



10-2013

## TASK Technical Report


Caroline Brayer Ebby  
*University of Pennsylvania, cbe@gse.upenn.edu*

Philip M Sirinides  
*University of Pennsylvania, sirinide@upenn.edu*

Jonathan A. Supovitz  
*University of Pennsylvania, JONS@GSE.UPENN.EDU*

Andrea Oettinger  
*University of Pennsylvania*

Follow this and additional works at: [https://repository.upenn.edu/cpre\\_workingpapers](https://repository.upenn.edu/cpre_workingpapers)

 Part of the [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Teacher Education and Professional Development Commons](#)

---

### Recommended Citation

Ebby, Caroline Brayer; Sirinides, Philip M; Supovitz, Jonathan A.; and Oettinger, Andrea. (2013). TASK Technical Report. *CPRE Working Papers*. Retrieved from [https://repository.upenn.edu/cpre\\_workingpapers/9](https://repository.upenn.edu/cpre_workingpapers/9)

[View on the CPRE website.](#)

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/cpre\\_workingpapers/9](https://repository.upenn.edu/cpre_workingpapers/9)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## TASK Technical Report

### Abstract

This report reviews the development, piloting, and preliminary results from the large-scale field trial of the TASK Instrument ([see cpre.org/task](https://cpre.org/task)). In the first section, we review the need for an assessment of teachers' capacity for learning trajectory-oriented instruction and the theoretical foundations that inform our work. We then describe the instrument and its development. Next, we detail the scoring process and the training of raters. The final section contains the analysis of the large-scale field trial conducted in 2012–13. We conclude with some directions for future work with this instrument.

### Disciplines

Curriculum and Instruction | Educational Assessment, Evaluation, and Research | Teacher Education and Professional Development

### Comments

[View on the CPRE website.](#)



CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

# Teacher Analysis of Student Knowledge (TASK): A Measure of Learning Trajectory-Oriented Formative Assessment

---

Caroline B. Ebby  
Phil Sirinides  
Jonathan Supovitz  
Andrea Oettinger

Consortium for Policy Research in Education  
University of Pennsylvania

Paper presented at the Annual Meeting of the American Educational Research Association,  
San Francisco, April 2013.

# Table of Contents

Overview.....	4
Theoretical Foundations.....	5
Learning Trajectories.....	5
Formative Assessment.....	5
New Directions, New Demands on Teaching.....	6
The TASK Instrument.....	7
Instrument Development.....	12
Pilot.....	12
Large-Scale Field Implementation.....	12
Scoring the TASK.....	13
Coding Procedure.....	13
Scoring of the Six Domains of Learning Trajectory-Oriented Formative Assessment Capacity.....	15
Training TASK Raters.....	24
Evidence of TASK Reliability and Validity.....	25
Inter-Rater Reliability Study.....	26
Internal Consistency Study.....	27
Test Criterion Validity Study.....	27
References.....	30
Appendix A. Coding Scheme Checklist for TASK Grades 3–5 Fractions.....	33
Appendix B. TASK Rater Training Procedure.....	34
Appendix C. TASK Domain Statistics by Subject.....	35
Appendix D. Correlation Matrices.....	36

## List of Figures and Tables

### *Figures*

Figure 1. The Formative Assessment Process.....	6
Figure 2. Problem and Designed Student Responses from the Grades 3–5 Fractions Task .....	8
Figure 3. Domains of Mathematical Knowledge for Teaching.....	10
Figure 4. TASK Rubric Levels .....	15
Figure 5. Analysis of Student Thinking: Sample Responses, Rubric Scores, and Rationale from Grades 3–5 Fraction TASK.....	18
Figure 6. Instructional Decision Making: Sample Teacher Responses, Rubric Codes, and Rationale from Grades 3–5 Fractions TASK .....	22
Figure 7. Simplified Learning Trajectory for Grades 3–5 Fractions TASK Training.....	25

### *Tables*

Table 1. TASK Prompts by Domain of Learning Trajectory-Oriented Formative Assessment and MKT ....	11
Table 2. Sample Composition .....	13
Table 3. TASK Prompts, Domains, and Scoring.....	14
Table 4. Rubric for Concept Knowledge .....	16
Table 5. Concept Knowledge: Sample Responses, Rubric Scores, and Rationale from Grades 3–5 Fractions TASK.....	16
Table 6. Rubric for Analysis of Student Thinking—Conceptual Understanding.....	17
Table 7. Automated Rubric for Learning Trajectory Orientation—Ranking .....	20
Table 8. Rubric for Learning Trajectory Orientation—Rationale .....	21
Table 9. Rubric for Instructional Decision Making.....	22
Table 10. Detailed Rubric for Instructional Decision Making .....	23
Table 11. TASK Inter-Rater Reliability .....	26
Table 12. Administration of TASK and MKT by Grade Band .....	28
Table 13. MKT Responses by TASK .....	28

## Overview

The Teacher Analysis of Student Knowledge (TASK) is an authentic, contextualized measure of teachers' ability to analyze students' mathematical thinking within a grade-specific content area in relation to research-based learning trajectories, and formulate effective instructional responses. Successful implementation of the Common Core State Standards in Mathematics (CCSSM) (Common Core State Standards Initiative, 2010), which were built upon existing research on student learning, will depend on teachers' ability to translate this information to instructional practice. The TASK instrument, therefore, has the potential to measure teacher instructional capacity in relation to the CCSSM.

TASK is a 25-minute, online assessment that asks teachers of mathematics to examine a set of carefully designed student responses to an assessment prompt, to explain what the responses show about student understanding, to order the student responses according to their developmental sophistication (i.e., the learning trajectories), and to suggest informed instructional responses. TASKs were originally developed in 2011 and refined over the past two years. They currently exist in six core mathematics content areas across grades K–12.

The TASK instrument has undergone two years of iterative development and field trials. Initial development and piloting of the TASK instrument has been supported by funding from the GE Foundation, as part of the Consortium for Policy Research in Education's (CPRE) evaluation of the GE Foundation's *Developing Futures™ in Education* program. The development work that CPRE has conducted to date includes instrument piloting, the development of scoring rubrics and procedures, and TASK administration to a random sample of approximately 1,400 mathematics teachers in 250 schools across five states. Using these data, we have conducted studies of inter-rater reliability and internal consistency, and correlated the TASK to a well-known measure of mathematical knowledge for teaching. The results point to the potential of the TASK instrument to measure important aspects of teachers' ability to implement the CCSSM in instruction that are not currently assessed by existing measures. Based on evidence of its reliability, such a measure could be used for evaluation research on the impacts of Common Core enactment and associated training and supports. Other potential uses of the TASK exist, such as identifying areas of professional development to target within a school or district, as the basis of professional learning community discussions, or as screening for teacher induction.

This report reviews the development, piloting, and preliminary results from the large-scale field trial. In the first section, we review the need for an assessment of teachers' capacity for learning trajectory-oriented instruction and the theoretical foundations that inform our work. We then describe the instrument and its development. Next, we detail the scoring process and the training of raters. The final section contains the analysis of the large-scale field trial conducted in 2012–13. We conclude with some directions for future work with this instrument.

## **Theoretical Foundations**

TASK is a measure of teachers' capacity for *learning trajectory-oriented formative assessment*. The idea of learning progressions, or "successively more sophisticated ways of thinking about a topic" (National Research Council, 2007), have recently become prominent in mathematics educational research, as well as conceptualizations of mathematics standards, assessment, and instruction (Battista, 2011; Clements & Sarama, 2004; Confrey, 2008; Daro, Mosher, & Corcoran, 2011; Empson, 2011; Szatjn, Confrey, Wilson, & Edgington, 2012). As we describe in this section, learning progressions, or *learning trajectories* as they are most often called in mathematics education literature, can provide a guiding framework for the process of formative assessment, one of the most powerful current educational practices in terms of improving student learning (Black & Wiliam, 1998; Crooks, 1988; Kluger & DeNisi, 1996; Natriello, 1987).

### ***Learning Trajectories***

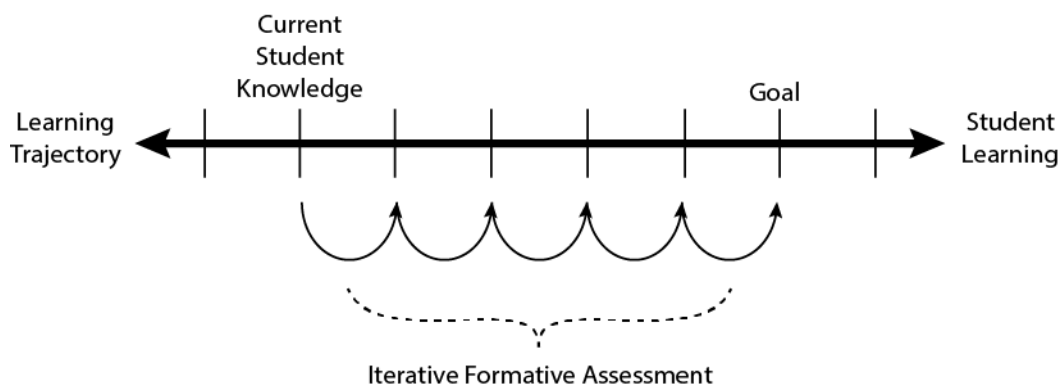
Research in mathematics education indicates that teachers who make sense of student thinking and incorporate this knowledge as a regular part of their instruction are better able to develop students' conceptual understanding of mathematics (Cobb, Boufi, McClain, & Whitenack, 1997; Stein, Engle, Smith, & Hughes, 2008). In the past decade, learning trajectories have been developed for several content topics, including early number, operations, geometry, measurement, multiplicative thinking, and rational number reasoning (Daro et al., 2011). While there are differences, particularly in grain size and level of detail, these trajectories are all research based and describe the development of conceptual understanding of core concepts, common student errors, and/or preconceptions or misconceptions that are precursors to more sophisticated thinking. As such, they can provide teachers with a clear articulation of learning goals, and the development of students' thinking, and inform learning activities that are likely to move students along the path toward achieving those goals, thereby linking research on learning with both assessment and instruction (Daro et al., 2011; Heritage, 2008; Szatjn et al., 2012). In the context of daily classroom practice, learning trajectories can provide teachers with a framework for analyzing the strategies students use to solve problems and then determining where students' current understanding is situated along the progression in order to make subsequent instructional decisions. This process is at the core of formative assessment.

### ***Formative Assessment***

Formative assessment involves assessing student understanding relative to a standard or goal, providing feedback to the student in the form of instructional guidance, and continuing to assess and, ideally, diminishing the gap between the student's performance and the goal (Ramaprasad, 1983; Sadler, 1989). The classical theory of formative assessment is based upon the theory of how teachers gain access to students' current state of understanding and move them toward a goal. According to Sadler (1989), "Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence" (p. 120). The essential activity of formative assessment is the use of some assessment mechanism to identify and change the gap between a learner's current knowledge and a desired goal (Ramaprasad, 1983; Sadler, 1989). That is, an assessment of student understanding becomes formative when its information is understood by the teacher and activated as feedback to the learner, altering the distance between the

present and desired state. As Ramaprasad (1983) states, “Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way” (p. 4). Though critical for improving student learning, formative assessment is challenging; numerous studies have concluded that teachers struggle to use assessment information to inform their own instructional practice (Black & William, 1998; Datnow, Park, & Wohlstetter, 2007; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Young, 2006).

**Figure 1. The Formative Assessment Process**



Well-designed student assessments can make student thinking visible, and through ongoing assessments, teachers can collect evidence and interpret student performance in relation to the known progression of student thinking toward the learning goal. Learning trajectories can provide a framework to guide teachers in making sense of the information collected from formative assessments. The formative assessment process is represented in Figure 1. In this figure, one can see the gap between the student’s current knowledge and the goal. The task of a teacher is to use information about the student’s current state of knowledge (often in the form of assessments) to understand where the student is currently. Then, utilizing knowledge of the appropriate learning trajectory, the teacher can provide instruction to move the student toward greater understanding. Subsequent information on the student’s state of knowledge (in the form of a subsequent assessment) can provide teachers with knowledge about whether the student has moved toward the goal, as represented by the progress being made in the student’s understanding and the sophistication of the strategies the student is using.

***New Directions, New Demands on Teaching***

The CCSSM, recently adopted by 45 states, have substantially increased expectations for both students and teachers. The CCSSM are designed to provide focus and coherence, balance of conceptual and procedural skills, emphasis on mathematical practices, and ambitious expectations for college and career readiness (Common Core State Standards Initiative, 2010). As stated in the introduction of the Standards, the “development of the standards began with research-based learning progressions detailing what is known today about how students’ mathematical knowledge, skill, and understanding develop over time” (p. 4). This focus on learning trajectories places new demands on teaching, as



teachers must not only understand the mathematical ideas and skills embodied in the Standards, but also 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. In other words, teachers must be able to implement formative assessment processes based on, and supportive of, the development of student thinking.

The ability to measure teacher knowledge, capacity, and growth in relation to the understanding and use of learning trajectories will therefore become increasingly important as states and districts train teachers to reach the goals of the CCSSM. While more established measures of "mathematics knowledge for teaching" expand the view of content knowledge to include several domains of both subject-matter knowledge and pedagogical content knowledge that are specifically required in the act of teaching mathematics (Hill, Schilling, & Ball, 2004), these multiple-choice measures have not been as useful in capturing teacher knowledge of learning trajectories or more subtle manifestations of teacher conceptual change, such as changes in the sophistication of teachers' mathematical analysis of student work (Goldsmith & Seago, 2007).

The current reform era of the Common Core State Standards (CCSS) requires a substantial increase in teachers' effective use of formative assessment based on learning trajectories. Mathematics teachers must have a deeper understanding of their students' conceptual understanding and the ability to teach with an explicit awareness of the developmental progressions that are embedded within the Common Core. The TASK instruments are designed to tap into these more ambitious expectations for teachers of mathematics.

## **The TASK Instrument**




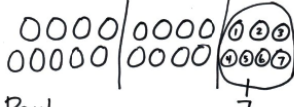
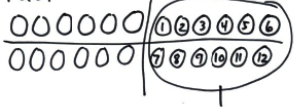
In order to measure teachers' capacity for learning trajectory-oriented formative assessment, the TASK presents a teacher with a carefully designed, grade-appropriate set of student responses to a mathematics problem. Student responses characterize different levels of sophistication of student thinking and misconceptions. Prompts ask the teacher to examine the mathematics problem and the students' solution strategies, to analyze students' thinking represented in their responses, and to provide subsequent instructional suggestions.

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 Figure 2. The problem involves reasoning about whether two fractional quantities combine to make a whole. As shown in Figure 2, 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, the TASK provides 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.

**Figure 2. Problem and Designed Student Responses from the Grades 3–5 Fractions Task**

**“Each carton holds 24 oranges. Kate’s carton is  $\frac{1}{3}$  full. Paul’s carton is  $\frac{2}{4}$  full. If they put all their oranges together, would Kate and Paul fill 1 whole carton? Solve the problem. Show your work.”**

<p>Abby</p> $\frac{2}{4} = \frac{1}{2}$  <p>if you put <math>\frac{1}{3}</math> with <math>\frac{1}{2}</math> it does not make a whole So the carton is not full.</p> 	<p>Brad</p> <p>Kate <math>\frac{1}{3} = \frac{4}{12}</math></p> <p>Paul <math>\frac{2}{4} = \frac{6}{12}</math></p> <hr style="width: 50%; margin-left: auto; margin-right: auto;"/> <p style="text-align: center;"><math>\frac{10}{12}</math></p> <p>it is not full, only <math>\frac{10}{12}</math>.</p>	<p>Carla</p> <p style="text-align: right;">NO</p> $\frac{1}{3} + \frac{2}{4} = \frac{3}{7}$  <p><math>\frac{3}{7}</math> is less than a whole</p>
<p>Devon</p> <p>Kate</p>  <p>Paul</p>  <p style="text-align: center;">12</p> <p><math>7 + 5 = 12</math> they need 5 more</p>	<p>Emma</p> <p>No, they do not have a full carton because <math>\frac{3}{4}</math> is <math>\frac{1}{2}</math></p> <p><math>\frac{1}{2} + \frac{1}{2} = 1</math> so <math>\frac{1}{2} + \frac{1}{3}</math> is not 1 because <math>\frac{1}{3}</math> is less than <math>\frac{1}{2}</math></p>	<p>Frank</p> <p>24 oranges</p> <p><math>\frac{1}{3}</math> is 3 oranges – Kate</p> <p><math>\frac{1}{4}</math> is 4 oranges so <math>\frac{2}{4}</math> is 8 oranges – Paul</p> <p><math>3 + 8 = 11</math> So it is not full, you need 24.</p>

Teachers are led through the following seven prompts:

- What is the correct answer to this problem? Briefly explain your answer.
- What does a student at your grade level need to know and/or understand in order to solve this problem?
- Examine the solutions of six students to the same math problem. Do you think each student’s solution process is mathematically valid? (Yes/No)

- Comment on each student's solution process in terms of what the work suggests about the student's understanding of *numbers and operations* (or for the high school TASKs, *algebraic thinking and reasoning* or *geometric thinking and reasoning*).
- Rank each student's solution process in order of the level of sophistication of the mathematical thinking that is represented.
- Explain why you ranked the mathematical thinking of [student name] number [rank] in relation to the responses of the other students.
- As a teacher, what would you do next with the two students below? [Abby and Devon] Please explain your rationale for the steps you suggest.

CPRE developed six versions of the TASK instrument in the following core mathematics content areas: *addition* for teachers in grades K–1, *subtraction* for teachers in grades 2–3, *fractions* for teachers in grades 3–5, *proportional reasoning* for teachers in grades 6–8, *algebraic reasoning* for teachers in grades 9–10, and *geometric reasoning* for teachers in grades 10–11. These content areas represent core or fundamental mathematical ideas at each of the grade levels and the TASKs are designed around key concepts in those domains (e.g., part/whole, equivalency, and magnitude for fractions). For the algebra and geometry TASKs, since the problems reflect a higher level of complexity that require longer student responses, the instrument contains only four examples of student work. 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.

Each online TASK takes a teacher approximately 25 minutes to complete. Teachers are led through a series of questions that measure six key domains of learning trajectory-oriented formative assessment related to the specific mathematical concept that is being assessed. These include:

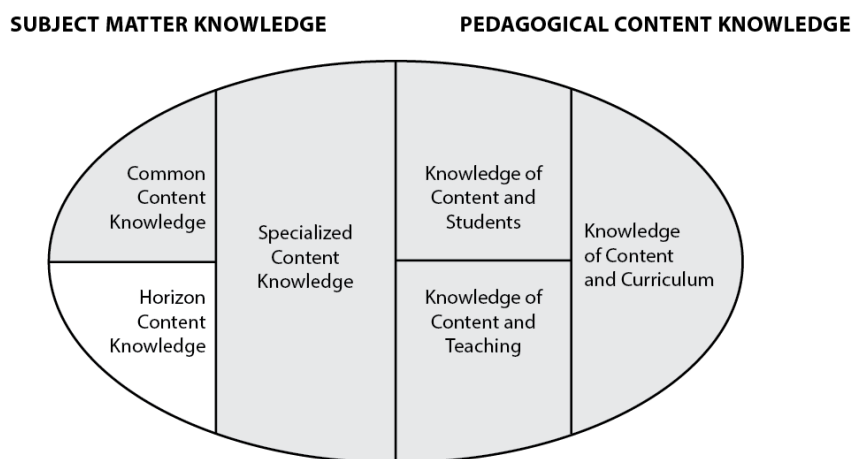
- 1) **Content Knowledge** - Teachers need to be able to understand and correctly solve mathematics problems that assess the content they are teaching.
- 2) **Concept Knowledge** - In order to assess student understanding, teachers must be able to identify and articulate the concept and related sub-concepts that a particular mathematics problem or item is assessing.
- 3) **Analysis of Student Thinking—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.
- 4) **Analysis of Student Thinking—Conceptual Understanding** - Teachers need to be able to identify the underlying conceptual understanding or misconceptions that are present in student work.
- 5) **Learning Trajectory Orientation** - 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 mathematics content. Thus, teachers must have a sense of what the developmental progress looks like for the particular mathematics concept and where to place students along that continuum. In addition to assessing teachers' ability to correctly order the student solutions

in terms of sophistication of thinking, the TASK assesses teachers' ability to use learning trajectories to justify that ranking.

- 6) **Instructional Decision Making** - 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 toward greater understanding.

The six domains measured by the TASK can be located in five of the domains in the framework of Mathematical Knowledge for Teaching (MKT)<sup>1</sup>, proposed by Ball, Thames, and Phelps (2008) (shown in Figure 3). While we do not aim to measure the MKT domains in their entirety, the TASK measures their application in the context of formative assessment that is informed by learning trajectories. Sztajn et al. (2012) also propose a Learning Trajectory-Based Instruction (LTBI) interpretation of the MKT categories that in many ways parallels our conceptualization of these domains. In order to show how the TASK aligns with MKT and LTBI interpretations of MKT, the specific prompts from the TASK are shown in Table 1 along with the corresponding domain of teachers' formative assessment capacity that are assessed by each set of prompts, as well as where these domains are located in MKT and the learning trajectory conceptualization of MKT. The coding and scoring of the prompts are described in the next section.

**Figure 3. Domains of Mathematical Knowledge for Teaching**



Note: Reproduced from Ball, Thames & Phelps (2008)

<sup>1</sup> TASK does not measure Horizon Content Knowledge, or the “awareness of how mathematical topics are related over the span of mathematics included in the curriculum” (Ball, et al, 2008, p. 403)

**Table 1. TASK Prompts by Domain of Learning Trajectory-Oriented Formative Assessment and MKT**

TASK Domain	Number of Prompts	TASK Prompt	MKT	LTBI interpretation of MKT
Content Knowledge <sup>2</sup>	1	Examine the mathematics problem and state the correct answer (grades 3 and up only),	Common Content Knowledge	Knowledge of concepts and procedures represented at each level of the trajectory to perform the tasks associated with each level, all the way to the overall mathematical generalization at the top of the trajectory.
	1	Explain how you solved the problem.		
Concept Knowledge	1	Explain what a student at your grade level needs to know and/or understand in order to solve the problem.	Knowledge of Content and Curriculum	Knowledge of how to use student voice to choose and adapt curricula that are built based on mathematical disciplinary perspectives.
Analysis of Student Thinking—Mathematical Validity	4–6 <sup>3</sup>	Examine the solutions of four to six typical students and determine if their solution processes are mathematically valid.	Specialized Content Knowledge	Knowledge of how to use one’s mathematical perspective to test the appropriateness of various solutions and representations learners propose in their own voice; unpacking each level of the trajectory, explaining the mathematical issues behind the levels.
Analysis of Student Thinking—Conceptual Understanding	4	Comment on four students’ solution processes in terms of what the work suggests about their understanding of number and operations (or algebraic/geometric reasoning).	Knowledge of Content and Students	Content knowledge intertwined with knowledge of how students think about, know, or learn particular content.
Learning Trajectory Orientation—Ranking	1	Rank each student’s solution in order of the level of sophistication of the mathematical thinking that is represented.	Knowledge of Content and Students	Knowledge of the various levels of the trajectories through which learners progress; knowledge of the cognitive steps that support development and of the ways learners approach certain tasks.
Learning Trajectory Orientation—Rationale	4–6 <sup>2</sup>	Explain the rationale for the rankings given to each student.		
Instructional Decision Making	2	Suggest instructional next steps and explain the rationale for those steps for a student who has a correct, but less-sophisticated response to the problem, and a student who demonstrates conceptual weakness in the response.	Knowledge of Content and Teaching	Knowledge of ways to support learners’ cognitive development along the trajectory to help students’ voices develop into mathematical perspectives; knowledge of how to select and target tasks to promote individual movement along the trajectory and content-rich classroom discourse.

<sup>2</sup> Not a central domain measured by the instrument.

<sup>3</sup> The number of prompts depends on the TASK, with addition, subtraction, fractions, and proportions having six and algebra having four.

A benefit of online administration of the TASK is that the system can target reminders to non-respondents to achieve a high response rate. Another benefit is that responses are entered directly into the master database, reducing errors from data entry, and increasing cost efficiency, and giving researchers real-time access to completed assessments. The TASK is markedly different from current teacher evaluation and assessment frameworks employed by researchers, states, and districts throughout the nation. Teachers who took the pilot form reported that they actually enjoyed completing the TASK because it presents them with real classroom situations and student work that is similar to what they see from their students. Teachers respond to prompts in their own words, not selections from fixed-choice items. As a result, the TASK allows teachers to demonstrate meaningful application of essential knowledge and skills, without leading them with prompts or intimidating them with questions that are perceived to have right or wrong answers.

## **Instrument Development**

### ***Pilot***

The initial pilot of the instrument occurred in Fall 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 as described in the next section. Adjustments to the instrument included changing the order of some prompts, rewording several prompts, and clarifying the instructions.

### ***Large-Scale Field Implementation***

In Spring 2012, we used the revised TASK instruments to conduct a larger field trial. This study was conducted by CPRE in partnership with five public school districts in five states. The districts varied in terms of size, student demographics, and programs of mathematics instruction. Table 2 presents the number of schools and the average number of students per grade, as well as student demographics. To achieve our final sample, we drew 1,851 teachers, of which 1,386 responded (a 74.9% response rate). Of completed TASKs, 6% were removed because teacher response data for two or more prompts were missing or so brief that it was not possible to be coded. Recruitment for this field trial used a stratified random sample of mathematics teachers by grade/subject, in which participation was voluntary, and we used incentives to reach the response goal. Teachers were stratified twice, first by district, and then by grade bands in elementary grades (i.e., K–1, 2–3, 4–5, 6–8) or mathematics subject in secondary grades (i.e., algebra, geometry). Teachers were sent several reminder post cards were offered a \$40 Amazon gift card upon completion of the survey. Data collection included 247 responses on the addition TASK,

185 responses on the subtraction TASK, 376 responses on the fractions TASK, 292 responses on the proportions TASK, 166 responses on the algebra TASK, and 83 responses on the geometry TASK.<sup>4</sup>

**Table 2. Sample Composition**

	District A	District B	District C	District D	District E
<b>District Size</b>					
# Schools	23	57	133	184	20
# Students	12,324	32,251	93,951	79,130	15,281
<b>Demographics</b>					
% White	47%	25%	53%	14%	39%
% Economically Disadvantaged	54%	73%	63%	82%	49%
% Limited English Proficient	8%	4%	6%	11%	14%
% Special Education	18%	20%	13%	20%	9%
<b>District Teachers</b>					
Sample Size <sup>b</sup>	325	376	394	438	318
# Teacher Respondents	273	274	268	329	242
TASK Response Rate	73%	84%	68%	75%	86%

Notes: District statistics reflect most recent published data from either 2010-2011 or 2011-2012 school year. <sup>b</sup> Random sample within district stratified by grade intervals; teachers with invalid e-mails were not replaced.

## Scoring the TASK

As described earlier and shown in Table 3, the TASK instrument collected teachers' responses to prompts or sets of prompts that assess six domains of teacher knowledge. Three of these response types were forced choice or short answer and could be scored automatically while the rest were constructed responses scored by trained raters with a rubric or a combination of a coding scheme and rubric. The procedures for scoring both types of responses are explained below.

### Coding Procedure

We developed a paperless process to link the web-based assessment to a web-based coding portal to score the teacher responses that required judgment by a trained rater. In the coding portal, the raters assigned codes and/or rubric scores to the teachers' written responses. The coding portal was designed to have raters score responses in batches of 10 by question, which increased efficiency and fostered independent assessments across the response of any individual respondent. Raters reported that they could score an average of five TASKs in one hour.

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 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

<sup>4</sup> We do not report geometry scoring in this report. It will be added in a later version.

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. 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

**Table 3. TASK Prompts, Domains, and Scoring**

Domain of Teacher Knowledge Assessed	Number of Prompts	Prompt	Scoring
Content Knowledge	1	Examine the math problem and state the correct answer.	Automated
	1	Explain how you solved the problem (grades 3 and up only).	Coded by Rater
Concept Knowledge	1	Explain what a student at that grade level needs to know and/or understand in order to solve the problem.	Scored and Coded by Rater
Analysis of Student Thinking—Mathematical Validity	4–6 <sup>5</sup>	Examine the solutions of four to six typical students and determine if their solution processes are mathematically valid.	Automated
Analysis of Student Thinking—Conceptual Understanding	4	Comment on four students' solution processes in terms of what the work suggests about the students' understanding of the mathematics.	Scored and Coded by Rater
Learning Trajectory Orientation—Ranking	1	Rank each student's solution in order of the level of sophistication of the mathematical thinking that is represented.	Automated
Learning Trajectory Orientation—Rationale	4–6 <sup>4</sup>	Explain the rationale for the rankings given to each student.	Scored by Rater
Instructional Decision Making	2	Suggest instructional next steps and explain the rationale for those steps for a student who has a correct, but less-sophisticated response to the problem, and a student who demonstrates conceptual weakness in the response.	Scored by Rater

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 promotes 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, in order for a response to be at the highest level of the rubric, we determined that a teacher's focus on conceptual understanding must be organized into a developmental framework. We then had four ordinal categories (general, procedural, conceptual, and learning trajectory) that applied to each

<sup>5</sup> The number of prompts depend on the TASK, with addition, subtraction, fractions, and proportions having six, and algebra having four.

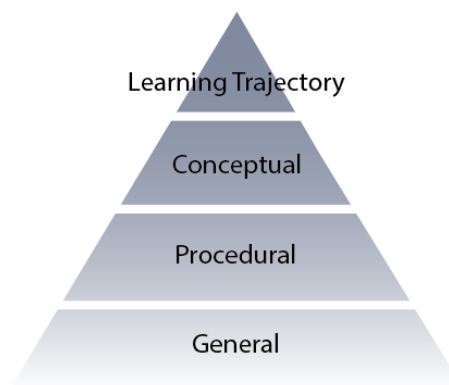


question on the TASK. The general rubric shown in Figure 4 describes each of the TASK rubric categories. These categories are seen as cumulative where each level builds on the one before it; therefore, a conceptual response might also contain some procedural focus.

As described below, the four domains were scored with more specific versions of this rubric: Concept Knowledge, Analysis of Student Thinking, Learning Trajectory Orientation—Rationales, and Informed Instructional Decision Making.

**Figure 4. TASK Rubric Levels**

Score	Category	Description
4	Learning Trajectory	Response draws on developmental learning trajectory to explain student understanding or develop an instructional response.
3	Conceptual	Response focuses on underlying concepts, strategy development, or construction of mathematical meaning.
2	Procedural	Response focuses on a particular strategy or procedure without reference to student conceptual understanding.
1	General	Response is general or superficially related to student work in terms of the mathematics content.



### ***Scoring of the Six Domains of Learning Trajectory-Oriented Formative Assessment Capacity***

#### ***Content Knowledge***

Teachers were asked to determine the correct answer to the mathematical problem in the TASK before they were shown any student responses, and in grades 3 and up, they were asked to describe how they would solve the problem. Particularly in the upper grades, a teacher’s answer to the problem provides a crude indicator of content knowledge. Since the responses were numerical, they could be automatically scored as correct/incorrect. Incorrect responses provide an indication of the teacher’s weakness in content knowledge; however, since this was the only item that addressed content knowledge, and we do not have a way to assess the range and variation in the teacher’s content knowledge, we do not consider it as one of the main domains measured by the TASK. In grades K–2, since the problems were simple addition and subtraction word problems, we did not ask teachers to describe how they solved the problem. (While children are likely to use a variety of strategies on these problems, adults would know the answer mentally.) However, on the fractions, proportions, algebra, and geometry TASKs (grades 3 and up), we did ask teachers to describe how they solved the problem. We developed a coding scheme to capture the range of strategies used and then trained raters to identify these strategies. The instrument therefore captures qualitative data on teachers’ strategies. While we have not included

these data in our current analyses, the codes could be useful for examining related questions (e.g., how do teachers' own strategies interact with the way they look at student strategies?).

### Concept Knowledge

As shown in Table 4, the general rubric was adapted so that teachers' written responses could be given an ordinal score from 1 to 4 in relation to the extent to which the response focused on the underlying conceptual understandings in the problem itself. For this question, the score level of 2 was subdivided into categories of 2P (procedural focus) and 2C (procedural with general conceptual focus). This distinction became salient during the coding process as raters frequently encountered responses that contained some reference to conceptual understanding but were not articulated sufficiently to warrant a score of 3. Thus, there are actually five levels to the rubric score for this prompt.

**Table 4. Rubric for Concept Knowledge**

<b>Prompt: What does a student at your grade level need to know and/or understand in order to solve this problem?</b>		
4	Learning Trajectory	<b>Informed focus on understanding.</b> Includes more than one articulated concept.
3	Conceptual Focus	<b>Focus on underlying concepts.</b> At least one concept is mentioned at an articulated level or multiple concepts at the general level.
2C	Procedural/Conceptual Focus	<b>Focus on how the problem is solved</b> but one concept is mentioned at the general level with no explication (e.g., part/whole)
2P	Procedural Focus	<b>Focus on how the problem is solved</b> using a procedure or specific strategy
1	General/Superficial	<b>Focus on general topic</b> (e.g., fractions, ratio, addition, subtraction, problem solving, reasoning, number sense, etc.)

Table 5 provides sample teacher responses with the scoring rationale from the grades 3–5 fraction TASK.

**Table 5. Concept Knowledge: Sample Responses, Rubric Scores, and Rationale from Grades 3–5 Fractions TASK**

<b>Prompt: What does a student at your grade level need to know and/or understand in order to solve this problem?</b>		
Each carton holds 24 oranges. Kate's carton is $\frac{1}{3}$ full. Paul's carton is $\frac{2}{4}$ full. If they put their oranges together, would Kate and Paul fill 1 whole carton? Solve the problem. Show your work.		
Score	Teacher Response	Explanation
4	They need to understand: A fraction is a part of a whole. A whole can be a group of things or one thing. 24 oranges is a whole, which is mentioned in this problem. $\frac{12}{12}$ is a whole. When adding fractions, you don't add the denominator. Either how to find $\frac{1}{3}$ and $\frac{2}{4}$ of 24 or how to make a common denominator.	<i>This learning trajectory response references multiple underlying concepts.</i>
3	They need to know that $\frac{2}{4} = \frac{1}{2}$ . They also need to know the relationship of fourths and thirds...which is bigger.	<i>This conceptual response focuses on underlying concepts and understanding.</i>
2C	Understand fractions, part of a whole, and know how to add.	<i>This response mentions the part/whole concept</i>

2P	They have to reduce fractions and be able to find common denominators and then add fractions.	<i>but only in a general way. This response is <b>procedural</b> because it focuses only on a specific procedure.</i>
1	Fractions. Reduction. Multiplication. Number order. Ability to count to 20. Addition. English comprehension.	<i>This <b>general</b> response uses broad terms and references only sub-skills.</i>

*Analysis of Student Thinking—Mathematical Validity*

After being shown the set of typical student responses to the problem, teachers were asked to determine whether each student’s solution process was mathematically sound by clicking “yes” or “no”. As Ball et al. (2008) state, determining whether a student’s solution will work in general is mathematical knowledge that is unique to teaching, and falls within the subset of mathematical knowledge of teaching that they call *specialized content knowledge*. It differs from content knowledge in that one would not expect a person trained in mathematics, but who is not a teacher, to necessarily have that expertise.

For each TASK, three of the solution processes were mathematically correct and two were incorrect. In all TASKs, one solution was somewhat ambiguous in that the student attempted to use an appropriate strategy, but either made a conceptual or computational error in the process. To score this question, we looked at the number of correctly identified responses out of the five that were unambiguous. This created a variable that could range from 0 to 1 and interpreted as percent correct. (If someone correctly identifies all five, they have a score of 1.0 or 100%).

*Analysis of Student Thinking—Conceptual Understanding*

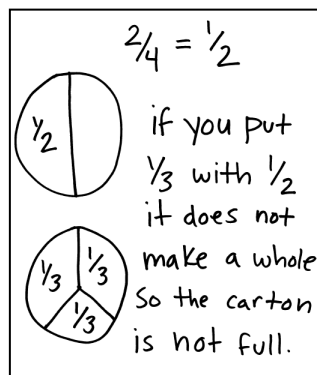
In scoring responses to the four prompts that asked teachers to comment on specific students’ solution processes in terms of the students’ understanding of number and operations (or algebraic or geometric reasoning), we used the same process that was used to score the domain of Concept Knowledge. Table 6 shows the rubric with specific descriptors that relate to the analysis of understanding shown in student the work and Figure 5 shows illustrative examples of each level of the rubric.

**Table 6. Rubric for Analysis of Student Thinking—Conceptual Understanding**

<b>Prompt:</b> <i>Please comment on each student's solution process in terms of what the work suggests about the student's understanding of numbers and operations.</i>		
4	Learning Trajectory	<b>Informed focus on understanding.</b> Identifies multiple concepts or sub-concepts (more than one at articulated level); analysis of student reasoning. Use of clear and appropriate terminology.
3	Conceptual Focus	<b>Analysis of what student did and understands.</b> Some reference to underlying concepts and/or reasoning at an articulated level and correct identification of student strategy.
2C	General Conceptual Focus	<b>Some reference to concepts</b> at a general level.
2P	Procedural Focus	<b>Description of what student did without specific reference to underlying concepts.</b> Describes student strategy or what is wrong about strategy or what student should have done.

1	General/Superficial	<b>No focus on understanding.</b> Focus is at general level, correct/incorrect answer only, or there is a failure to interpret student strategy. “Numbers and operations” falls in this category since it is referenced in the question stem.
---	---------------------	---

**Figure 5. Analysis of Student Thinking: Sample Responses, Rubric Scores, and Rationale from Grades 3–5 Fraction TASK**



<b>Prompt:</b> Please comment on Abby's solution process in terms of what the work suggests about her understanding of numbers and operations.		
Score	Teacher Response	Explanation
4	Abby understands that the size of fractions is determined by the denominator, and they represent breaking the whole into equal parts. She also understands equivalent fractions. Thereby, she is able to compare the two fractions, and, ultimately, compare her results to one whole.	<i>This <b>learning trajectory</b> response references multiple underlying concepts in the student work.</i>
3	She shows that she understands the concepts of fractional part of a whole.	<i>This <b>conceptual</b> response focuses on underlying concepts and understanding.</i>
2C	Abby understands how fractions make a whole part.	<i>This response references conceptual understanding, but it is not articulated.</i>
2P	She drew 2 pies and was able to figure out that the 2 different fractions didn't equal a whole together.	<i>This response is <b>procedural</b> because it describes what the student did.</i>
1	Abby has a basic understanding of fractions.	<i>This <b>general</b> response does not provide any specific evidence.</i>

To increase rater reliability in the domains of Concept Knowledge and Analysis of Student Thinking—Conceptual Understanding, we developed a coding scheme to guide raters in their categorization of responses in relation to different dimensions of the particular mathematical concept. Using the pilot

data, we developed a checklist of frequently mentioned components organized under each of the four rubric categories. A team of three to four researchers then used the coding scheme and rubrics to code 40 pilot responses in each content area. Codes and rubric scores were compared in order to clarify and/or modify the codes and rubric descriptions. Raters then used this list of components as a guide when making the rubric judgments. These checklists provided details to guide the raters and partially automated the scoring process to distinguish level 2P from level 2C. They also provided us with qualitative or descriptive data about what teachers were paying attention to when examining the problem and the student work samples. Appendix A provides an example of the coding scheme for the grades 3–5 fractions TASK. Specific coding schemes were developed for each of the other content areas in a similar format. The “other” category in the coding scheme was used only for the analysis of student thinking cluster, where the rater could indicate if the teacher misunderstood the mathematics behind the strategy or the teacher’s explanation contained a mathematical error or misconception. We also noted instances where teachers were focusing on student explanation or communication skills or referenced the process of scoring student work. These “other” codes were not used to determine a rubric score, but noted for possible later analysis.

The coding scheme is organized in the form of an online checklist, with descriptors under three main categories: general/superficial, procedural, and conceptual. Raters are able to view the original question and teacher response on the top half of the screen, while the codes are presented in the lower half next to check boxes. After the raters have assigned the relevant codes, they are presented with a summary of selected codes and then proceed to determine a rubric score.

The coding scheme shows how the conceptual category is further divided into general and articulated responses. If the reference to conceptual understanding remains at the general level (e.g., the teacher mentions part/whole understanding with no further explication or support to allow the rater to know how the teacher determined this), the response is given the general conceptual code CP. If the response articulates the student’s conceptual understanding (e.g., the teacher explains how the student’s work shows understanding of fractional quantities as a relationship between part and whole), it is given the articulated conceptual code CP1. Raters are trained that the response must contain at least one articulated conceptual code in order to receive a rubric score of 3 (conceptual focus). When raters assign a score of 2 to the response, we are able to use the codes selected to automatically determine whether it should be given a score of 2P or 2C by looking at the codes the raters have checked off. The coding scheme supports the rubric; as raters check off general, procedural, or conceptual elements in the response, they are documenting evidence to support the rubric score. Coding produces a tabulation of the specific concepts and procedures that are referenced by teachers. In this way, coding is both holistic and analytic, and we have a robust base of information on which to score rubrics. We also envision these checklists to be useful for future analysis to decompose patterns of teacher responses within each of the rubric categories.

Teachers were also asked to rank the four or six (depending on the TASK) responses in order of the level of sophistication. As described earlier, the student work was deliberately constructed to represent different levels of sophistication of student thinking as well as common misconceptions or conceptual weaknesses. While there was no one correct way to rank the responses, for each TASK, the student solutions fell into three categories: whether the response contained evidence of solid numerical, fractional, proportional, or algebraic reasoning; evidence of transitional thinking in relation to numerical, fractional, proportional, or algebraic reasoning; or no evidence of numerical, fractional, proportional, or algebraic reasoning. If the teacher correctly ordered the student responses in relation to these categories (e.g., students falling into the solid category were ranked at the top, students falling into the transitional category were ranked in the middle, and students falling in the lowest or "no evidence" category were ranked at the bottom), then the response was considered to be ordered correctly. Additionally, within some of these levels, finer distinctions could be made in terms of sophistication of reasoning. We were therefore able to devise the following rubric and an automated procedure for assigning a rubric score to the different rankings that were present in the teacher responses (see Table 7).

**Table 7. Automated Rubric for Learning Trajectory Orientation—Ranking**

Rubric Score	Ranking	Explanation
4	Correct order and most sophisticated thinking identified.	<i>Advanced learning trajectory orientation.</i>
3	Correct order.	<i>Evidence of learning trajectory orientation.</i>
2	Incorrect order but the lowest two responses were in the bottom two-thirds of the ranking.	<i>Able to identify correct and incorrect reasoning.</i>
1	Incorrect order and one of the lowest two responses was ranked in the top two.	<i>No emphasis on student reasoning or prioritizing use of specific method over conceptual or procedural understanding.</i>

The lowest score of 1 was assigned to rankings where one of the lowest-level responses was ranked first or second. Since those responses reflected very superficial understanding or a complete lack of understanding on the part of the student, but often contained a traditional procedure, this could be an indication that either the teacher prioritized the use of a specific algorithm or that the teacher did not have the ability to distinguish correct from incorrect reasoning in student work. In either case, it represented the lowest level of knowledge of learning trajectories in that it was not based on students' thinking.

#### *Learning Trajectory Orientation—Rationale*

As described earlier, the ranking provided by the teacher allowed us to assign an automated Learning Trajectory Orientation score. We also scored the rationales provided by the teacher. The scorer first examined the rationales provided by the teacher for the student solutions they ranked in the top three to determine what the teacher was paying attention to when evaluating successful student work. The scorer then examined the rationales for the student solutions ranked in the bottom three to see how

the teacher was evaluating weaknesses in student work. Using the rubric shown in Table 8, two scores were assigned for each respondent.

**Table 8. Rubric for Learning Trajectory Orientation—Rationale**

<b>Prompt: Please explain why you ranked _____'s solution process number ____ in relation to the responses of the other students.</b>	
4	<p><b>Learning Trajectory</b></p> <p><b>Ranking is based on level of sophistication of thinking or reasoning</b> behind method used in relation to learning trajectory or framework (i.e., the movement from concrete to abstract understanding of number); makes sense of student thinking or errors and/or identifies common misconceptions in student work</p>
3	<p><b>Conceptual</b></p> <p><b>Ranking is based primarily on understanding or not understanding underlying concepts (or students' algebraic reasoning),</b> and those concepts are named or explained. Efficiency is connected to understanding or applicability to a wide range of problems. Reference is made to the mathematics behind student's reasoning.</p>
2	<p><b>Procedural</b></p> <p><b>Ranking is based primarily on using/not using a specific preferred method,</b> communication or clarity, or demonstrating a specific skill or method without reference to specific underlying concepts or understanding. <b>Overall focus is on what student did.</b> Efficiency is referenced in terms of speed or ease without regard to understanding or thinking. Conceptual understanding may be referenced but without mention of specific concepts.</p>
1	<p><b>General</b></p> <p><b>Ranking is based primarily on getting correct/incorrect answer, format of answer, or other external aspects</b> (labeling, neatness). Uses general terms to describe errors in students' strategies, such as "doesn't understand" or "doesn't make sense" or there is a failure to make sense of students' strategies.</p>

*Instructional Decision Making*

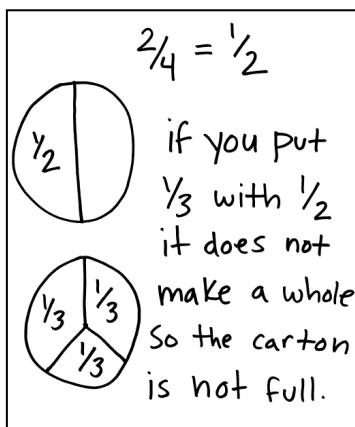
Teachers were asked to provide next steps and explain their rationale for a student who has a correct, but less-sophisticated response to the problem, and a student who demonstrates conceptual weakness in the response. We initially developed a rubric for instructional implications to reflect the four levels that form the core of our measures (general, procedural, conceptual, and developmental/learning trajectory). However, because teacher responses were often not well developed or articulate enough to make finer distinctions, and raters often had difficulty agreeing whether the method described actually built on student understanding, we decided to combine levels 3 and 4, as shown in Table 9.

**Table 9. Rubric for Instructional Decision Making**

<b>Prompt: Examine [Abby or Brad's] work. As a teacher, what would you do next? Please explain your rationale for the steps you suggest.</b>		
3	Conceptual	Help student <b>develop strategy or mathematical meaning</b> , see flaw in thinking/reasoning, and develop conceptual understanding of a particular concept. May build up from <b>current student understanding</b> either toward a more sophisticated or generalizable strategy through an incremental or gradual approach, or to address misconceptions, solidify current strategy, or deepen mathematical understanding.
2	Procedural	Teach student <b>how to use a particular strategy or procedure</b> without mention of conceptual understanding.
1	General/ Superficial	<b>Not directly or superficially related to student work</b> in terms of the mathematics content

The examples shown in Figure 6 illustrate teacher responses at each level of the rubric in response to Abby's work. Abby used a drawing to represent the fractions, and while her reasoning is correct and shows some understanding of part/whole concepts, it is not the most sophisticated method as it relies on the drawing rather than fractional reasoning.

**Figure 6. Instructional Decision Making: Sample Teacher Responses, Rubric Codes, and Rationale from Grades 3–5 Fractions TASK**



**Prompt: Examine Abby's work. As a teacher, what would you do next? Please explain your rationale for the steps you suggest.**

Score	Teacher Response	Explanation
3	I would first ask Abby to look at her representation of $1/3$ and ask her to explain how it is indeed $1/3$ . Abby needs to understand that the circle must be divided into 3 <b>equal</b> parts. Next, I would ask how can she prove it does not make one whole when it is added to one half. I would guide her in seeing that $1/3$ (a whole divided into 3 equal parts) is less than one whole divided into 2 equal parts and, therefore, when added to $1/2$ , it could not equal one whole.	<i>This response is <b>conceptual</b> because it focuses on strengthening Abby's understanding of the concepts of part/whole and magnitude.</i>
2	Abby would be directed into writing fractions, determining a common	<i>This response is</i>



	denominator, and making equivalent fractions and solving the problem.	<i>procedural, focusing on teaching Abby about a specific procedure not based on her understanding.</i>
1	Abby should practice to enrich her understanding.	<i>This is a <b>general</b> instructional response and contains no specifics.</i>

We further clarified this rubric by presenting it to raters in the form of a checklist with specific examples to help guide them in their decision making, as shown in Table 10.

**Table 10. Detailed Rubric for Instructional Decision Making**

<i>To what extent is the instructional response described based upon the student’s thinking and understanding? (Check one.)</i>	
<input type="checkbox"/>	<b>General/Superficial</b> (Not directly or superficially related to student work in terms of the mathematics content) <ul style="list-style-type: none"> <li>• Teach or re-teach the general topic (e.g., proportions, word problems).</li> <li>• Provide enrichment.</li> <li>• Provide remediation or one-on-one tutoring.</li> <li>• Work on explanation or have student explain.</li> <li>• Ask student to do it another way or double check work.</li> <li>• Word problem interpretation (re-read the problem, work on word problems).</li> </ul>
<input type="checkbox"/>	<b>Procedural</b> (Teach a particular strategy or procedure without mention of conceptual understanding) <ul style="list-style-type: none"> <li>• Show student how to use a specific strategy or procedure</li> <li>• Practice specific skills or sub-skills.</li> <li>• Correct a misconception or explain/show why student’s strategy or thinking is wrong.</li> <li>• Help student move away from particular strategy or model (counting fingers, drawing pictures).</li> </ul>
<input type="checkbox"/>	<b>Conceptual</b> (Help student develop strategy or mathematical meaning) <ul style="list-style-type: none"> <li>• Help student see flaw in thinking/reasoning or correct misconception through questioning or explanation.</li> <li>• Ask questions to get more information about student’s conceptual understanding.</li> <li>• Develop conceptual understanding of a particular concept through representations (drawings, manipulatives).</li> <li>• Develop conceptual understanding of a related concept.</li> <li>• Relate current understanding to new understanding.</li> </ul>

## Training TASK Raters

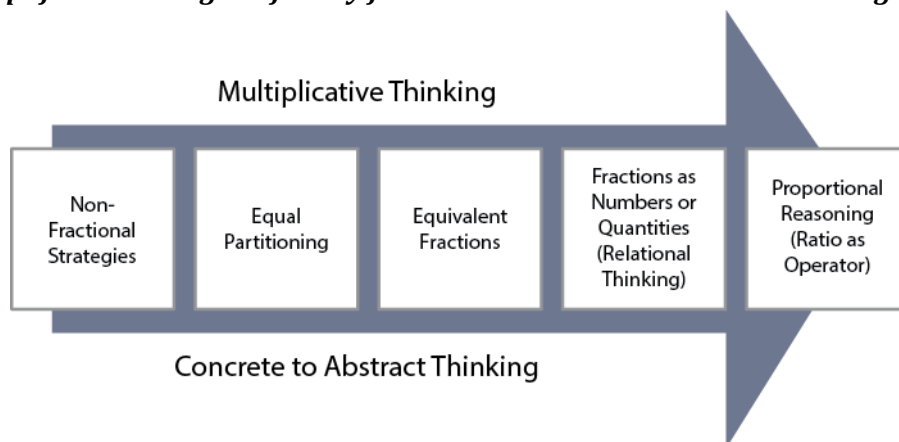
Raters for each of the six TASKs were trained by the TASK developers during Summer 2012. Fifteen mathematics educators were trained to code and score the TASKs in three half-day training sessions tailored to the content of the particular TASK they would be responsible for coding. The raters included experienced mathematics teachers and coaches from school districts in New Jersey, New York, and Pennsylvania and CPRE researchers with mathematics content expertise from the University of Pennsylvania. For each grade band (i.e., K–2, 3–5, 6–8, and 9–11), two outside raters were trained along with a CPRE researcher.

In the first training session, raters were introduced to the TASK and the conceptual work that underlay its development. In the second and third sessions, they received focused training on one specific content domain (i.e., addition, subtraction, fractions, proportions, algebra, and geometry) and were trained to code for references to procedures and concepts particular to that domain in order to characterize the instructional perspective of the respondent. Raters were also introduced to the online scoring system, in which they recorded their observations about particular references to mathematics instruction that were present in teachers' written responses and then assigned a rubric score for each set of items. In the online system, raters went through a batch of 10 responses, one question at a time. The teacher response was presented at the top of the screen followed by the relevant rubrics. For the questions that used a coding scheme, raters completed the checklist and were presented with a summary of their selections before choosing a rubric score.

All raters went through 10 hours of on-site training and additional retraining as needed until reliability was reached, as summarized in Appendix B. The on-site training began with an overview of all grade levels, focused on getting to know the TASK instrument, the problems and solutions in each TASK, and the content-specific coding schemes. Each grade-level group then received an additional five hours of on-site training focused on the specific content of their assigned TASK.

During the subject-specific training (see Appendix B), we presented research findings on student understanding of the content, common student errors or difficulties, and a simplified learning trajectory such as the one shown for fractions in Figure 7. Examples of strategies that fit into the categories of this trajectory were explored so that raters could understand them in terms of the development of multiplicative thinking as well as the move from concrete to abstract thinking. For each question on the TASK instrument, we also reviewed several examples from the pilot data of teacher responses at each level of the rubric, explaining components in the responses that justified the score. We then gave each rater a packet of five teacher responses to each prompt on the TASK, which were intentionally selected to represent a range in levels according to our rubric. Each rater scored the responses independently, and then convened to discuss and reach consensus on the codes and rubric scores for each respondent.

**Figure 7. Simplified Learning Trajectory for Grades 3–5 Fractions TASK Training**



After completing the training, the raters were provided with a set of 20 TASK responses to code via the online system. Approximately 10 days later, they participated in facilitated discussions to examine their agreement and calibrate their scoring with each other. The CPRE trainer facilitated this meeting and provided notes to the raters following the meeting. These notes were also used to revise our training manual for future administration. After this training session, we continued to have the two outside raters double score until they reached over 75% agreement on the rubric scores on all questions. If the agreement was lower than 75%, we provided targeted retraining to recalibrate scoring. Depending on the TASK, reaching 75% agreement took between one and three rounds of retraining.

## **Evidence of TASK Reliability and Validity**

This section presents findings of three reliability studies of the TASK's inter-rater reliability, internal consistency, and test criterion validity. Each study provides a different type of validity evidence. Collectively, the results speak to the construct validity of TASK performance as a meaningful measure of mathematical instructional capability and provide evidence that suggests the scores from the instrument are reliable and valid for the purposes of evaluating teachers' learning trajectory-oriented formative assessment capacity in mathematics. The three studies address the following questions:

*Question #1 (Inter-rater Reliability):* How consistent are the results of an assessment coded by different raters? This question is addressed using evidence based on inter-rater reliability estimates.

*Question #2 (Internal Consistency):* How well does the instrument represent the domain of teacher competencies to be measured? Evidence based on internal reliability statistics was used to address this question.

*Question #3 (Test Criterion Validity):* How well does teacher performance on the assessment predict performance on a criterion measure? Evidence based on the correlations of teachers' TASK scores with concurrent performance on a similar established assessment (the mathematical knowledge for teaching, or MKT) was used to address this question.

The design of the instrument and validation methods are also directly influenced by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999), which provides strong guidelines for high-quality and technically sound assessments. Our methods of instrument validation were chosen to demonstrate evidence of construct validity for consistency, internal structure, and relation to other measures (American Educational Research Association et al., 1999; Cronbach & Meehl, 1955; Kane, 2006). We are mindful that construct validation is a never-ending process wherein new findings may lead to modifications of underlying conceptual models or new interpretation of the measured constructs (Messick, 1980, 1988). Validity speaks to utility and appropriateness of results for a specific purpose in a given context. We are validating the TASK as an observational snapshot of teachers' ability to recognize and respond to students' mathematical understanding.

### ***Inter-Rater Reliability Study***

To assess the consistency of TASK scores across multiple raters, we examined percent agreement based on at least 40 double-coded TASKs per instrument (i.e., addition, subtraction, etc.). After raters demonstrated at least 75% agreement on the last 40 TASKs, they were deemed reliable to individually code and score teacher responses.

Data for these analyses were collected in Spring 2012. Analyses of all five of the TASKs suggest strong evidence for inter-rater reliability. Not only was the raw coding of teacher responses highly consistent, but final rubric scores showed high agreement. Inter-rater reliability on rubric scores, as measured by direct agreement, was 75%. Overall inter-rater agreement was 75%. Agreement on rubric scores across the five TASKs ranged from 69% to 83%, and agreement on items within each TASK ranged from 70% to 85%. Table 11 presents inter-rater statistics, including percent agreement by domain, overall agreement, and Pearson correlation statistic.

***Table 11. TASK Inter-Rater Reliability***

	TASK					
	Addition	Subtraction	Fractions	Proportion	Algebra	Overall
Number Double Coded	40	40	40	40	70	230
Minimum, Maximum	{1,4}	{1,4}	{1,4}	{1,3}	{1,3}	{1,4}
Percent Agreement by Domain <sup>6</sup>						
Concept Knowledge	88	75	85	98	69	83
Analysis of Student Thinking	86	64	66	91	88	79
Learning Trajectory Orientation	71	73	60	76	71	70
Instructional Decision Making	81	73	64	76	50	69
Overall Agreement	82	71	70	85	70	<b>75</b>
Overall Pearson <i>r</i>	.64	.46	.60	.73	.60	<b>.63</b>

<sup>6</sup> See Table 3 for explanation of these domains.

### ***Internal Consistency Study***

To assess the internal consistency of some TASK domains, we used final coded data and estimated the extent to which the rubric scores of Analysis of Student Thinking, Learning Trajectory Orientation, and Instructional Decision Making prompts within theorized dimensions are related to each other through item-total correlations and Cronbach's alpha. Scale reliability analysis was performed overall and for each TASK separately, based on the theorized scale structure. Neither ceiling nor floor effects were observed although variation in scores did vary by subject (i.e., responses on the fractions TASK showed the most variation). Each of the scales was shown to have moderate to high internal consistency with no item-total correlation of less than 0.20 and scale reliability (coefficient alpha) for the domain scales ranged from 0.56 to 0.68. These findings are only preliminary but indicate a moderate level of association between rubric scores within domain, and suggest that alternative scale structures or scoring methods may yield more reliable and meaningful ways of representing coded responses from this instrument.

### ***Test Criterion Validity Study***

An instrument high in criterion-related validity assists test users in decisions of classification, selection, and assessment. To examine the test criterion validity of TASK scores as a measure of formative assessment capacity, we examined concurrent validity based on its statistical association with another similar established measure, the MKT assessment (Hill, Schilling, & Ball, 2004; Schilling, Blunk, & Hill, 2007). The MKT assesses two strands of mathematical knowledge for teaching: Common Content Knowledge and Specialized Content Knowledge (see Figure 3), or the kind of mathematical knowledge that is particular to the teaching profession. Topics include methods for representing mathematical content to students, identifying adequate mathematical explanations, and evaluating unusual solution methods, all of which teachers need to teach mathematics effectively. Items may ask teachers which methods or answers are appropriate for solving a particular mathematics problem or which word problems accurately describe certain equations. The MKT provides a strong criterion related to the content of the TASK in that it is well constructed and validated, and is publicly available. The MKT is most closely aligned with the TASK domains of Content Knowledge and Analysis of Student Thinking—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.

The MKT has multiple equivalent scaled forms available in the following areas (but not limited to): number and operations; proportional reasoning; patterns, functions, and algebra; and geometry. Each MKT form is a multiple-choice assessment of 25 to 30 items that fall under 13 to 18 super-stems. While the assessments are content specific, they are not grade-level specific. For example, the *Elementary Number Concepts and Operations* assessment can be administered to teachers in kindergarten through sixth grade. Both the MKT and TASK are targeted to skills that students learn at certain grade levels. Table 12 presents the specific TASK and MKT instrument used for six teacher groups in the pilot study.

**Table 12. Administration of TASK and MKT by Grade Band**

	TASK	MKT
Kindergarten–Grade 1	Addition	Elementary Number Concepts and Operations
Grades 2–3 <sup>a</sup>	Subtraction	Elementary Number Concepts and Operations
Grades 3–5 <sup>a</sup>	Fractions	Elementary Number Concepts and Operations
Grades 6–8	Proportions	Proportional Reasoning (Grades 4–8)
Grades 9–11	Algebra	Middle School Patterns, Functions, and Algebra

Note: <sup>a</sup> Grade 3 teachers were recruited for either instrument based on random sample.

We used paired teacher data across the five districts to correlate the TASK domain scores with the MKT assessment. Teachers were administered the MKT in Summer 2012 after completing the TASK in May 2012. For the test criterion study, we recruited a sample of 486 teachers to complete one of the six TASKs and the associated MKT as shown in Table 13.

**Table 13. MKT Responses by TASK**

	TASK Responses	MKT Responses
Addition (K–2)	247	89
Subtraction (Grades 2–3)	185	74
Fractions (Grades 3–5)	376	134
Proportions (Grades 6–8)	292	108
Algebra (Grades 9–11)	166	81
<i>Total</i>	<i>1,266</i>	<i>486</i>

Appendices B and C present descriptive statistics and correlation matrices for domain scales separately for each TASK. We report the correlations between each of the six domains based on the full analytic sample, while correlations between TASK domains and the MKT are based on the smaller sample of teachers for whom we have paired data. We find that the statistical associations of MKT and TASK domains reflect a low and moderate relationship and note that correlations are largest for TASK domains that had the most variance.

There are several things worth noting in the correlation table in Appendix D. First, almost all of the correlations are in a positive direction, as expected, with no significant negative correlations. Second, in terms of magnitude, all correlations are below 0.5 (with only one exception of 0.56), indicating there are no highly associated domains (Cohen, 1988). The direction and magnitude of the statistics across TASK instruments suggests that the domains we are measuring are related, but also distinct, and that these constructs are related but distinct from MKT. Across the subject areas, we observe that the largest correlations are between the domains of Analysis of Student Thinking—Conceptual Understanding and Learning Trajectory Orientation (both Ranking and Rationale). In addition, Analysis of Student Thinking—Conceptual Understanding 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. However, it

should also be noted that there is significant variation in the patterns between domains across TASKs, a phenomenon we hope to explore through further research.

## References

- American Educational Research Association, American Psychological Association, and 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.
- 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.
- Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education, 28*, 258–277.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers.
- Confrey, J. (2008, July). *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.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438–481.
- Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- 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.
- Empson, S. (2011). On the idea of learning trajectories: Promises and pitfalls. *The Mathematics Enthusiast, 8*, 571–596.



- Goldsmith, L. T., & Seago, N. (2007, July). *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, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Hiebert, J. (Ed.) (1986). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, *105*, 11–30.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). New York: American Council on Education and Praeger Publishers.
- 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. (1986). 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 and London: Routledge.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*, 1012–1027.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequence of measurement. In H. Wainer & H. I. Vraun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Lawrence Erlbaum Associates.
- National Council of Teachers of Mathematics. (1988). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: The National Academies Press.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*, 155–175.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4–13.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.
- 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.

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

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

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.

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

## Appendix A. Coding Scheme Checklist for TASK Grades 3–5

### Fractions

General/Superficial			
<input type="checkbox"/>	G1	Fractions	
<input type="checkbox"/>	G2	General (e.g., "Problem Solving," "Reasoning," or "Number Sense")	
<input type="checkbox"/>	G3	Sub-skills: Division, multiplication, addition, computation, or other sub-skills	
Procedural (Focus is on how the problem is solved or what the student does)			
<input type="checkbox"/>	P1	Using a drawing or visual representation	
<input type="checkbox"/>	P2	Finding fraction of a quantity numerically	
<input type="checkbox"/>	P3	Finding or making equivalent fractions	
<input type="checkbox"/>	P4	Finding or using common denominators or least common denominators to add or compare	
<input type="checkbox"/>	P5	Simplifying fractions	
<input type="checkbox"/>	P6	Adding $\frac{1}{2}$ and $\frac{1}{3}$ (without common denominators) and comparing to 1	
<input type="checkbox"/>	E1	Counting or calculation error	
Conceptual			
<input type="checkbox"/>	<b>CD Partitioning/Division</b>		
<input type="checkbox"/>	<b>CP Part/Whole</b>	<input type="checkbox"/>	<b>CE Equivalence</b>
<input type="checkbox"/>	<b>CM Magnitude (Comparison/Value)</b>	<input type="checkbox"/>	<b>CO Operation/Addition</b>
<input type="checkbox"/>	CP1 Understanding of fractions as a relative quantity, relationship of part to whole	<input type="checkbox"/>	CE1 Knowledge and ability to make use of basic equivalencies
<input type="checkbox"/>	CP2 Partitioning—understanding of the need for equal-sized parts	<input type="checkbox"/>	CE2 Understanding of equivalent fractions
<input type="checkbox"/>	CP3 Understanding of the meaning of numerator and denominator	<input type="checkbox"/>	CM1 Size of fraction pieces determined by denominator
<input type="checkbox"/>	CP4 Understanding the meaning of the whole	<input type="checkbox"/>	CM2 If the numerator is less than the denominator, less than one whole
<input type="checkbox"/>	CP5 Use of a model (area or set) to represent part/whole	<input type="checkbox"/>	CM3 Ability to compare fraction (or total) to one whole
		<input type="checkbox"/>	CM4 Comparison of benchmark fractions ( $\frac{1}{2}$ and $\frac{1}{3}$ )
<input type="checkbox"/>	CO1 In order to add fractions, you need a common denominator or reference		
"Other" Codes (For Q2 Only)			
<input type="checkbox"/>	O1	Focus on student explanation or communication skills	
<input type="checkbox"/>	O2	Reference to scoring or grading student work	
<input type="checkbox"/>	O3	Teacher misunderstands student method or mathematics behind strategy—evidence of a mathematical misconception	

## Appendix B. TASK Rater Training Procedure

### I. Initial On-Site Training

#### a. Common Training for All TASK Subjects (2 1/2 hours)

- Introduction to the TASK project.
- Overview of the general rubric: Conceptual vs. procedural understanding, balance of conceptual/procedural understanding in Common Core State Standards, levels of rubric.
- Getting to know the instrument: In grade-level groups, participants solve problem and examine ways students have solved the problem.
- Getting to know the coding scheme: Participants examine coding scheme to locate their observations and add further observations about student work.
- Introduction to online coding interface.

#### b. Grade-Level Training, Part I (2 1/2 hours)

- Presentation on research findings on student understanding of concept, learning trajectory, and common misconceptions.
- Review rubric and coding scheme for Concept Knowledge cluster. Review examples of responses at each level.
- Participants code training set (five responses).
- Discussion and calibration of coding.
- Repeat above for Analysis of Student Thinking cluster (four prompts).

#### c. Grade-Level Training, Part II (2 1/2 hours)

- Review Learning Trajectory Orientation rationale rubric and examples of responses at each level. Review learning trajectory.
- Participants code training set (five responses).
- Discussion and calibration of coding.
- Repeat above process for Informed Instructional Decision Making cluster (two prompts).

### II. Off-Site: All raters triple code a set of 20 responses online

### III. On-Site Retraining Session (2 1/2 hours)

- Raters examine triple-coded data for each question and discuss disagreements in codes and rubric scores to reach agreement. Facilitated by CPRE researcher.
- Notes are written and distributed to all raters following the meeting.

### IV. Off-Site: Raters double code a set of 40 responses online

### V. Follow-Up Retraining Sessions (1 hour): Scheduled as needed

- Raters discuss disagreements on questions where inter-rater reliability is less than 70%. Facilitated by CPRE researcher.
- Notes are written and distributed to all raters following the meeting.

*Note: Parts IV and V are repeated until overall reliability is over 75% at which point raters are considered reliable and single code the remainder of the data.*

## Appendix C. TASK Domain Statistics by Subject

		N	Mean	Standard Deviation	Minimum	Maximum
<b>Addition</b>						
	Concept Knowledge	246	2.38	0.80	1.00	5.00
	Analysis of Student Thinking—Mathematical Validity	247	0.89	0.15	0.00	1.00
	Analysis of Student Thinking—Conceptual Understanding	247	2.23	0.54	1.00	4.25
	Learning Trajectory Orientation—Ranking	244	2.70	0.97	1.00	4.00
	Learning Trajectory Orientation—Rationale	247	2.10	0.32	1.00	4.00
	Instructional Decision Making	238	1.82	0.41	1.00	3.00
	Mathematical Knowledge for Teaching	89	-0.34	0.84	-1.97	1.83
<b>Subtraction</b>						
	Concept Knowledge	185	2.45	0.82	1.00	5.00
	Analysis of Student Thinking—Mathematical Validity	185	0.92	0.14	0.50	1.00
	Analysis of Student Thinking—Conceptual Understanding	185	2.47	0.73	1.00	4.00
	Learning Trajectory Orientation—Ranking	180	2.78	0.95	1.00	4.00
	Learning Trajectory Orientation—Rationale	185	2.16	0.56	1.00	4.00
	Instructional Decision Making	185	1.82	0.55	1.00	3.00
	Mathematical Knowledge for Teaching	74	0.00	0.64	-1.12	2.05
<b>Fractions</b>						
	Concept Knowledge	376	2.59	1.13	1.00	5.00
	Analysis of Student Thinking—Mathematical Validity	376	0.84	0.22	0.00	1.00
	Analysis of Student Thinking—Conceptual Understanding	376	2.67	0.83	1.00	4.75
	Learning Trajectory Orientation—Ranking	375	2.27	0.84	1.00	4.00
	Learning Trajectory Orientation—Rationale	376	1.93	0.50	1.00	4.00
	Instructional Decision Making	369	1.76	0.55	1.00	3.00
	Mathematical Knowledge for Teaching	134	0.28	0.73	-1.32	1.83
<b>Proportion</b>						
	Concept Knowledge	291	1.80	0.76	1.00	5.00
	Analysis of Student Thinking—Mathematical Validity	292	0.87	0.21	0.00	1.00
	Analysis of Student Thinking—Conceptual Understanding	292	1.85	0.52	1.00	3.25
	Learning Trajectory Orientation—Ranking	291	1.96	0.77	1.00	4.00
	Learning Trajectory Orientation—Rationale	290	1.80	0.45	1.00	3.50
	Instructional Decision Making	283	1.69	0.55	1.00	3.00
	Mathematical Knowledge for Teaching	108	0.34	0.79	-1.88	2.65
<b>Algebra</b>						
	Concept Knowledge	163	2.40	1.15	1.00	5.00
	Analysis of Student Thinking—Mathematical Validity	166	0.80	0.37	0.00	1.00
	Analysis of Student Thinking—Conceptual Understanding	166	2.10	0.54	1.00	3.75
	Learning Trajectory Orientation—Ranking	166	3.22	1.04	1.00	4.00
	Learning Trajectory Orientation—Rationale	166	2.00	0.62	1.00	4.00
	Instructional Decision Making	164	1.94	0.59	1.00	3.00
	Mathematical Knowledge for Teaching	81	0.53	0.83	-1.60	2.19

## Appendix D. Correlation Matrices

	Task Domains					MKT	
Concept Knowledge (CK)	Analysis of Student Thinking—Mathematical Validity (AST-MV)	Analysis of Student Thinking—Conceptual Understanding (AST-CU)	Learning Trajectory Orientation Ranking (LTO-R1)	Learning Trajectory Orientation Rationale (LTO-R2)	Instructional Decision Making (IDM)		
<b>Addition</b>							
CK	1.00	.	.	.	.	.	
AST-MV	0.05	1.00	.	.	.	.	
AST-CU	<i>0.19</i>	<i>0.18</i>	1.00	.	.	.	
LTO-R1	0.05	<i>0.13</i>	<i>0.22</i>	1.00	.	.	
LTO-R2	0.07	0.03	<i>0.33</i>	0.08	1.00	.	
IDM	0.06	0.06	<i>0.16</i>	<i>0.16</i>	<i>0.13</i>	1.00	
MKT	0.00	0.14	<i>0.33</i>	<i>0.03</i>	<i>0.23</i>	<i>0.24</i>	1.00
<b>Subtraction</b>							
CK	1.00	.	.	.	.	.	
AST-MV	0.07	1.00	.	.	.	.	
AST-CU	<i>0.33</i>	<i>0.24</i>	1.00	.	.	.	
LTO-R1	0.04	<i>0.16</i>	<i>0.20</i>	1.00	.	.	
LTO-R2	<i>0.25</i>	<i>0.16</i>	<i>0.43</i>	<i>0.22</i>	1.00	.	
IDM	<i>0.15</i>	<i>0.23</i>	<i>0.37</i>	<i>0.15</i>	<i>0.43</i>	1.00	
MKT	0.03	0.07	0.22	0.18	<i>0.26</i>	<i>0.24</i>	1.00
<b>Fractions</b>							
CK	1.00	.	.	.	.	.	
AST-MV	<i>0.17</i>	1.00	.	.	.	.	
AST-CU	<i>0.32</i>	<i>0.41</i>	1.00	.	.	.	
LTO-R1	<i>0.16</i>	<i>0.34</i>	<i>0.33</i>	1.00	.	.	
LTO-R2	<i>0.22</i>	<i>0.29</i>	<i>0.48</i>	<i>0.33</i>	1.00	.	
IDM	<i>0.14</i>	<i>0.21</i>	<i>0.39</i>	<i>0.12</i>	<i>0.29</i>	1.00	
MKT	<i>0.28</i>	<i>0.32</i>	<i>0.37</i>	<i>0.40</i>	<i>0.23</i>	<i>0.17</i>	1.00
<b>Proportion</b>							
CK	1.00	.	.	.	.	.	
AST-MV	0.11	1.00	.	.	.	.	
AST-CU	<i>0.40</i>	<i>0.25</i>	1.00	.	.	.	
LTO-R1	0.07	<i>0.24</i>	0.05	1.00	.	.	
LTO-R2	<i>0.28</i>	<i>0.22</i>	<i>0.47</i>	0.05	1.00	.	
IDM	<i>0.12</i>	0.10	<i>0.28</i>	-0.04	<i>0.14</i>	1.00	
MKT	<i>0.24</i>	<i>0.26</i>	<i>0.35</i>	-0.04	<i>0.27</i>	<i>0.33</i>	1.00
<b>Algebra</b>							
CK	1.00	.	.	.	.	.	
AST-MV	<i>0.24</i>	1.00	.	.	.	.	
AST-CU	<i>0.35</i>	<i>0.33</i>	1.00	.	.	.	
LTO-R1	0.10	0.05	0.08	1.00	.	.	
LTO-R2	<i>0.43</i>	<i>0.23</i>	<i>0.56</i>	0.03	1.00	.	
IDM	<i>0.31</i>	0.08	<i>0.42</i>	0.06	<i>0.34</i>	1.00	
MKT	<i>0.30</i>	0.21	0.20	0.01	0.20	<i>0.28</i>	1.00

Note: Italicized correlations  $p < .05$