional Decision Making.

Table 3 Percent of teacher responses by TASK domain, content, and score

Domain/content/grade	*n*	General	Procedural	Conceptual	Learning trajectory
AST					
Addition (K-1)	246	4	93	3	0
Subtraction (1–2)	185	13	76	11	0
Fractions (3–5)	376	7	73	18	1
Proportions (6–8)	291	21	79	0	0
Algebra (9–10)	163	9	88	3	0
LTO (rationale)					
Addition (K-1)	246	0	83	17	0
Subtraction (1–2)	185	4	69	22	5
Fractions (3–5)	376	9	76	14	1
Proportions (6–8)	291	14	78	7	1
Algebra (9–10)	163	15	57	26	2
IDM					
Addition (K-1)	246	13	79	8	0
Subtraction (1–2)	185	19	59	20	2
Fractions (3–5)	376	22	60	16	2
Proportions (6–8)	291	28	55	14	3
Algebra (9–10)	163	15	46	30	9

Note: 1,261 teacher responses to the TASK were collected in spring 2012

only general analyses of student work (e.g., "understands proportions" or "demonstrates strong reasoning."). The results highlight the widespread lack of a conceptual focus in teachers' analysis of student thinking around proportions among middle grades teachers.

Learning trajectory orientation (**LTO**). Again the vast majority of teachers explained their ranking of student work by pointing to procedural aspects of student work rather than what students understood or how that understanding was situated in a learning trajectory. It should be noted that teachers were somewhat more successful in choosing the ranking than they were in providing a reasoned rationale for that ranking, though fewer than half of teachers in grades K-8 were able to correctly order student strategies in terms of sophistication.

Instructional decision making (**IDM**). Across all grade levels, the majority of teachers' instructional suggestions for specific students focused on teaching a student a particular strategy or procedure rather than on developing mathematical meaning or understanding. The percentage of teachers who gave conceptual or learning trajectory responses was highest for algebra and lowest for addition in grades K-1.

Together, these results suggest that there is a great deal of room for growth in relation to teacher's ability to interpret and respond to conceptual understanding in student work, and even more so in relation to learning trajectories. We also used these results to provide information back to the participating districts in the form of reports that detailed the relative proportion of teachers at each grade level who responded at each level of the rubric in each domain. These reports were designed to allow districts to view both strengths and weaknesses in their teachers' capacity for learning trajectory-oriented formative assessment.

Instrument Properties

The design of the instrument and validation methods were directly influenced by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), which provides strong guidelines for high-quality and technically sound assessments. Our methods of ongoing instrument validation were chosen to supply evidence that the resulting scores from this theoretically grounded instrument are reliable and valid for the purposes of evaluating teachers' capacity for learning trajectory-oriented formative assessment in mathematics. Unless otherwise noted, data for these analyses were collected from the large-scale field trial described above. The technical report (Ebby et al., 2013) provides more details about the measurement studies for the TASK.

To examine the validity of TASK scores as a measure of pedagogical content knowledge, we have analyzed its association with another similar established test, the measures of Mathematical Knowledge for Teaching (MKT) (Ball et al., 2008; Hill, Schilling, & Ball, 2004; Schilling, Blunk, & Hill, 2007). The MKT is a measure of the Common Content Knowledge and Specialized Content Knowledge that teachers need for effective mathematics instruction. The MKT is most aligned with the TASK domains of Content Knowledge and Mathematical Validity, but we expect that there would still be a positive, though smaller, relationship with the other domains, for which no validated measures exist. In the technical report (Ebby et al., 2013), we present descriptive statistics and correlation matrices for domain scales and the MKT separately for each TASK based on a sample of 486 teachers across the five districts. We find that the statistical associations of MKT and TASK domains reflect a low relationship and note that correlations are largest ($r=0.56$) for TASK domains with the most variance.[4] The positive direction and low magnitude of the statistics suggests that the constructs are related but distinct from MKT.

Collectively, results from a series of ongoing instrument validation studies are generating evidence that the instrument yields reliable and valid scores of teachers' learning trajectory-oriented formative assessment capacity in mathematics, is feasible for widespread use in a variety of settings, and provides useful reporting of results. Ongoing research studies will focus on how TASK can be used to measure change in teacher knowledge over time and whether it can predict student outcomes.

[4] We are mindful that score reliabilities for the TASK are still under investigation and that correlations may be underestimated in the presence of measurement error (i.e., attenuation) (Lavrakas, 2008).

Pathways Analyses

In this section, we highlight an investigation of the relationships between the various dimensions of teachers' ability to analyze student work in mathematics and their instructional decision making. More specifically, we investigate the relationships between: (1) mathematical validity; (2) analysis of student thinking; (3) learning trajectory orientation; and (4) instructional decision making.[5] The theoretical framework is based on the research literature in mathematics education and our hypothesis that analyzing student work for underlying conceptual understanding should contribute to a more sophisticated instructional response. Given the current focus on learning trajectories in mathematics education research, and standards, we also investigated whether the ability to place student work in a learning trajectory would have an effect on instructional decision making, and if so, how strong that relationship is compared to other dimensions.

The conceptual framework guiding the empirical study is summarized by the structural pathways in Fig. 3. This framework includes a series of relationships among independent and dependent constructs, which characterize the mechanism through which the analysis of student work influences instructional decision making. The analysis of student work in terms of Mathematical Validity (MV), Analysis of Student Thinking (AST) and Learning Trajectory Orientation (LTO) is theorized to affect instructional decision making (IDM). Additionally, AST is theorized to be

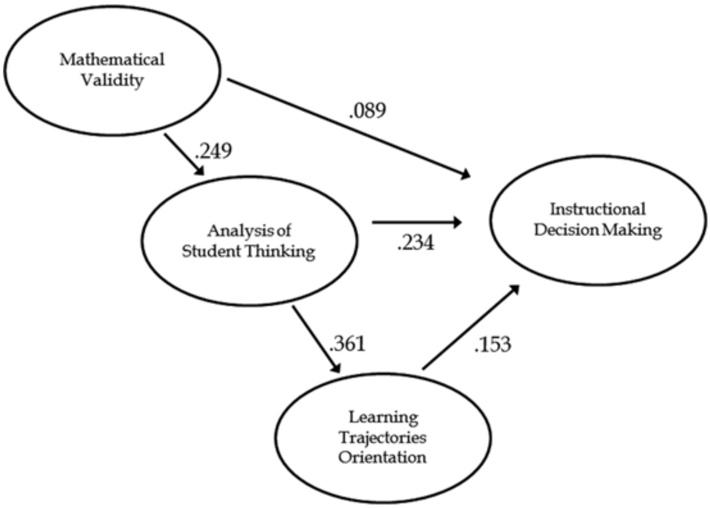

Fig. 3 Conceptual model of teachers' assessment of student knowledge and instructional decision making

[5] We do not include the domains of content knowledge or concept knowledge in this analysis as we do not expect them to have as strong of an influence on instructional decision making.

indirectly predictive of IDM through its effect on LTO. Finally, MV indirectly affects IDM through AST's direct and indirect paths to IDM. It is important to note that our method of analysis (described below) cannot be employed as a causal modeling approach because it cannot satisfy assumptions of directionality (Duncan & Hodge, 1963) or spurious correlation (Simon, 1957). As such, study findings do not adopt a causal interpretation, in the sense of confirming a presumed hypothesized network of causation. Rather this study sheds light on the tenability of the theorized causal model and results may be used as grounds for future research to further investigate causal mechanisms that are implied by this correlational study.

For this study, a statistical modeling approach was needed to meet several analytic goals. First, the study of indirect effects required the modeling of mediating variables. Path analysis (Wright, 1934) met this need because it offered a single framework for a system of multiple equations. Another analytic goal was the inclusion of latent variables in the model. This study examined relationships among theorized dimensions that pertain to learning trajectory-oriented formative assessment, which are not measured directly, but rather are measured by the TASK using a set of indicators and rubrics. Structural equation modeling (SEM) expands the path analysis framework to include a measurement model. In the measurement component of the SEM, latent variables are modeled as exogenous predictors of multiple observed items. The structural component of SEM specifies relationships among latent or observed variables. A benefit of using SEM is that both the measurement model and the structural model are estimated as one system of equations.

An empirical model was specified according to the structural pathways in the conceptual model (Fig. 3). Each of the four hypothesized domains were modeled as unobserved factors represented by the ten constituent rubric scores using all 1,261 complete TASK records. The latent factors were identified in the model by assigning each a variance of one, making the factor covariance interpretable as a factor correlation. The observed data were analyzed as continuous outcomes and the pathways between factors were freely estimated parameters. The model was estimated using Full Information Maximum likelihood using MPlus 7.1.

The full structural equation model defined by the structural pathways specified by our conceptual model was estimated and did not meet conventional thresholds for model fit, with a significant overall model chi square statistic and $\text{RMSEA} = 0.18$ (recommended < 0.10). Despite the marginal fit of the model, we find that all path coefficients in the structural model are statistically significant and consistent with the hypothesized direction of the relationships. The three antecedent dimensions accounted for 23 % of variation in instructional decision making and the estimated direct, total indirect, and total effects presented in Table 4 are all statistically significant at $p < 0.05$. Table 5 presents the standardized estimated correlation matrix for the latent variables.

Estimated correlations between TASK domains are positive, as expected. Further, all correlations are low suggesting that the measured domains are not highly associated (Cohen, 1988). The direction and magnitude of the statistics across TASK instruments suggests that the domains we are measuring are distinct. Across the subject areas, we observe that the largest correlations are between

Table 4 Standardized direct, indirect, and total effects on instructional decision making

	Direct	Total indirect	Total
MV	0.089	0.072	0.161
AST	0.234	0.055	0.289
LTO	0.153	–	0.153

Note: All estimates are significant at $p < 0.05$

Table 5 Standardized estimated correlation matrix for the latent variables

	MV	AST	LTO	IDM
MV	1.00			
AST	0.224	1.00		
LTO	0.058	0.259	1.00	
IDM	0.118	0.179	0.114	1.00

the domains of Analysis of Student Thinking (AST) and Learning Trajectory Orientation (LTO). In addition, Analysis of Student Thinking (AST) is more strongly correlated with each of the other domains, particularly in grades K-5, suggesting that overlap exists between this domain and the other domains.

These results are interpreted as preliminary findings that will inform ongoing instrument development as well as alternative conceptual models that may improve the fit of the measurement and structural components using a second round of multi-district TASK data. The ability of a teacher to analyze student thinking in terms of conceptual understanding was the largest predictor of instructional decision making in both its direct and total effect. A teacher's ability to assess the mathematical validity of student work is also predictive of IDM with nearly half of the total effect being mediated by their analysis of students' thinking and learning trajectory orientation. Overall, these findings provide preliminary evidence that a teacher's depth of understanding of student thinking may have the largest total effect on instructional decision making in terms of the degree to which these decisions draw upon learning trajectories with a significant amount of that relationship being mediated by the teachers' learning trajectory orientation.

These results confirm and add to some of the existing findings of qualitative studies of the relationship between teachers' interpretation of student work and their ability to develop informed instructional responses. In studying teachers' use of interim test data, Goertz, Oláh, and Riggan (2009) found that teachers who interpreted student errors conceptually, rather than only procedurally, were more likely to generate substantive instructional responses. Similarly, in analyzing teacher logs, Riggan and Ebby (2013) found a clear linkage between the way teachers analyze their student work and the nature of the instructional responses they develop. Teachers who described student work in terms of conceptual understanding were more likely to state that they would reteach the content differently using strategies that were tailored to the individual student. Adding to this research base, our analysis of TASK suggests that the depth of teachers' interpretation of student work is moderately related to their tendency to develop a learning trajectory-oriented instructional response.

Building Capacity for Effective Mathematics Instruction

TASK was developed as a tool for researchers and evaluators to assess teacher capacity for learning trajectory-oriented formative assessment and the impact of initiatives that seek to develop that capacity. The development of TASK and the various ongoing studies described in this chapter has led to some key findings about the instrument itself, the current capacity of teachers to interpret student thinking in relation to learning trajectories, and the nature of the knowledge that teachers need for effective mathematics instruction. TASK was developed to explore and measure an understudied component of mathematical knowledge for teaching: teacher knowledge in the context of formative assessment. Taken together, our analyses highlights three key domains—analysis of student thinking, learning trajectory orientation, and instructional decision making—that advance the conceptualization of the teacher knowledge required for learning trajectory-oriented formative assessment.

Our results offer empirical evidence that teachers' tendency to analyze student thinking for underlying conceptual understanding is related to their ability to develop instructional responses that build on the current state of students' thinking to move them towards more sophisticated understanding. Yet the vast majority of teachers surveyed across grade levels analyzed student work procedurally, in terms of what students did to solve the problem, rather than in relation to underlying conceptual understanding. Given the current emphasis in mathematics education on rigor as a balance between conceptual and procedural understanding, this suggests that there is a great deal of room for growth in teacher capacity to identify, interpret, and respond to students' conceptual understanding. Furthermore, results point to understanding a learning trajectory orientation, or the ability to order different student strategies in terms of the sophistication of mathematical thinking, as rooted in analysis for conceptual understanding and an immediate predictor of instructional decision making.

The TASK instrument is thus an important step towards identifying more precisely the components of teacher knowledge that can influence and potentially improve classroom instruction. TASK also represents an important new tool for researchers in mathematics education that has the capability to gauge more than just content knowledge or general pedagogical content knowledge. The ability to measure teacher knowledge, capacity, and growth in relation to the understanding and use of learning trajectories will become increasingly important as states and districts work to train teachers to reach the goals of new and more rigorous standards.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: Authors.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 398–407.

Battista, M. T. (2011). Conceptualizations and issues related to learning progressions, learning trajectories, and levels of sophistication. *The Mathematics Enthusiast, 8*, 507–570.
Bennett, R. E. (2014). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.
Black, P. B., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.
Clements, D., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81–89.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Confrey, J. (2008). *A synthesis of the research on rational number reasoning: A learning progression approach to synthesis*. Paper presented at The International Congress of Mathematics Instruction, Monterrey, Mexico.
Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. (Research report no. 68). Madison, WI: Consortium for Policy Research in Education.
Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles: Center on Educational Governance, University of Southern California.
Duncan, O. D., & Hodge, R. W. (1963). Education and occupational mobility: A regression analysis. *The American Journal of Sociology, 68*, 629–644.
Ebby, C. B., Sirinides, P., Supovitz, J., & Oettinger, A. (2013). *Teacher analysis of student knowledge (TASK): A measure of learning trajectory-oriented formative assessment*. Technical report. Philadelphia: Consortium for Policy Research in Education.
Goertz, M., Oláh, L., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction.* CPRE research report #RR-65. Philadelphia: Consortium for Policy Research in Education.
Goldsmith, L. T., & Seago, N. (2007). *Tracking teachers' learning in professional development centered on classroom artifacts*. Paper presented at the Conference of the International Group for the Psychology of Mathematics Education, Seoul, Korea.
Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Los Angeles: National Center for the Research on Evaluation, Standards, and Student Testing (CRESST).
Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practices, 28*(3), 24–31.
Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum.
Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.
Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*, 11–30.
Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496–520.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.
Lamon, S. J. (2012). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers* (3rd ed.). New York: Routledge. National Council of Teachers of Mathematics, 1988.
Lavrakas, P. J. (2008). *Encyclopedia of survey research methods: AM (Vol. 1)*. Thousand Oaks, CA: Sage.

National Council of Teachers of Mathematics. (1988). *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM.

National Research Council (Ed.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: The National Academy Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*, 4–15.

Riggan, M., & Ebby, C. B. (2013). *Enhancing formative assessment in elementary mathematics: The ongoing assessment project* (Working paper). Philadelphia: Consortium for Policy Research in Education.

Schilling, S. G., Blunk, M., & Hill, H. C. (2007). Test validation and the MKT measures: Generalizations and conclusions. *Measurement: Interdisciplinary Research and Perspectives, 5*(2–3), 118–127.

Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279–303). New York: Lawrence Erlbaum.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Simon, H. A. (1957). *Models of man*. New York: Wiley.

Steffe, L., & Olive, J. (2010). *Children's fractional knowledge*. New York: Springer.

Supovitz, J., Ebby, C. B., & Sirinides, P. (2014). *Teacher analysis of student knowledge: A measure of learning trajectory-oriented formative assessment*. Retrieved from http://www.cpre.org/sites/default/files/researchreport/1446_taskreport.pdf

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Towards a theory of teaching. *Educational Researcher, 41*(5), 147–156.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics, 5*(3), 161–215.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521–548.