Measuring the Power of Learning.™

Research Memorandum
ETS RM–16-06

Eliciting Student Thinking: Definition, Research Support, and Measurement of the *ETS*® National Observational Teaching Examination (NOTE) Assessment Series

Yi Qi

Gary Sykes

August 2016

ETS Research Memorandum Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Eliciting Student Thinking:
Definition, Research Support, and Measurement of the
***ETS*® National Observational Teaching Examination (NOTE) Assessment Series**

Yi Qi and Gary Sykes
Educational Testing Service, Princeton, New Jersey

August 2016

Corresponding author: Yi Qi, E-mail: yqi@ets.org

Suggested citation: Qi, Y., & Sykes, G. (2016). *Eliciting student thinking: Definition, research support, and measurement of the* ETS® *National Observational Teaching Examination (NOTE) assessment series* (Research Memorandum No. RM-16-06). Princeton, NJ: Educational Testing Service.

Action Editor: Heather Buzick

Reviewers: Tenaha O'Reilly and Caroline Wylie

Abstract

This report describes and provides research and scholarly support for a core practice of teaching—eliciting student thinking (EST)—that is the target for a performance assessment contributing 1 component of the *ETS*® National Observational Teaching Examination (NOTE) assessment series. The purpose of this report is to review the evidence supporting the importance of this practice for teaching, including how it has been conceptualized, studied, and represented in consensus literature on teaching. Then, the report describes the EST performance assessment together with the rationale for its design.

Key words: high-leverage teaching practice, teaching performance assessment, measurement of teaching

Acknowledgments

The authors thank Andrew Croft for his contribution to this report. Some of the content in this report is used in the following companion reports: *Modeling and Explaining Content: Definition, Research Support, and Measurement of the* ETS® *National Observational Teaching Examination (NOTE) Assessment Series* (RM-16-07) by Leslie Stickler and Gary Sykes and *Leading a Classroom Discussion: Definition, Supporting Evidence, and Measurement of the* ETS® *National Observational Teaching Examination (NOTE) Assessment Series* (RM-16-09) by Margaret Witherspoon, Gary Sykes, and Courtney Bell.

Table of Contents

This paper describes and provides support for an assessment of a critical practice of teaching—eliciting student thinking (EST). While this practice may be modeled as a separate construct, decomposed into constituent elements or aspects, and studied for its association with student learning, it typically operates in tandem with other teaching practices in the ongoing flow of instruction, conceived as the interaction of teachers and students around academic content. The practice of elicitation is featured prominently in the seminal report on human learning issued by the National Research Council (Bransford, Brown, & Cocking, 2000). This report portrays children as actively engaged in making sense of their worlds. In particular domains, such as physical causality, number, and language, children bring a range of predispositions and entering conceptions that may both enable and impede learning. Based on this broad theoretical orientation to learning, Bransford et al. (2000) argued that teaching must be "learner-centered," pointing explicitly to how "teachers attempt to get a sense of what students know and can do as well as their interests and passions . . ." (p. 136). This process involves both the ways in which teachers probe for student understanding and the ways in which they interpret what they see, hear, and observe from their students.

Although this construct is evidently associated with the use of questioning as a means of EST, teachers also design and pose tasks for purposes of eliciting thinking, observe students engaged in academic work in order to understand their thinking, and examine student work for insights into how students are thinking. The construct then is fundamental to the range of ways that teachers seek to understand their students for the purpose of instructing them.

A Performance Assessment

This paper provides backing for the EST construct and its measurement in a performance assessment for licensure, the *ETS*® National Observational Teaching Examination (NOTE) assessment series, developed by Educational Testing Service (ETS). Before describing EST in greater detail, then, we provide a brief preview of the assessment ETS is developing, for the reader to keep in mind. This performance assessment is a simulation during which a candidate is first provided with information about a single simulated student, a content-based task, and the goals for an interchange. Following a 10-minute period to review this information, the candidate engages with the "student" represented by an avatar in a computer-based environment. The avatar is controlled by a trained and certified simulation specialist who operates from a script together with general guidelines for responses to the candidate. In this 10-minute interaction, the candidate demonstrates ability to elicit the simulated student's thinking about the assigned academic content. The simulation specialist can see and hear the candidate. In the final part of the assessment, candidates are given 10 minutes to answer task-specific questions about their understanding of the simulated student's knowledge, based on evidence from the interaction. The simulation specialist's actions and words are standardized to be similar across candidates, and the specialist's performance—that is the actions and words of the avatar—is rated on a rubric designed to assess the specialist's adherence to the standardization guidelines. Further details of this assessment are provided in the following sections.

The NOTE Assessment Series

This performance assessment is one part of a new licensure examination, known by its acronym NOTE, that includes a combination of performance assessments and assessments of the common and specialized knowledge used in teaching. The assessment is being developed by ETS; TeachingWorks at the University of Michigan; and Mursion, a firm that is pioneering the uses of avatar-based interactive simulations for training, preparation, and assessment.

The NOTE performance assessments[1] are oriented around the high-leverage teaching practices identified by teams of scholars and teachers convened by TeachingWorks:

> High-leverage practices are the basic fundamentals of teaching. These practices are used constantly and are critical to helping students learn important content. The high-leverage practices are also central to supporting students' social and emotional development. These high-leverage practices are used across subject areas, grade levels, and contexts. They are "high-leverage" not only because they matter to student learning but because they are basic for advancing skill in teaching. (2016b, para. 2)

High-leverage practices are consensus representations across many efforts to define good teaching.

The NOTE assessments concentrate on what is termed high-leverage content, defined as the particular topics, practices, and texts that have been proposed by TeachingWorks as foundational to the K–12 curriculum and vital for beginning teachers to be able to teach (TeachingWorks, 2016a). Such content, organized by subject area and grade level, is anchored in national and state standards for student learning that have been developed with input from key professional groups (see for example the Common Core State Standards [CCSS] Initiative, 2015a, 2015b). NOTE assessments focus on practices of teaching content based on samples of tasks and items from the relevant content domain. The initial focus of the NOTE assessments is on mathematics and English language arts (ELA) teaching at the elementary level.

EST and the License to Teach

Assessment of EST is part of a licensure examination for teaching. The purpose of licensure is to assure the public that individuals who practice an occupation have met certain standards (Clauser, Margolis, & Case, 2006; Raymond & Leucht, 2013). The focus is on standards of competence needed for effective performance (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 175). “Licensure requirements are imposed by federal, state, and local governments to ensure that those who are licensed possess knowledge and skills in sufficient degree to perform important occupational activities safely and effectively” (AERA, AMA, & NCME, 2014, p. 174). Licensure examinations cover what is necessary but insufficient for practice, meaning that not all of the competencies are assessed, but those that are assessed are critical to effectiveness on the job. The content of licensure examinations typically is derived from job analyses that may be conducted in a variety of ways, usually involving current practitioners who judge how critical competencies are to effective practice.

Validation of tests, including those for licensure, rely on what has been termed, following Toulmin (2003), an argument-based approach (M. Kane, 2004; Papageorgiou & Tannenbaum, 2016). In this approach, the claims for a licensure test are based on data or information provided by warrants, defined as the justification for intended inferences from the data to the claims. Warrants, according to M. Kane (2004), are generally not self-evident and so must be justified: “The evidence supporting the warrant is referred to as the *backing* for the warrant,” (p. 149) as may be derived from theory or empirical research. Validating constructs for licensure is necessarily an ongoing enterprise because validity is a process through which evidence is accumulated and evaluated, not an end state or property of a measure or test (M. T. Kane, 2006).

An important question for licensure concerns how to establish the standard for entry to an occupation. Here, the warrant for a scoring rule “relies on an analysis of the likely consequences (positive and negative) of using the rule. The warrant for the scoring rule may be based mainly or exclusively on expert judgment” (M. Kane, 2004, p.149); a variety of methods for standard setting have been established (see Tannenbaum & Katz, 2013). ETS is conducting standard-setting studies to determine what is required for entry-level practice for EST, which will not be taken up in this report. While researchers have surveyed practitioners on the importance of EST and other critical practices assessed in the NOTE assessments, the purpose of this report is to marshal the research and scholarly literature that provides backing for the EST construct and for its measurement.

The paper unfolds in the following manner. First, the construct of interest is defined together with the rationale for its importance and explication of key aspects of the construct definition. Next, the paper reviews how the construct has been measured in prior studies. Then, the paper describes general studies that have explored this construct, first with respect to student questioning, then with attention to studies of the construct in broad contexts of use. The next section of the paper reviews studies that have explored critical dimensions of the construct, leading to a section that describes the measurement approach adopted by ETS for the NOTE assessments, followed by brief concluding observations.

Construct Definition and Rationale

EST is proposed as one of the "tasks and activities that are essential for skillful beginning teachers to understand, take responsibility for, and be prepared to carry out in order to enact their core instructional responsibilities" (Ball & Forzani, 2009, p. 504). TeachingWorks describes this high-leverage teaching practice as follows:

> Teachers pose questions or tasks that provoke or allow students to share their thinking about specific academic content in order to evaluate student understanding, guide instructional decisions, and elicit ideas that will benefit other students. To do this effectively, a teacher draws out a student's thinking through carefully chosen questions and tasks, attends to what the student says and does, and considers and checks alternative interpretations of the student's ideas and methods (2016b, 3. Eliciting and Interpreting Individual Students' Thinking).

Teachers' knowledge and use of student thinking in their instruction has long been of interest to researchers and policy makers. The construct proceeds from a conception of how people learn, particularly the cardinal principle that "students come to the classroom with preconceptions about how the world works. If their initial understanding is not engaged, they may fail to grasp the new concepts and information that are taught, or they may learn them for purposes of a test but revert to their preconceptions outside the classroom" (Bransford et al., 2000, pp. 14–15). Teachers[2] then do not simply deliver knowledge absent attention to students' entering thoughts and ideas but rather interact with students to probe for their prior and current understanding as a basis for guiding subsequent instruction. In this sense, EST relies on basic principles of constructivist learning insofar as the practice rests on the claim that student prior knowledge exerts strong influence in shaping their continued learning. This claim enjoys strong empirical support in the research on human learning (Hewson & Hewson, 1983; Pazzani, 1991; Pressley et al., 1992; Sidney & Alibali, 2015; Spires & Donley, 1998).

Effective teaching involves making student thinking explicit and integrating students' preconceptions, existing knowledge, and understanding of the content being taught in instruction (Bransford et al., 2000; Fuson, Kalchman, & Bransford, 2005; Kilpatrick, Swafford, & Findell, 2001; Sleep & Boerst, 2012). In order to incorporate student thinking into their instruction, teachers need first to elicit such information from students. For example, working with a student on a problem in mathematics or listening to a student read a passage out loud, the teacher probes for the student's understanding of the underlying mathematical ideas or the import of the passage in terms of its main ideas and supporting evidence. EST requires teachers to be skillful in questioning, using productive prompts and employing appropriate student tasks, to gather information about students' understanding of the academic content (Forzani & Ball, 2016).

The idea of grounding teaching in student thinking and reasoning increasingly draws attention from researchers across content areas and disciplines (Jacobs, Lamb, & Philipp, 2010; Levin, Grant, & Hammer, 2012; Monte-Sano, 2011; Sherin, Jacobs, & Philipp, 2011). Researchers who study teacher preparation and professional development argue that the practice of exploring and interpreting students' understanding of subject matter is essential to instructional decision-making, including deciding what clarifying questions to ask, what new tasks to pose, and what instructional moves constitute this practice (Bautista, Brizuela, Glennie, & Caddle, 2014).

Rationale for Construct

EST is represented in both public and professional standards for teaching and learning. The CCSS Initiative (2015a) for ELA and literacy calls for teachers to ask students questions that push them to demonstrate critical thinking and analytical stills. For mathematics, the CCSS

Initiative (2015b) requires that teachers assess students' understanding of mathematics by asking them to explain and justify their reasoning on whether a mathematical statement is correct, probing for the mathematical thinking behind their solutions.

Prominent standards for teaching also highlight this practice. For example, the InTASC standards (Council of Chief State School Officers, 2013) call for teachers to pose questions that "elicit thinking about information and concepts in the content areas as well as learner application of critical thinking skills such as inference making, comparing, and contrasting (8f; 8g; 8q)" (p. 40). Subject-specific approaches also feature elicitation. For example, the National Council of Teachers of Mathematics (2014) includes "elicit and use evidence of student thinking" (para. 8) as one of the eight mathematics teaching practices in its *Principles to Action*.

Operating from the same premise—that student thinking is critical to access and understand—an important line of scholarship has begun to explore the kinds of knowledge teachers draw on in the practice of EST. This work serves as a companion to direct measures of elicitation, supplying practice-based measures of teachers' knowledge. Shulman (1986) launched this work with his concept of *pedagogical content knowledge*. Based on the assumption that students' minds are not "blank slates," Shulman argued that an essential component of such knowledge includes an understanding of "the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons" (p. 9). How though would teachers gain access to this knowledge about student understanding?

EST is a cardinal practice in response to this question. Building on Shulman's (1986) work, Ball, Thames, and Phelps (2008) conceptualized this domain as knowledge of content and students, defined as a combination of knowledge about students and the subject matter. And, in their tasks of teaching framework, D. H. Gitomer, Phelps, Weren, Howell, and Croft (2014) proposed that teachers bring both common and specialized content knowledge to this interactive work with students that can be measured. The approach that ETS took toward the assessment of EST moves beyond knowledge measures to capture this construct as an interactive performance the teacher engages in with a simulated student.

Explication of the Construct

The definition of EST suggests a logical sequence of steps teachers undertake when implementing this practice. Teachers begin by formulating questions and tasks in advance of instruction; then, they employ these questions and tasks to launch instruction. As students respond, teachers *attend*—listen carefully—to what students say and do in their oral and written responses, then follow up with further questions and tasks that probe students' thinking and understanding.

The literature on this construct makes use of a variety of descriptors for elicitation that are intimately bound up with one another in interactions. In practice, elicitation is dynamic, iterative, and sensitive to context in the sense of always involving a "conversation with the situation," where the situation involves the specific students, subject matter, curriculum, and larger context. Teachers ask questions or pose tasks, probe for student responses, listen carefully to what students say, ask further clarifying questions, explore their own interpretations of what students are saying, and continue the process with an eye on developing sufficient understanding to enact the next round of instructional moves. Although this set of moves is useful in decomposing EST, in practice, there is no single, simple sequence for these subpractices, and separating into parts what is a complex, interactive process risks reification and distortion, as if EST were a linear process that may be scripted.

Further underscoring the importance of EST, this practice is an important component or building block for other critical practices of teaching including formative assessment, leading classroom discussions, and building relationships with students. EST in this sense is a foundational skill that undergirds a conception of teaching in which the teacher seeks to make connections with all students through an understanding of their thinking and reasoning about the subject matter of instruction. EST also provides support for equity goals insofar as teachers attend carefully to how students' social, cultural, and linguistic backgrounds influence their reasoning about and understanding of subject matter. This practice emphasizes how teachers listen to students and draw out their thinking, rather than moving quickly to provide answers that may not be responsive to students' current understanding.

As Forzani and Ball (2016) describe:

> Doing this well depends on skillful questioning, using questions that are open and curious, rather than leading and discouraging. These questions require a tone, an intentness, and a manner that communicates to students that their ideas are interesting and worthy of attention. But eliciting and interpreting well, in ways that respect and value students, is more than a set of specific professional skills.

> The practice depends on a stance toward students, a commitment to them as thinkers and knowers, and an appreciation of their constant ability to apprehend the world and to interpret and make sense of it. The practice also depends on a positioning that communicates that students are the experts about their own thinking and that teachers seek to learn from them about what they know and believe. (p. 7)

We review the scholarly and research literature supporting EST in two sections: first, a general overview of the research on this practice, and second, a review of the studies that support the specific component practices that make up the construct. Throughout we describe how EST has been measured across various studies and traditions of inquiry.

Background Studies Related to EST

Two general lines of inquiry bear on interactions between teachers and students that involve elicitation. One concentrates on student questions and how teachers prompt and promote student questioning. Here, the main hypothesis holds that when students are taught to ask productive questions, in settings ranging from one-on-one tutoring to collaborative problem solving groups, they learn more and become more proficient at regulating their own learning. The second line of scholarship concentrates attention on how teachers elicit student thinking, treating this as a core practice of teaching. For purposes of the NOTE assessments, research in this second tradition is most relevant, but we begin with a brief review of elements in the first tradition that have implications for teachers' eliciting practices.

Studies of Student Questioning

One context in which questioning has been studied is individual tutoring. Studies have supported the efficacy of this practice in producing student learning (Bloom, 1984; Cohen, Kulik, & Kulik, 1982), prompting inquiry into the specific aspects of tutoring that contribute to its effectiveness. A leading hypothesis in this line of research concentrates on the questioning behavior of students in the tutoring relationship. Here, we do not review these studies in detail, which have included efforts over many years to create computer simulations as well as live tutoring (for examples in ELA, see McNamara, Ozuru, Best, & O'Reilly, 2007, Section 4; for examples in mathematics, see Christman & Badgett, 2003; Kulik, Kulik, & Bangert-Drowns, 1985; Mevarech & Rich, 1985; Roschelle, Pea, & Hoadley, 2000; Seo & Bryant, 2009). Rather, we point to a number of findings from this work that bear on teachers' elicitation of student thinking.

One prominent study in this line of inquiry revealed that while students asked very few questions in regular classroom settings, they asked considerably more questions in tutoring situations (Graesser & Person, 1994). But it proved not to be the case that high-achieving students asked more questions. Rather, qualitative aspects of questioning and answering looked to be more influential on learning (their sample included college students and 7th graders).

More particularly, this work discovered some differences between good and poor learners. Good learners tend to provide answers to questions that are more complete and specific. In contrast, "poor students were more likely to provide vague answers, error-ridden answers, or no answer at all to tutor-posed questions" (Person, Graesser, Magliano, & Kruez, 1994, p. 223). These researchers continued, noting that "robust correlations between student achievement and student answers suggest that student understanding is best assessed through analyses of students' answers rather than analyses of student questions" (Person et al., 1994, p. 223).

This study also found that students' answers to comprehension-gauging questions were frequently misleading (Person et al., 1994, p. 225). Students who were poor learners often did not have a good grasp on what they did not understand and so might not respond to such invitations as, "Do you understand that . . ." or "Do you have questions about . . ." In fact, Person et al. (1994) found a positive correlation between student achievement and students' negative answers to comprehension-gauging questions, which they interpreted as an indication that good students realize they do not understand more often than poor students. Likewise, the study found that student knowledge-deficit questions are a poor information source for assessing comprehension, again theorizing that poor students cannot calibrate their own understanding and may be reluctant to reveal ignorance in social settings.

From studies of this kind came a number of suggestions for the practice of tutoring. First, asking more long answer and deep reasoning questions supports comprehension and understanding. A long answer question might be, "How is X similar to Y?" or "What does X mean?" A deep reasoning question might inquire about interpretation ("What concept or claim can be inferred from a pattern in the data?") or about causality ("What were the causes or consequences of X?"). Person et al. (1994) found that relatively few such questions were present in the interactions they studied (undergraduate and graduate students) and hypothesized that questions of this character have value in promoting learning. Further, they argued that questions featuring a high degree of specification are likely to be productive. Rather than inquiring generally about understanding, more specific questions formulate what is to be understood with particular content. Finally, this work suggested that students can be trained to ask good questions: “In fact, there is substantial empirical evidence that there are robust improvements in comprehension, learning, and memory of technical material after students are trained how to ask good questions” (Graesser & Person, p. 131). Other studies on student comprehension have also highlighted the use of questions that students pose to themselves and to peers as fostering problem-solving and comprehension capabilities (see for example M. Chi, 1996; M. T. Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Fuchs & Fuchs, 2007; McNamara et al., 2007).

Following this suggestion about training, another line of research explored student questions in the context of collaborative work among pairs of students. Over a series of studies, King (1991, 1992, 1994, 2007; King & Rosenshine, 1993) developed and refined an intervention around the kinds of questions students were trained to ask of one another as they worked on problems or interpreted texts. This work found that as questioning became increasingly elaborated, while encouraging students to tap into and access their prior knowledge, gains in learning improved, relative to control groups that were trained in less elaborated approaches to questioning. In the elaborated condition, learners were taught to ask five different kinds of questions, calling for explanations, inferences, and predictions, sequenced according to partner responses, and comprehension level.

In this program of research, simple “why” and “how” questions progressed in subsequent studies to more elaborate, thought-provoking questions, and then from text-based to experience-based questions that tapped into students’ prior knowledge. Results of these studies reported that students generated significantly more explanations and inferences during peer interactions and on written tests than did students in various control conditions. Although these studies were conducted with fourth and seventh graders on science and social studies content, the investigators surmise that the strategies would be useful for any subject matter and for any grades above fourth grade.

Returning to our focal interest, the EST construct concentrates on how teachers elicit student thinking rather than on how teachers may train students to, in effect, elicit their own thinking and that of their peers. These constructs are not the same, and both may be valuable. But this work on tutoring and on student-paired collaborations contains some pertinent insights. One concerns attention to the types of question that teachers ask, with emphasis on questions that provoke thought, call on more extended responses, and require deep reasoning from students. This point, we hypothesize, applies to teachers' elicitation strategies. Another finding draws attention to how teachers, as with tutors, might press students to provide more complete, thorough, and accurate responses to questions that are well specified in relation to the particular content of instruction. And a third point takes note that teachers are likely to learn more about student understanding from their answers to questions than from the questions students pose, insofar as students have been shown to be untrustworthy guides to their own understanding. "It has been well documented that students have difficulty in identifying contradictions and inconsistencies in orally presented messages, scientific texts, mathematical word problems, and other text forms" (Person et al., 1994, pp. 206–207). This documentation suggests a mediating role for the teacher, which ties directly to the EST construct.

We next organize the research and scholarly literature bearing more directly on EST according to the following logic. High-leverage practices in general, including EST, often operate in tandem with other practices in achieving learning outcomes. Consequently, some of the studies and measures that include EST also include other constructs such that teasing out the particular contribution of EST is not possible. Rather, the claim in such cases is more modest, that EST makes some (unknown) contribution to the efficacy of the practices being measured. But just as EST is often studied as a component among teaching practices, so EST itself may be decomposed into constituent aspects or dimensions, each hypothesized to contribute to the overall effect of EST. Some of the literature on this construct examines how such hypothesized dimensions contribute to the efficacy of the practice, and we review these studies as well.

We divide our review then into these two parts: studies that include EST either focally or as part of more complex sets of practices and studies that examine how particular features or dimensions of EST have been theorized, described, and studied. This second part of the review prefigures how we are measuring EST in the performance assessment: aspects of the scoring rubric reflect dimensions of EST derived from the research literature on the construct.

Prior Research and Measurement of EST in Context

Here, we order studies bearing on EST from the theoretical and general to increasingly focused on the construct. In some cases, investigators sought to associate EST and related practices to outcomes, whereas in other cases they sought to provide more detailed descriptions of EST. A starting point for such review begins with theoretical interest that in turn began to prompt empirical studies of various kinds. Such theory sinks deep roots in cognitive and sociocognitive studies of learning and of the broad turn to more constructivist-oriented conceptions of learning, manifest for example in the National Research Council report (Bransford et al., 2000) referenced previously.

One influential contribution in this vein, drawing on naturalistic studies of learning in out-of-school-contexts, posited the notion of "cognitive apprenticeship" (Brown, Collins, & Duguid, 1988). This idea came to influence a good deal of thinking about teaching and learning, but here the connection drew attention in particular to the practices of deliberately making one's thinking visible, and an essential principal of designing cognitive apprenticeship environments included how teachers encourage students to articulate their knowledge and thinking. In order to make students' thinking visible, teachers must probe for how students are thinking about a mathematics problem or a reading passage. Teachers might model their own thinking interspersed with queries that ask students to present their thinking, to think out loud in response to which teachers might then probe further. The leading metaphor here—apprenticeship—positions the student as apprentice to a master (the teacher), where what is learned is not a physical skill but a cognitive one. The master *scaffolds* apprentice performance, a term indicating activities and tasks such as generating interest, simplifying the task, maintaining pursuit of the learning goal via motivation and direction, marking critical discrepancies between what a child has produced and an ideal solution, controlling frustration, and demonstrating an idealized version of the act to be performed (Bransford et al., 2000, p. 103). Elicitation then plays a role in scaffolding, even as the term points to other kinds of support for learning. Strategies of elicitation then owe theoretical backing to the cognitive apprentice model of teaching and learning.

Although this model invokes the tutoring relationship, its import was generalized to the work that teachers do with students in regular classrooms. As investigators began developing observational protocols that sought to identify important features of instruction that could be inferred by raters watching live or videotaped lessons or lesson segments, attention to elicitation appeared in several of the most prominent instruments. In Danielson's (2013) *The Framework for Teaching*, the component, using questioning and discussion techniques, specifically evaluates teachers' ability to design and use questions to elicit high-level thinking and understanding. Desired practice emphasizes that the "teacher uses open-ended questions, inviting students to think." The framework also includes an element that measures the extent to which the teacher asks students to explain or justify their reasoning and the extent to which the teacher makes adequate use of wait time for students to respond.

Likewise, in the Classroom Assessment Scoring System (Pianta, Karen, Paro, & Hamre, 2008), some dimensions under the instructional support domain attend to teachers' abilities to elicit student thinking. For example, the instructional dialogue dimension measures teachers' skills in carrying out cumulative, content-driven exchanges with students and using facilitation strategies, including asking open-ended questions. Although these practices are beneficial in EST, the construct is defined and measured in a much broader way than in our assessment.

An emphasis on EST also appears in subject-specific observation protocols. The Protocol for Language Arts Teaching Observations (Stanford University, 2013) identifies *intellectual challenge* as one of its 13 dimensions of quality, including "the level of analytical and inferential thinking demanded in the questions asked by the teacher within discussion or classroom activities" (para. 3). And, the Mathematical Quality of Instruction (MQI) instrument (Hill, Blunk, et al., 2008) is designed to capture characteristics of mathematics teaching conjectured to yield conceptual understanding among students. The dimension, working with students and mathematics, measures the extent to which the teacher "elicits and develops students' own mathematical thinking and reasoning, and engages students in cognitively demanding activities such as questioning and reasoning" (Hill et al., 2012, p. 5). In addition to its focus on eliciting students' mathematical thinking, MQI also captures the extent to which students are asked to explain and justify their answers using mathematical concepts. Among the few well-researched observation protocols, only MQI has an element that measures whether teachers can accurately interpret students' mathematical ideas.

This small set of studies helps to provide validity evidence for these observation protocols, but much remains to be understood (e.g., Grossman et al., 2010; Hill et al., 2012; T. Kane, Taylor, Tyler, & Wooten, 2011; La Paro, Pianta, & Stuhlman, 2004). The validation work does not specify the contribution made by particular dimensions of instruction to outcomes. In these instances, the construct of EST tends to be part of larger pedagogical or instructional frameworks suggesting how the construct works with other aspects of instruction to promote learning.

EST took up a more prominent and central place in several programs of research based on theory and findings from cognitive science. As the tutoring research described previously capitalized on, detailed studies of student learning in particular topical areas in mathematics and science offered the prospect that such knowledge might be of direct use to teachers. One hypothesis held that if teachers were provided information on typical student misconceptions with respect to particular topics in mathematics, they could make good use of this information in their instruction. This hypothesis was the premise of the Cognitively Guided Instruction (CGI) program that produced some significant results. In a series of studies spanning 20 years, researchers and educators gathered evidence that explicated the particular instructional moves that teachers use when eliciting and making use of student thinking, which in turn was linked to improved student learning (Carpenter, Fennema, & Franke, 1996; Carpenter, Fennema, Franke, Levi, & Empson, 2000; Fennema et al., 1996; Franke, Kazemi, & Battey, 2007; Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Sfard & Kieran, 2001).

In a study of the CGI professional development program, investigators focused on the strategies that a sample of elementary teachers, trained on the CGI program, used in asking questions to help children make their mathematical thinking explicit. Using videotaped lessons on a prescribed mathematical topic, researchers conducted qualitative coding on teacher questioning and student participation. Results suggested that although these CGI teachers consistently directed students to share their initial explanations and thinking, their practices in following up and eliciting students' elaboration varied. Certain moves teachers made appeared to be more effective in revealing details of students' strategies and thinking. In particular, probing sequences of specific questions were the most successful in eliciting further elaborations (Franke et al., 2009). Researchers found that teachers participating in CGI discovered the need to invite students' mathematical explanations by using questions and strategies to elicit their thinking. Students in these teachers' classrooms were found to have higher achievement gains on a mathematics test than their control group counterparts (Fennema et al., 1996; Peterson, Fennema, & Carpenter, 1991; Villaseñor & Kepner, 1993).

In addition to CGI, another program of research in mathematics, known as QUASAR, also drew on insights from cognitive psychology to design and test an intervention that employed elicitation (Stein, Smith, Henningsen, Edward, & Silver, 2009). These investigators implemented an approach to mathematics instruction that focused on the development of conceptual understanding and thinking for students in economically disadvantaged urban schools. In conjunction with carefully designed instructional tasks that emphasize cognitive complexity, QUASAR teachers asked middle school students to provide explanations and justifications when solving mathematical problems (Silver, Smith, & Nelson, 1995). Through skillful questioning—the eliciting aspect of the intervention—teachers were able to uncover students' understandings and misunderstandings of the mathematical concepts, which then entered into subsequent and ongoing instruction. Results of students' performance on the QUASAR Cognitive Assessment Instrument demonstrated learning gains. QUASAR students' performance on the National Assessment of Educational Progress (NAEP) tasks also indicated that they performed better in comparison with a comparable NAEP sample (Silver & Stein, 1996).

Several other studies continued this effort to probe aspects of EST. In one study, Fraivillig, Murphy, and Fuson (1999) proposed a framework for advancing children's thinking (ACT) that facilitates the development of students' conceptual understanding of mathematics. As part of a larger study on the implementation of the *Everyday Mathematics* curriculum, they analyzed classroom observation and interview data to examine how first grade teachers promoted students' mathematical thinking and then developed a pedagogical framework based on one teacher's instructional practices. Eliciting students' solution methods was among the three major components of the effective teaching practices that were identified. This practice included probing students to describe their solution methods to ascertain what students already knew as well as how they understand mathematical concepts, listening to students' descriptions and explanations, encouraging students to elaborate their explanations, and using challenging follow-up questions. Using the ACT framework with a small sample (n = 5) for cross-teacher analysis, these researchers found that teachers would likely require directed training and support in this practice as their elicitation of students' thinking and solution methods appeared less frequently than expected.

Another piece of evidence bearing on EST was the study by Jacobs et al. (2007). In evaluating the effects of a professional development program focused on students' algebra reasoning, these investigators used a structured interview protocol to assess teachers' knowledge of students' problem-solving strategies as well as their knowledge of the specific mathematical concept being studied. Results uncovered no significant difference in the performance on the mathematical knowledge test, but the participating teachers demonstrated more knowledge of students' strategies than did nonparticipating teachers. (The investigators noted that these results were reported after the first year of the professional development program, suggesting that with more time and training, more robust effects on student achievement were likely.)

The studies reported to this point, moving from the most general (measuring EST among other practices) to studies that move closer to the construct, employ classroom observations, structured interviews, and review of student work samples. Yet another influential strand of research involved descriptive studies that provided fine-grained descriptions of the specific moves teachers make when they are probing for student understanding. Such cases explore not only teachers' efforts to uncover where students may misunderstand key ideas, concepts, and strategies, but also where student thinking may be original and provocative, serving as a springboard for discussions that explore novel solutions as well as moving toward canonical ones. Both Magdalene Lampert (e.g., Lampert, 1986, 1990, 2003) and Deborah Ball (Ball, 1993a, 1993b; Ball & Bass, 2000) pioneered an innovative approach by studying their own practice of elementary mathematics teaching. Their inquiries supplied provocative and widely cited examples of how a teacher explores students' thinking as they work on particular mathematics problems.

For example, in studying her own elementary mathematics teaching, Lampert (1986) revealed the deep purposes of EST. She writes, "In my discussions with students, I almost always followed an unexplained answer with a question to probe how the student 'figured it out.' This strategy has two purposes. One is to give me some sense of the procedures students are using to arrive at their answers and how they are warranted; the other is to develop a habit of discourse in the classroom in which work in mathematics is referred back to the knower to answer questions of reasonability" (p. 317).

Another example helps to uncover the challenges teachers face as they interpret or make sense of students' responses to problems that are posed to them. An exploration along these lines imagined the following:

> Suppose you posed four numbers—7, 38, 63, and 90—to a class and asked the students to identify which of the numbers was even. And suppose, further, that you got this paper back from one of the students, with none of the numbers circled

> . . . What would you make of this? Is this answer surprising or predictable? What might the student actually know? What number or numbers would you pose next to find out with more precision what the student thinks? Why would that selection be useful? (Ball & Bass, 2000, p. 83)

Examples of this kind, drawn from qualitative case studies, clarify the interpretive aspect of EST together with the process of setting tasks for students that aim specifically to uncover their understanding and their reasoning.

Finally, case studies of this kind helped give rise to the prospect that what expert teachers know as they engage in common tasks of teaching might be measured. This line of work sought to develop so-called practice-based knowledge items that moved conventional assessments (e.g., multiple choice questions) close to the practice of teaching. One such program identified a design framework for tasks of teaching (D. H. Gitomer et al., 2014; D. Gitomer & Zisk, in press). One such task involved "developing questions, activities, tasks, and problems to elicit student thinking" (p. 498). Appendix A provides a sample item that demonstrates how such content knowledge for teaching may be measured. The item requires knowledge of ordering fractions, the ability to construct a set of fractions that meet a given set of conditions, and an understanding of the concept of a counter-example. The eliciting aspect involves the teacher in understanding how to continue eliciting the student's understanding through the use of a mathematics problem that addresses the particular confusion exhibited by the student in this case.

The preponderance of these studies have explored EST in the context of mathematics teaching and learning. EST in the context of ELA has been less studied, although practices such as having students read aloud are commonplace and constitute at once an opportunity for students to practice and a form of elicitation in which teachers are gauging student reading abilities (for extensive review, see Snow, Griffin, & Burns, 2005).

Prior Research and Measurement of Dimensions of EST

In this next section of the report, we examine dimensions of EST that have been described and studied in the empirical and scholarly literature. These dimensions directly informed development of the rubric used to score performances on the EST assessment.

In a simple-minded way, teachers set tasks and ask students questions in part to gain access to their thinking, probe their responses with further questions, then undertake to make sense of what students are saying. In a class full of students, this is a complex process, but it may be informed by results from cognitive studies that have uncovered common misconceptions, misunderstandings, and errors in student reasoning and background knowledge (as demonstrated in the CGI studies reported previously). Armed with this knowledge about student thinking, the hypothesis holds: Teachers are better able to elicit student thinking in productive ways.

A further point to note concerns teachers' purposes in EST. A primary purpose may involve guiding students to correct, accurate, and complete understanding, moving from a student's initial responses to improved understanding. But teachers also elicit student thinking to explore promising or interesting student conceptions and ideas. For example, some discursive practices involve helping students to generate novel or creative solutions, to learn how to argue and debate, or how to engage in collaborative work with other students. Rather than funneling down to some correct understanding, elicitation may open up to multiple possible strategies, for example, of solving a mathematical problem or interpreting a text. As teachers engage with students in exploring their ideas, they may simultaneously be opening up the interchange to promote student thinking and move students toward conceptual and procedural understanding of the mathematics or the genre and other elements of a text.

Several clarifications are useful here. First, although questions may be the basic unit for eliciting thinking, teachers also use short prompts that may involve providing some information to help students along. A prompt may be something as simple as "OK, but explain that a bit further," "You are on the right track, but . . . ," or "I like what you said, but what did you mean by . . . ?" Second, teachers also use tasks to elicit thinking. In the midst of a mathematics lesson, a teacher might pose a problem somewhat different from the ones students have been working on, in order to observe their responses and see if their knowledge transfers to a somewhat different problem. Or they may pose hypothetical tasks such as interpreting a scenario in which a character in a story behaves differently to extend student understanding of character and circumstance. We turn next to a description of the primary dimensions of EST together with the theoretical and empirical literature in support.

Primary Dimensions of EST

Aspect 1: Using Questions, Prompts, and Student Tasks to Elicit Student Thinking

Teacher questioning has been studied as an essential element of effective teaching (Dillon, 1988, 1990). Researchers have described frameworks for analyzing teachers' questions, providing techniques for asking effective questions, and illuminating how norms in the classroom are established and instructional goals are achieved by the use of questioning (Goos, 2004; Martino & Maher, 1999; Mewborn & Huberty, 1999; Van Zee & Minstrell, 1997; White, 2003). Harris (2000) described several goals of teacher questioning, including "checking understanding, starting a discussion, inviting curiosity, beginning an inquiry, determining student's prior knowledge, and stimulating critical thinking" (p. 25). Various question typologies have been proposed and studied, often in the context of live and simulated tutoring (see Dillon, 1990; Graesser & Person, 1994).

In studying what questions to ask and how to ask questions in classroom instruction across disciplines, researchers have examined the frequency of questions asked in classrooms, correlated the types of questions teachers ask with the desired student responses, analyzed the cognitive level of teacher questions (a measure of complexity in terms of mental operations), and investigated teachers' other strategies in orchestrating question–answer interactions (Appalachia Education Laboratory, 1994; Cotton, 1989; Gall, 1970, 1984; Ornstein, 1988; Redfield & Rousseau, 1981). For example, researchers of the Questioning and Understanding to Improve Learning and Thinking (QUILT) program studied participating teachers' behavioral change in their use of wait time when questioning students and the extent to which they used more cognitively complex questions to probe student thinking (Barnette, Orletsky, & Sates, 1994). In one early study, Gall (1984) found that although teachers ask a lot of questions in the classroom, only about 20% of them require thinking at higher levels. In another early study, Redfield and Rousseau (1981) conducted a meta-analysis on 20 experimental or quasi-experimental studies to examine the relationship between teachers' instructional use of questions and student achievement. They found a positive overall effect for teachers' use of higher cognitive questions in instruction. Based on considerations of the experimental design and sample size, these investigators suggest that students' achievement gains can be expected when the teachers receive training on questioning skills.

Studies have identified features of effective questioning when teachers check for student understanding and prompt student thinking and reasoning (Araceli Ruiz-Primo & Furtak, 2006; Walsh & Sattes, 2011). One of the fundamental purposes teachers pursue in EST is to ensure that the student understands the questions he or she is being asked (Walsh & Sattes, 2011). This practice calls on teachers' ability to design and frame questions that the student clearly understands (Ellis, 1993). Walsh and Sattes (2011) proposed that in order to formulate quality questions, teachers need to determine the content focus, consider the instructional function of the questioning, stipulate the expected cognitive level required from the student, and match the questions to the appropriate social context, which determines the level of challenge of the questions as well as the potential for supporting and scaffolding student learning.

Another aspect of questioning examines whether a series of questions is logically sequenced and complete. And in response to English language learners and students with special needs, an additional aspect is whether the language that teachers use is academically and developmentally appropriate and well-adapted for such students (Charlesworth, 1998; Kostelnik, Soderman, & Whiren, 2007).

Soter et al. (2008) analyzed features of discourse that are indicative of successful learning across different types of discussions. Two qualities of discourse, critical-analytic and expressive, stood out as the most strongly related to high-level student reasoning. From the perspective of EST and reasoning, we note that these approaches place emphasis on asking authentic questions and incorporating a high degree of uptake or use of student responses in the discussion. This type of question "elicited high level thinking (analysis, generalization, and speculation) together with high incidence of elaborated explanations and/or exploratory talk" (p. 389).

While studies of this kind examine general features of elicitation, other research has examined this practice in the context of particular subject matter, making clear that EST has strong subject-specific aspects. For example, in ELA, most research examining teacher questioning has focused on the teaching of literature. In one meta-analysis that synthesized results of 42 studies, researchers found a positive association between certain discussion approaches and measures of teacher-student talk as well as students' reading comprehension outcomes (Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). Results suggested that student comprehension is enhanced through discussion approaches and questions that are more *efferent* in nature, defined as having a focus on reading for knowledge acquisition and information retrieval. The authors conjecture that increasing opportunities for students to share their thoughts in the context of critical thinking about text during reading instruction is related to increased comprehension.

Several studies have explored decompositions of EST in elementary and secondary mathematics teaching. Boaler and Brodie (2004) conducted qualitative analyses to categorize the questions teachers used in mathematics classrooms across three schools. Of the nine types of questions they codify, one type they labeled as "gathering information, leading students through a method" elicited answers from students while providing opportunities for students to explain facts and procedures. Another type, labeled as "probing, getting students to explain their thinking" pressed students to articulate, elaborate, and clarify their ideas. This work was primarily descriptive in identifying how types of questions serve a range of pedagogical purposes.

In investigating how scaffolds (i.e., questions provided to preservice teachers for use in working with students) support beginning teachers' practice of eliciting and interpreting students' mathematics thinking, Sleep and Boerst (2012) coded teacher questions and prompts, grouping them into four main categories: initial posing, eliciting student thinking, neutral, and leading. Among these categories, initial posing introduced a task, whereas "the eliciting student thinking category aimed at prompting students to elaborate their thinking as they respond to a task" (Sleep and Boerst, 2012, p. 1042). Examples of this category of questions were: "What is 25+10?" and "How high can you count?" (Sleep and Boerst, 2012, p. 1041). This questioning move asked the student to predict the answer; describe his or her method or reasoning; explain the underlying rationale for a response; discuss connections between mathematical ideas, problems, and representations; extend the problem to other conditions where similar ideas apply or generalize; and orient the student on key aspects of the problem to enable problem solving. This investigation was a descriptive study that explored how scaffolds help teaching interns pose fruitful questions, so it did not trace the practice to student learning outcomes.

In sum, then, investigators have proposed various typologies of questions—both general and subject-specific—that teachers use in gaining access to and pressing students to elaborate on their thinking and reasoning. Work of this kind has been productive in directing teachers to features of questioning and prompting that assist them in understanding student thinking as the basis for instructional moves and interventions.

Aspect 2: Attending to Student Talk and Actions

The term *attending* refers to cognitive activity on the part of the teacher as he or she listens carefully to what students say in order to detect their reasoning and their understanding, for example, of content, concepts, and mental processes. During instruction, students speak and act, and attending refers to both behaviors. Student actions may be a response to something a teacher says, such as, “Show me how you would do this problem.” In this case, the teacher observes the student’s actions in response to an eliciting move. As well, teachers also notice student nonverbal behavior as this might signal doubt, uncertainty, or anxiety, or alternatively, confidence, determination, or pride. As teachers attend to these aspects of student responses, this process merges into how teachers render their understanding in terms of interpretations of student thinking and reasoning. That is, attending to student utterances and actions constitutes the evidence upon which teachers base their interpretations. In this process, teachers bring to interpretations what they understand generally about student reasoning at this developmental stage and in relation to this specific content—and what they hear and see from students. Interpretations then are informed both by teacher knowledge and student responses.

Some investigators use the term *noticing* in this context, referring to teachers’ ability to pick up on and make sense of student thinking in the midst of instruction. In 2011, van Es developed a three-element definition of noticing that involves “attending to noteworthy events, reasoning about such events, and making informed teaching decisions on the basis of the analysis of these observations” (p. 135). Similarly, Jacobs, Lamb, Philipp, & Schappelle’s (2011) analysis of “professional noticing” identified three elements: “attending to children’s strategies, interpreting children’s understandings, and deciding how to respond on the basis of children’s understandings” (p. 99). Walsh and Sattes (2011) developed a process for determining gaps in student understanding and pointed out that teachers need to “listen actively and carefully to a student’s answer for the purpose of identifying the student’s current understanding or skill level regarding thinking” (p. 88). While the precise terminology associated with noticing differs across these descriptive studies, the underlying construct is relatively clear.

As teachers interact with students to understand their thinking, they attend simultaneously to what students are saying and doing, and they clarify student responses with further prompts, questions, and tasks. In this sense, the two practices of questioning and attending operate in tandem (Araceli Ruiz-Primo & Furtak, 2006; Inoue & Buczynski, 2011; Lobato, Clarke, & Ellis, 2005; Ramnarain, 2011). A number of studies provide more detailed description and examples of these interconnected practices. For example, in studying teacher–student conversations when solving story problems in elementary classrooms, Jacobs and Ambrose (2008) identified one of the teacher moves as exploring “what the student has already done,” which includes asking the student to explain a partial or incorrect strategy and asking “specific questions to explore how what he has already done relates to the quantities and relationships in the problem” (p. 261).

Further, researchers who emphasize using the information gathered from student responses to inform subsequent instructional decision-making practices recommend the use of follow-up questions as well as posing additional problems (Franke, Fennema, & Carpenter, 1997). In a 2014 study, Gilson, Little, Ruegg, & Bruce-Davis, investigated elementary teachers’ use of follow-up questions in the literacy field and identified several types of follow-up questions that attend to student thinking, including “ask[ing] student[s] to give more examples or reasons demonstrating literal comprehension or recall of the text; ask[ing] student[s] to provide additional justification, opinion, or examples that demonstrate inferential comprehension of the text [*elaborate*]; and ask[ing] student[s] to support or explain his or her opinion or inference [*justify*]” (p. 114). Walsh and Sattes (2011) also recommend that at times, the “teacher needs to check with a student to ensure that they understand what the student is attempting to express” while asking clarifying questions to reveal “the thinking behind the actual response” (p. 88).

A final aspect of attending involves how teachers pace their questions, prompts, and responses sensitively, allowing wait time for students to think when responding so that they have better opportunities to use higher levels of thinking (Ellis, 1993; Goos, 2004; Martino & Maher, 1999). The widely studied concept of wait time (Rowe, 1986) operates as an element in teachers’ attentive interactions with students around academic content, and some research indicates that greater use of wait time decreases the amount of teacher talk and of lower-level questions (Tobin, 1987).

The emphasis on attention or noticing then incorporates teachers’ responses to student utterances and actions. Such responses occur in the flow of instruction so that as teachers notice something from a student’s response, they seek to probe, clarify, and connect what students are saying in order to effect a better grasp of student thinking and reasoning. Here, the emphasis is on teachers’ active listening and probing as that interacts continuously with their moves to provide further information to support their inferences about student thinking and reasoning.

As Forzani and Ball (2016) noted:

> It is important that teachers’ own ideas about a student’s thinking not crowd out the student’s actual thinking. Having hunches, knowing the content or the context well, or having a good sense of what might be typical can focus the teacher’s

> listening. But it can also overtake it. Skill with eliciting and interpreting student thinking depends on responsible use of professional knowledge about students and the content, but even more so from careful listening and recognizing that this student might be thinking differently from other students. (p. 16)

Aspect 3: Interpreting Student Thinking

Attending to or noticing student utterances and actions constitutes the evidence upon which teachers base their interpretations. In this process, teachers bring to interpretations what they understand generally about student reasoning at this developmental stage and in relation to this specific content—and what they hear and see from students. Interpretations then are informed both by teacher knowledge and student responses.

One challenging task of EST in classrooms is to interpret and analyze student thinking “on the fly” during instruction. Researchers in teacher education as well as professional development regard interpreting students’ understandings as one of the basic practices that informs teachers’ instructional decision-making (Bautista et al., 2014). Researchers who have done extensive work on noticing in mathematics education (for example, Sherin & van Es, 2002, 2005, 2009; van Es & Sherin, 2008), demonstrated the critical role that interpretation plays based on teachers’ ability to notice what happens in mathematics classrooms.

They argued that how teachers analyze what they notice is as important as what they pay attention to. For example, Sherin, Linsenmeier, and van Es (2009) studied how the use of video clips led to productive teacher discussions about student thinking. Teachers in the study demonstrated various levels of understanding about student thinking and engaged in the discussion of substantive mathematical ideas based on what they observed from the video clips. Results indicated that teachers’ discussion quality was influenced by factors including depth and clarity of student thinking, as well as windows into student thinking, that is, the extent to which students’ thinking is visible. This finding supports the claim that effective elicitation of student thinking is the basis for teachers’ understanding and interpretation.

A number of studies provide further details about how teachers interpret student thinking. For example, in the Jacobs et al. (2007) study that focused on students’ reasoning in algebra, teachers participating in the intervention were asked to generate possible strategies students might use in solving relational thinking problems, including strategies that reflect students’ correct conceptions as well as incorrect ones. Results indicated that participating teachers were able to generate multiple strategies students might use while the majority of nonparticipating teachers only provided single-strategy responses, reflecting a lack of knowledge about student thinking.

In Fernández, Llinares, and Valls's (2013) study, preservice teachers' ability to notice students' mathematical thinking was evaluated based on the extent to which they could adequately analyze students' answers to mathematical problems. Teachers were required to demonstrate the ability to describe in detail what they think each student did in response to each problem and to indicate what they learned about "students' understandings related to the comprehension of the different mathematic concepts implicated" (p. 447).

Sleep and Boerst (2012) asked interns to provide "assertion sketches" about their students' understanding or strategies supported by evidence obtained from interaction with the student. Figure 1 is an example of such an assertion. Criteria for a warranted assertion specified that it not be too broad and that it be based on evidence specific to the mathematical problem or strategy being studied.

Kelly can use the standard algorithm to add and subtract two- and three-digit numbers; however her conceptual understanding of why these algorithms work is still developing.

- Correctly computed 46 + 28, 50 - 14, and 302 - 197 using the standard algorithms
- Could use bundling sticks to solve 46 + 28, 50 - 14
- Could not explain how the numbers match what she did with the sticks—asked her to show the "little 1" with the sticks and she picked up one loose stick

Figure 1. Assertion sketch example from Sleep and Boerst, 2012, p. 1048.

Research of this kind provides careful descriptions of how teachers interpret student thinking and make appropriate inferences from evidence they have gleaned through their attending and prompting. Clearly such inferential work also depends on teachers' content knowledge for teaching because interpretation is based on an implicit or explicit comparison between what the student demonstrates and what the specific mathematics or literacy involves. Interpretation then relies both on the evidence teachers collect in interactions with students and on their content knowledge for teaching (Hill, Ball, & Schilling, 2008). "Research demonstrates," noted Russ, Sherin, and Sherin (2016), "that teachers' ability to effectively make sense of students' ideas rests firmly in their specialized knowledge" (p. 417).

Aspect 4: Understanding the Content

Early studies of teacher questioning, particularly in the context of both teacher and peer tutoring (see the Aspect 1 subsection) treated this practice as domain-independent, in concert with prevailing views in cognitive science (see D. Gitomer & Zisk, in press). As we have been describing, more recent studies have modeled EST as involving a close partnership between general and specialized knowledge (Perkins & Solomon, 1989). As Bennett (2011) has described in relation to formative assessment, domain-independent strategies are powerful but weak, serving well in handling routine problems. In contrast, subject-specific knowledge is powerful but brittle, working well under constrained conditions. From these observations, Bennett drew two implications: "A teacher who has weak cognitive-domain understanding is less likely to know what questions to ask of students, what to look for in their performance, what inferences to make from that performance about student knowledge, and what actions to take to adjust instruction" (p. 15). The second implication, he continued, is that the tools and resources supplied to teachers may differ from one domain to the next because they should be attuned to the domain in question.

This perspective on the domain dependency issue is represented as well in the practice-based measures of content knowledge for teaching described previously. Researchers argue, for example, that to be effective with questioning, teachers need strong content knowledge and pedagogical skill in many possible mathematics curriculum topics (Ball et al., 2008; Hill, Ball, & Schilling, 2008; Shulman, 1986; White, 2003).

Work in this vein also is exemplified in the teaching case studies supplied by Ball (1993a, 1993b), Lampert (1986, 1990, 2003), and others. Teachers might be equipped with tool kits of general probing or prompting questions, but without content knowledge for teaching they could not make the critical "in the moment" decisions that connect students to the subject matter. Without both common and specialized knowledge of the subject matter, teachers will struggle to interpret students' responses, unpack their reasoning, and recognize potential misconceptions and confusions.

Measurement Approach for EST Performance Assessment

This section of the paper provides a description of how EST will be measured in the NOTE assessments. It is important to be clear on a set of key design decisions, which we identify and discuss as we describe the measurement approach.

Delivery Mode

Measurement of the EST construct relies on a single digitally simulated student, known as an avatar, to be the agent that interacts with candidates. Avatars are flexible and easily adaptable for tasks on different subject matters or for use with various types of materials. Using an avatar controlled by a live individual, known as a simulation specialist, we capture the verbal and written interactions in response to a structured instructional task. Candidates elicit student thinking around specified high-leverage content as they interact with the avatar. For each task, standardized materials are developed for the simulation specialist to guide the avatar’s responses during the interaction. In these materials, a student profile is created to specify the student’s thinking process, knowledge, and conceptions in regard to the specific content topic, misconceptions, and general orientation to responses. In addition, an interaction protocol is developed to describe the steps or procedures the student takes to arrive at an answer, including recommended responses to the candidate’s possible prompts.

The primary reason for relying on an avatar as opposed to recording a live performance with a student was to standardize the testing conditions and minimize construct irrelevant variance. The task type, background materials, student profile, and simulation specialist performance are all standardized. Of these elements, simulation specialist performances are potentially the most variable. To counteract for construct irrelevant variance, protocols are developed that script responses to the greatest degree practicable. Simulation specialists are carefully trained on the tasks they will use and are certified on those tasks, with their performances audited periodically to ensure reliable and comparable performance. Rubrics are employed to evaluate the performance of the simulation specialist. The aim in these procedures is to strike a balance between maintaining a semblance of authentic interaction while providing the benefits of standardization necessary for licensure decisions.

Content

The tasks for assessing teacher candidates' performances can be carried out in various contexts, across subject areas and grade levels. Information included in the assessment scenarios often notes the grade-level of the student, the specific content to be addressed, and the student work and task. As indicated above, the NOTE assessment initially focuses on mathematics and ELA at the elementary level. The high-leverage content (referenced previously) is derived from the CCSS for the relevant grade levels included in the assessment. The tasks on EST include content and material presented in a range of formats, including mathematics problems that use formulas, graphs, charts, word problems, or text-based materials. The ELA tasks involve content that employs different genres including informational and literary texts. Appendix B includes a sample ELA task for EST.

Structural Features

The structure of the EST assessment includes three parts. First, candidates are provided with a task and a description of the goals for an interaction with a simulated student. Information about the student is brief: typically a grade level and description of a reading or mathematics assignment completed. The reason for this design feature is that tasks are driven by the subject matter rather than non-content-related aspects of students. Candidates have 10 minutes to prepare for interaction with the student avatar.

The second part of the task involves a recorded interaction with a single simulated student—an avatar in a computer-based environment—in order to elicit the "student's" thinking related to a specific task. Candidates are allowed to use notes from their planning time, and the student has access to a shared digital workspace around which he or she may interact with the candidate. Interaction then includes both verbal and written aspects. Interactions may last up to 10 minutes. A candidate can conclude an interaction before 10 minutes has elapsed if he or she decides enough of the student's understanding has been elicited.

In the third and final part of the task, candidates are given 10 minutes to answer task-specific questions about their understanding of the simulated student's thinking and knowledge based on evidence from the interaction. In math, these queries prompt candidates to provide descriptions of the student's procedures for solving mathematics problems, evaluate whether the student's answers are correct, and include explanations for why the student's procedures are or are not valid. For the ELA tasks, queries prompt candidates to offer examples of the student's understanding and/or misunderstanding of the ELA excerpt the student read and to explain how the student used the text features to achieve understanding. Candidates may use notes they may have taken during their preparation and interaction with the student when they answer these questions.

In addition, candidates are required to provide evidence from the interaction with the student to support their interpretation and analysis and to demonstrate their understanding of the student's thinking through predicting the student's possible responses to other similar problems. This aspect of the assessment is important because candidates might simply look at the student work sample provided in the preparation materials and base interpretations just on this information. The postperformance questions provide evidence that the candidate has used the interaction itself to elicit student thinking.

Scoring Criteria

The scoring rubric for this assessment will be placed on the ETS website (www.ets.org) when finalized. The aspects and indicators map onto the features of EST identified in the literature as described above. An assumption underlying this rubric is that these are distinctive, recognizable dimensions of the construct that each contributes to an overall score. Based on pilot and field test data, this assumption will be tested and the rubric will be modified accordingly.

The four criteria used to score EST performances include:

- Aspect 1: Using questions, prompts and student tasks to elicit student thinking
- Aspect 2: Attending to student talk and actions
- Aspect 3: Interpreting student thinking
- Aspect 4: Understanding the content

The first three criteria map to the key dimensions of EST identified in the research and scholarly literature. Because EST is hypothesized to rely on knowledge of the content that forms the basis for an interaction between teacher and student, the fourth aspect draws attention to this feature.

Limitations and Boundary Conditions

A number of boundary conditions and limitations are involved in this assessment task. First, EST fits into a flow of practices and also interacts with other dimensions of teaching.

Teachers plan and carry out other instructional activities before and after they elicit and interpret student thinking. For instance, they plan how to elicit information based on prior knowledge of their students. And, they use the information they gather from the elicitation to inform their subsequent instructional moves. In this case, candidates are provided minimal information about the simulated student and they are instructed to engage only in elicitation-based interactions, removing efforts to respond with instruction. In both these senses, this task is artificial when judged against its fully authentic enactment in real classrooms.

As well, interaction is limited to one student in order to keep scoring manageable and to avoid the complexities associated with how EST is implemented in small and whole group instruction. Here, too, the basic claim is that if a candidate is unable to demonstrate EST with one student, he or she is unlikely to be able to do so with groups of students, where this practice is more difficult to manage. Further, teachers do interact one on one with students at times, so from this perspective, this assessment replicates at least one form of interaction with students.

In early trials of this assessment, candidates reported that the interaction with the avatar felt reasonably authentic. In the first trial with candidates, they mostly agreed with the statement, "The student's responses were authentic to the way I think elementary school students would respond" (n = 41, M = 3.4 on a 4-point Likert scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). In a subsequent trial, candidates mostly agreed the teaching practice was authentic: "Eliciting student thinking as required for these tasks felt authentic to me" (n = 13, M = 3.2 on a 4-point Likert scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Phrased differently in another trial, 35 candidates who each completed four tasks mostly agreed with the statement, "The kind of interaction required by this task felt authentic to me" (n = 140, M = 3.6 on a 4-point Likert scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*).

The task also required bounding as to its starting and ending points. Candidates are provided time to study the task and other background materials as a proxy for teacher planning, given time to interact, then evaluated on their interpretation and understanding of what the simulated student said. Crucially, candidates are instructed not to provide instruction during the interaction. Literature discussing this practice closely connects its theoretical foundations and enactment with other competencies and instructional moves that include, for example, extending student thinking, providing quality feedback, remediating student errors, and engaging in extended discussions (Jacobs & Ambrose, 2008; Soter et al., 2008). Consequently, this restriction limits the authenticity of the task but is necessary in order to provide an accurate gauge of this practice, which we isolated to avoid confounding with other instructional practices.

We chose to limit our attention to the four criteria identified in the scoring rubric as these relate most directly to the targeted construct. The hypothesis underlying this design choice is that if teacher candidates are unable to elicit and interpret student thinking, they are more likely to build on incomplete or incorrect information as the basis for their instructional decisions and actions. While in real interactions teachers undoubtedly blend eliciting with instructing, the claim for this assessment is that EST is a separable practice that is necessary to effective instruction.

Conclusion

Support for the EST construct rests on advances in cognitive science, instructional design, and consensus views of good teaching. Important in the process of validating this construct is transparency so that those affected by measurement understand and have confidence in the reasoning and the evidence that supports its measurement and use in high-stakes decisions such as licensure. The intent of this paper is to contribute to this transparency by setting forth the research base and related evidence supporting the construct together with the approach that ETS is taking to its measure for use in NOTE.

Most noticeable in the ETS approach is a concentration on practice. Traditionally, licensure for teaching has involved tests of knowledge thought to underlie practice. The warrant is in the knowledge, with the assumption that the absence of such knowledge undercuts the claim about readiness for practice. This emphasis on knowledge and skills makes good sense and is a feature in licensure for all professions and occupations. The NOTE assessments continue this tradition with a substantial battery of knowledge measures that extend deeply into the use of such knowledge in the practice of teaching. Here, we further extend the warrant for licensure to direct measures of teaching practices themselves that we argue are central to teaching. The claim is not that these practices alone make up effective teaching but rather that they are central and must be included. In all fields the warrant for entry combines licensure assessments with the content of accredited programs such that between them the full complement of knowledge and skill is conveyed and assessed. Working out the division of responsibility between licensure examinations and programs of preparation is a matter for professional judgment rendered by experts and stakeholders in a field of practice.

EST meets the standard for centrality to effective teaching practice due to its consensus representation across many efforts to define good teaching. Such efforts have included research studies, observation instruments, accounts of best practice, teaching standards, and others. At the same time, the details associated with this construct matter and are under continuous negotiation as new knowledge accumulates, terms are redefined, new aspects are highlighted, and new evidence amassed. ETS intends to contribute to this ongoing conversation through its validity work on this construct. We invite readers to enter this conversation, recognizing that what is described here is neither the first word nor the last, but a contribution that sets a stake in the ground requiring at once robust defense and openness to new developments, new knowledge, and new challenges rising from the field of teaching research, policy, and practice.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (4th ed.). Washington, DC: American Educational Research Association.

Appalachia Education Laboratory. (1994). *Questioning and understanding to improve learner thinking (QUILT): The evaluation results*. Charleston, WV: Author.

Araceli Ruiz-Primo, M., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 237–263. doi:10.1080/10627197.2006.9652991

Ball, D. L. (1993a). Halves, pieces, and twoths: Constructing and using representational contexts in teaching fractions. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 157–195). New York, NY: Routledge.

Ball, D. L. (1993b). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal, 93*(4), 373–397. doi:10.1086/461730

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511. doi:10.1177/0022487109348479

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. doi:10.1177/0022487108324554

Barnette, J., Orletsky, S., & Sates, B. (1994). *Evaluation of teacher classroom questioning behaviors*. Retrieved from ERIC database. (ED377188)

Bautista, A., Brizuela, B., Glennie, C. R., & Caddle, M. C. (2014, July). Mathematics teachers attending and responding to students' thinking: Diverse paths across diverse assignments. *International Journal for Mathematics Teaching & Learning*, 1–28.

Bennett, R. (2011). Formative assessment: A critical review. *Assessment in education: Principles, policy, and practice, 18*(1), 5–25.

Bloom, B. (1984). The two-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13,* 4–16.

Boaler, J., & Brodie, J. (2004, October). *The importance, nature, and impact of teacher questions.* Paper presented at the proceedings of the 26th annual meeting of the North American Chapter of the International Group for Psychology of Mathematics Education, Toronto, Ontario.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. Washington, DC: National Academy Press.

Brown, J. S., Collins, A., & Duguid, P. (1988). *Cognitive apprenticeship, situated cognition and social interaction*. Cambridge, MA: Bolt Beranek and Newman.

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

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. (2000). *Cognitively guided instruction: A research-based teacher professional development program for elementary school mathematics.* Retrieved from ERIC database. (ED470472).

Charlesworth, R. (1998). Developmentally appropriate practice is for everyone. *Childhood Education, 74*(5), 274–282. doi:10.1080/00094056.1998.10521951

Chi, M. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10*(7), 33–49.

Chi, M. T., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

Christman, E. P., Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual*, *1,* 91–104.

Clauser, B. E., Margolis, M. J., & Case, S. M. (2006). Testing for licensure and certification in the professions. *Educational Measurement*, *4,* 701–731.

Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*, 237–248.

Common Core State Standards Initiative. (2015a). *English language arts standards.* Retrieved from http://www.corestandards.org/ELA-Literacy/

Common Core State Standards Initiative. (2015b). *Standards for mathematical practice.* Retrieved from http://www.corestandards.org/Math/Practice/

Cotton, K. (1989). *Expectations and student outcomes*. Portland, OR: Northwest Regional Educational Laboratory.

Council of Chief State School Officers. (2013). *Interstate Teacher Assessment and Support Consortium InTASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development.* Washington, DC: Author.

Danielson, C. (2013). *The framework for teaching.* Princeton, NJ: The Danielson Group.

Dillon, J. T. (1988). *Questioning and teaching.* New York, NY: Teachers College Press.

Dillon, J. T. (1990). *The practice of questioning*. New York, NY: Routledge.

Ellis, K. (1993, February). *Teacher questioning behavior and student learning: What research says to teachers*. Paper presented at the annual meeting of the Western States Communication Association, Albuquerque, NM.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*(4), 403–434. doi:10.2307/749875

Fernández, C., Llinares, S., & Valls, J. (2013). Primary school teacher's noticing of students' mathematical thinking in problem solving. *The Mathematics Enthusiast*, *10*(1/2), 441–468.

Forzani, F. M., & Ball, D. L. (2016). Eliciting and interpreting individual student's thinking. In F.M. Forzani & D. Ball (Eds.), *High leverage practices for competent beginning teaching.* Manuscript in preparation.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in everyday mathematics classrooms. *Journal for Research in Mathematics Education, 30*(2), 148–170. doi:10.2307/749608

Franke, M. L., Fennema, E., & Carpenter, T. (1997). Teachers creating change: Examining evolving beliefs and classroom practice. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 255–282). Mahwah, NJ: Lawrence Erlbaum.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Charlotte, NC: Informationa Age.

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education, 60*(4), 380–392. doi:10.1177/0022487109339906

Fuchs, D., & Fuchs, L. (2007). Increasing strategic reading comprehension with peer-assisted learning activities. In D. S. McNamara (Ed.), *Reading comprehension strategies* (pp. 175–198). New York, NY: Lawrence Erlbaum.

Fuson, K. C., Kalchman, M., & Bransford, J. D. (2005). Mathematical understanding: An introduction. In M. Donovan & J. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 217–256). Washington, DC: National Academy Press.

Gall, M. (1970). The use of questions in teaching. *Review of Educational Research, 40*(5), 707–721. doi:10.3102/00346543040005707

Gall, M. (1984). Synthesis of research on teachers' questioning. *Educational Leadership, 42*(3), 40–47.

Gilson, C. M., Little, C. A., Ruegg, A. N., & Bruce-Davis, M. (2014). An investigation of elementary teachers' use of follow-up questions for students at different reading levels. *Journal of Advanced Academics, 25*(2), 101–128. doi:10.1177/1932202x14532257

Gitomer, D., & Zisk, R. (in press). Knowing what teachers know. *Review of Research in Education.*

Gitomer, D. H., Phelps, G., Weren, B. H., Howell, H., & Croft, A. J. (2014). Evidence on the validity of content knowledge for teaching assessments. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 493–528). San Francisco, CA: Jossey-Bass.

Goos, M. (2004). Learning mathematics in a classroom community of inquiry. *Journal for Research in Mathematics Education, 35*(4), 258–291. doi:10.2307/30034810

Graesser, A., & Person, N. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137.

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores*. Retrieved from http://www.nber.org/papers/w16015

Harris, R. L. (2000). Batting 1,000: Questioning techniques in student-centered classrooms. *The Clearing House, 74*(1), 25–26. doi:10.1080/00098655.2000.11478634

Hewson, M. G., & Hewson, P. W. (1983). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching, 20*(8), 731–743. doi:10.1002/tea.3660200804

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking "pedagogical content knowledge": Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H. C., Blunk, M. L., Charalambous, C., Lewis, J. M., Phelps, G., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. doi:10.1080/07370000802177235

Hill, H. C., Charalambous, C. Y., Blazar, D., McGinn, D., Kraft, M. A., Beisiegel, M., . . . Lynch, K. (2012). Validating arguments for observational instruments: Attending to multiple sources of variation. *Educational Assessment, 17*(2–3), 88–106. doi:10.1080/10627197.2012.715019

Inoue, N., & Buczynski, S. (2011). You asked open-ended questions, now what? Understanding the nature of stumbling blocks in teaching inquiry lessons. *The Mathematics Educator, 20*(2), 10–23.

Jacobs, V. R., & Ambrose, R. C. (2008). Making the most of story problems. *Teaching Children Mathematics, 15*(5), 260–266.

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education, 38*(3), 258–288.

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202.

Jacobs, V. R., Lamb, L. L., Philipp, R. A., & Schappelle, B. P. (2011). Deciding how to respond on the basis of children's understandings. In M. G. Sherin, V. Jacobs, & R. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 97–116). New York, NY: Routledge.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement*, *2*(3), 135–170.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (pp. 17–64). Westport, CT: American Council on Education and Praeger.

Kane, T., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587–613. doi:10.1353/jhr.2011.0010

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.

King, A. (1991). Effects of training in strategic questioning on children's problem solving performance. *Journal of Educational Psychology, 83*(3), 307–317.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teachingchildren how to question and how to explain. *American Educational Research Journal, 31*(2), 338–368.

King, A. (2007). Beyond literal comprehension: A strategy to promote deep understanding oftext. In D. S. McNamara (Ed.), *Reading comprehension strategies* (pp. 267–290). New York, NY: Lawrence Erlbaum.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education, 61*, 127–148.

Kostelnik, M. J., Soderman, A. K., & Whiren, A. P. (2007). *Developmentally appropriate curriculum: Best practices in early childhood education*. Upper Saddle River, NJ: Prentice Hall.

Kulik, J. A., Kulik, C.-L., & Bangert-Drowns, R. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior,* 1(1), 59–74.

Lampert, M. (1986). Knowing, doing, and teaching multiplication. *Cognition and Instruction, 3*(4), 305–342. doi:10.1207/s1532690xci0304_1

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29–63. doi:10.3102/00028312027001029

Lampert, M. (2003). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409–426. doi:10.1086/499760

Levin, D. M., Grant, T., & Hammer, D. (2012). Attending and responding to student thinking in science. *The American Biology Teacher, 74*(3), 158–162. doi:10.1525/abt.2012.74.3.6

Lobato, J., Clarke, D., & Ellis, A. B. (2005). Initiating and eliciting in teaching: A reformulation of telling. *Journal for Research in Mathematics Education, 36*(2), 101–136.

Martino, A. M., & Maher, C. A. (1999). Teacher questioning to promote justification and generalization in mathematics: What research practice has taught us. *The Journal of Mathematical Behavior, 18*(1), 53–78. doi:10.1016/s0732-3123(99)00017-6

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-pronged comprehension strategy framework. In D. S. McNamara (Ed.), *Reading comprehension strategies* (pp. 465–496). New York, NY: Lawrence Erlbaum.

Mevarech, Z. R., & Rich, Y. (1985). Effects of computer-assisted mathematics instruction on disadvantaged pupil's cognitive and affective development. *Journal of Educational Research, 79*(1), 5–11.

Mewborn, D. S., & Huberty, P. D. (1999). Questing your way to the standards. *Teaching Children Mathematics, 6*(4), 226–227, 243–246.

Monte-Sano, C. (2011). Learning to open up history for students: Preservice teachers' emerging pedagogical content knowledge. *Journal of Teacher Education*, *62*(3), 260–272. doi:10.1177/0022487110397842

Murphy, P. K., Wilkinson, I. A., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*(3), 740–764. doi:10.1037/a0015576

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* Retrieved from http://www.nctm.org/PtA/

National Council of Teachers of Mathematics. (2016). *Using a journal article as a professional development experience.* Retrieved from http://www.nctm.org/Conferences-and-Professional-Development/Professional-Development-Guides-(Reflection-Guides)/Principles-to-Actions_-Ensuring-Mathematical-Success-for-All

Ornstein, A. C. (1988). Questioning: The essence of good teaching—Part II. *NASSP Bulletin, 72*(505), 72–80. doi:10.1177/019263658807250514

Papageorgiou, S., & Tannenbaum, R. J. (2016). Situating standard setting within argument-based validity. *Language Assessment Quarterly, 13*(2), 109–123.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 416–432. doi:10.1037/0278-7393.17.3.416

Perkins, D., & Solomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*(1), 16–35.

Person, N., Graesser, A., Magliano, J., & Kruez, R. (1994). Inferring what the student knows inone-to-one tutoring: The role of student questions and answers. *Learning and Individual Differences, 6*(2), 205–229.

Peterson, P. L., Fennema, E., & Carpenter, T. P. (1991). Using children's mathematical knowledge. In B. Means (Ed.), *Teaching advanced skills to educationally disadvantaged students* (pp. 103–128). Menlo Park, CA: SRI International.

Pianta, R. C., Karen, M., Paro, L., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual, pre-K*. Baltimore, MD: Paul H. Brookes.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*(1), 91–109. doi:10.1207/s15326985ep2701_7

Ramnarain, U. (2011). Teachers' use of questioning in supporting learners doing science investigations. *South African Journal of Education, 31*(1), 91–101.

Raymond, M. R., & Luecht, R. M. (2013). Licensure and certification testing. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (Vol. 3, pp. 391–414). Washington, DC: American Psychological Association.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, *51*(2), 237–245. doi:10.3102/00346543051002237

Roschelle, J. M., Pea, R., & Hoadley, C. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children, 10*(2), 76–101.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education, 37*(1), 43–50. doi:10.1177/002248718603700110

Russ, R. S., Sherin, B. L., & Sherin, M. G. (2016). What constitutes teacher learning? In D. Gitomer & C. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 391–438). Washington, DC: American Educational Research Association.

Seo, Y.-J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*(3), 913–928.

Sfard, A., & Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students' mathematical interactions. *Mind, Culture, and Activity, 8*(1), 42–76. doi:10.1207/s15327884mca0801_04

Sherin, M. G., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York, NY: Routledge.

Sherin, M. G., Linsenmeier, K. A., & van Es, E. A. (2009). Selecting video clips to promote mathematics teachers' discussion of student thinking. *Journal of Teacher Education, 60*(3), 213–230. doi:10.1177/0022487109336967

Sherin, M. G., & van Es, E. A. (2002). *Using video to support teachers' ability to interpret classroom interactions.* Paper presented at the Society for Information Technology & Teacher Education International Conference, Chesapeake, VA.

Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, *13*(3), 475.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20–37. doi:10.1177/0022487108328155

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. doi:10.3102/0013189x015002004

Sidney, P. G., & Alibali, M. W. (2015). Making connections in math: Activating a prior knowledge analogue matters for learning. *Journal of Cognition and Development*, *16*(1), 160–185. doi:10.1080/15248372.2013.792091

Silver, E. A., Smith, M. S., & Nelson, B. S. (1995). The QUASAR project: Equity concerns meet mathematics reforms in the middle school. In W. G. Secada, E. Fennema, & L. B. Adjanian (Eds.), *New directions for equity in mathematics* (pp. 9–56). New York, NY: Cambridge University Press.

Silver, E. A., & Stein, M. K. (1996). The Quasar Project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education, 30*(4), 476–521. doi:10.1177/0042085996030004006

Sleep, L., & Boerst, T. A. (2012). Preparing beginning teachers to elicit and interpret students' mathematical thinking. *Teaching and Teacher Education, 28*(7), 1038–1048. doi:10.1016/j.tate.2012.04.005

Snow, C., Griffin, P., & Burns, M. S. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco, CA: Jossey-Bass.

Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research, 47*(6), 372–391. doi:10.1016/j.ijer.2009.01.001

Spires, H. A., & Donley, J. (1998). Prior knowledge activation: Inducing engagement with informational texts. *Journal of Educational Psychology*, *90*(2), 249–260. doi:10.1037/0022-0663.90.2.249

Stanford University. (2013). *The Protocol for Language Arts Teaching Observation (PLATO).* Retrieved from http://platorubric.stanford.edu/Elements.html

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). New York, NY: Teachers College Press.

Tannenbaum, R. J., & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol 3. Testing and assessment in school psychology and education* (pp. 455–478). Washington, DC: American Psychological Association.

TeachingWorks. (2016a). *High-leverage content.* Retrieved from http://www.teachingworks.org/work-of-teaching/high-leverage-content

TeachingWorks. (2016b). *High-leverage practices.* Retrieved from http://www.teachingworks.org/work-of-teaching/high-leverage-practices

Tobin, K. (1987). The role of wait time in higher cognitive learning. *Review of Educational Research, 57*(1), 69–95.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press.

van Es, E. A. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. Jacobs, & R. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 134–151). New York, NY: Routledge.

van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching & Teacher Education, 24*(2), 244–276. doi:10.1016/j.tate.2006.11.005

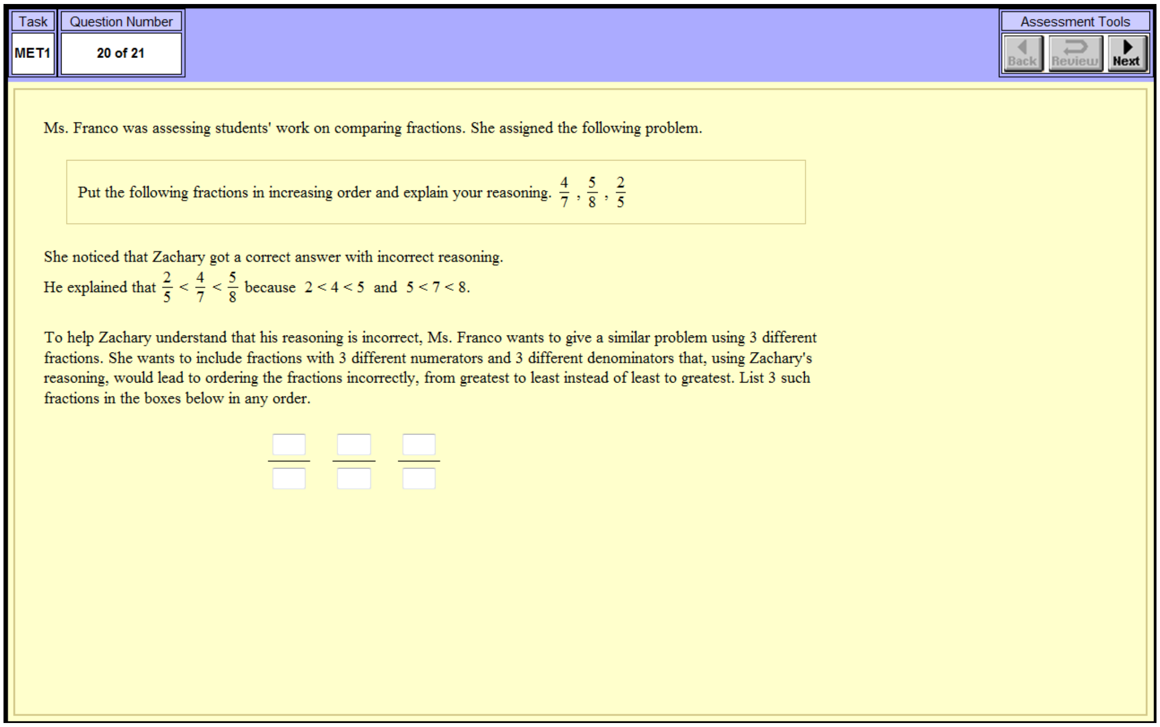
Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences, 6*(2), 227–269. doi:10.1207/s15327809jls0602_3

Villaseñor, A., & Kepner, H. S. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24*(1), 62–69. doi:10.2307/749386

Walsh, J. A., & Sattes, B. D. (2011). *Thinking through quality questioning: Deepening student engagement*. Thousand Oaks, CA: Corwin Press.

White, D. Y. (2003). Promoting productive mathematical classroom discourse with diverse students. *The Journal of Mathematical Behavior, 22*(1), 37–53. doi:10.1016/s0732-3123(03)00003-8

Appendix A. Sample EST Task


Task | Question Number | Assessment Tools

MET1 | 20 of 21 | Back Review Next

Ms. Franco was assessing students' work on comparing fractions. She assigned the following problem.

Put the following fractions in increasing order and explain your reasoning. $\frac{4}{7}, \frac{5}{8}, \frac{2}{5}$

She noticed that Zachary got a correct answer with incorrect reasoning.

He explained that $\frac{2}{5} < \frac{4}{7} < \frac{5}{8}$ because $2 < 4 < 5$ and $5 < 7 < 8$.

To help Zachary understand that his reasoning is incorrect, Ms. Franco wants to give a similar problem using 3 different fractions. She wants to include fractions with 3 different numerators and 3 different denominators that, using Zachary's reasoning, would lead to ordering the fractions incorrectly, from greatest to least instead of least to greatest. List 3 such fractions in the boxes below in any order.

Appendix B. Sample Released Task for EST

ELICITING STUDENT THINKING
UPPER ELEMENTARY ENGLISH LANGUAGE ARTS
Reading Narrative Text (Grade 5)

Candidate Materials

This task is part of an assessment designed to collect evidence of a prospective teacher's ability to elicit and interpret an individual student's thinking. Each task in the assessment is designed to focus on a particular student's thinking about a specific high-leverage content topic in mathematics or English language arts at the elementary level.

Teaching Practice and Content Topic Focus of This Task

High-Leverage Practice: Eliciting and interpreting individual students' thinking

Teachers pose questions or tasks that provoke or allow students to share their thinking about specific academic content in order to evaluate student understanding, guide instructional decisions, and elicit ideas that will benefit other students. To do this effectively, a teacher draws out a student's thinking through carefully chosen questions and tasks, attends to what the student says and does, and considers and checks alternative interpretations of the student's ideas and methods.

High-Leverage Content: Comprehension of narrative text

Assessment Task Overview:

There are three parts to this assessment.

Part 1: Preparation for interaction (10 minutes)—In the first part of this task, you will be provided with a student task and a description of the goals for your interaction with a simulated student (i.e., an avatar).

Part 2: A recorded interaction with the simulated student (10 minutes)—You will interact with a simulated student in order to elicit the student's thinking and understanding related to a specific task. You and the student will have access to a shared digital workspace through the use of a tablet (iPad).

Part 3: Post-performance questions (10 minutes)—After you have completed your interaction, you will have ten minutes to answer questions about your understanding of the student's thinking and knowledge based on evidence from your interaction. You are permitted to use any notes you may have taken during Part 1 and Part 2 to answer the questions.

General Instructions:

- You should assume that you have already established a relationship with the student and that you do NOT need to begin the interaction by "breaking the ice." You should begin by asking about the work right away.
- If you realize that the student has a misunderstanding, you should ask probing, follow-up questions to learn more about that misunderstanding. You should NOT provide instruction. It is NOT your goal to change the student's thinking or understanding.
- You should NOT indicate future instructional steps based on what you elicit.
- You may refer to notes that you make during the preparation time while you are interacting with the student and when responding to the post-performance questions. All materials will be turned in at the end of the task.

- During the interaction, you and the student will have the ability to read a shared document on the iPad. You will both be able to write on this document and see one another's writing. However, you cannot write on the document on the iPad as you prepare for the interaction.
- You should stop the interaction when you feel you have sufficiently elicited the student's thinking about the student task. You do NOT need to use all of the time allowed for the interaction.
- You do NOT need to include any closing remarks to the student.

Assessment Task Evaluation Criteria:

Your performance on the assessment task will be evaluated on the following criteria.

- Your ability to ask appropriate questions and to probe the student's thinking
- Your ability to attend to the student's ideas during the interaction
- Your ability to interpret the student's thinking accurately

Part 1: Preparation (10 minutes maximum)

Scenario:

A fourth-grade student has been asked to read the story "Frog Aisle" independently and think about what Sam and Violet think of each other. You meet with the student to elicit the student's ideas about what these two characters think of one another.

Your goals:

When you are eliciting the student's thinking, be sure to do the following.

- Ask questions about both characters and what they think of each other.
- Elicit how and why the student came to these ideas.
- Ask for specific language that the student used from the text.
- Probe for breakdowns in the student's understanding.
- Pay attention to the general strengths and weaknesses of the student's comprehension and ability to respond to your questions.

Directions:

You have ten minutes to prepare for the interaction, during which time you should do the following.

- Read the story "Frog Aisle" closely. The story can be found on the next page and will be displayed on the shared workspace. As you read, think about how the student might read and understand the text, and identify aspects that might be challenging for the student.
- Generate potential questions you might ask that will elicit the student's understanding of the story.
    - The questions could include initial opening questions and also probing follow-up questions about the text.
    - Consider probing and follow-up questions you might ask that will elicit *how* or *why* the student came to that understanding.
    - Use these questions to guide your interaction, but be flexible enough to respond to the student.

You are welcome to take notes on the paper provided, but you cannot mark up the shared document on the shared workspace during the preparation time. If you choose to take notes, you are free to use them during the interaction and when responding to the post-performance questions in Part 3 of the assessment.

Frog Aisle

Since last year when she landed in our class like a meteor, Violet Chang had inspired in me a deep but secret admiration. I knew certain things about her. She could burp the whole alphabet, her bike was sparkly purple, and she kept an old bottle cap in her backpack for luck. But we'd never spoken to each other. At least, we hadn't until one summer Saturday in the public library. Who goes to the library on summer Saturdays? I do. And so, it turned out, did Violet Chang.

One peep and the librarian, Mrs. Gonzalez, could appear with a scolding finger before you could blink. Unless you were Violet. Whenever she muttered to herself or laughed at what she was reading, Mrs. G would simply ask a question about what Violet was reading. They'd end up having a big discussion about the metric system or, say, whether Pluto is still a planet.

That particular afternoon, I knew Violet was camped out on the floor in the Reptiles aisle, draped over an oversized book. I usually skipped that whole area (my brother watched way too many animal programs), but, that day, I dawdled on the next aisle.

I must be a terrible spy. Violet reached across the books and poked me in the leg. "Hey, Sam," she said. "Come over here." I obeyed, collapsing near her on the floor. She leafed through a few pages full of frog pictures, pointing to some of her favorites.

Before I had time to think, I blurted, "Did you know there are only three types of amphibians? It's a special class of animal."

She examined me for a moment, then replied, "Yes, well, there are only three orders, but thousands of species. I think they're discovering new ones all the time."

I swallowed loudly, wondering if Mrs. G. was within earshot. "If you had a pet frog, what would you name it?"

"Peppy as a tadpole. Pepper as an adult."

I liked it that she hadn't hesitated. "Wouldn't that confuse him?"

"Oh, he would understand. Sometimes you're one thing when you're young, and another when you're older. You have to keep up with yourself."

"I hope you won't tell him, 'When you're older, you'll understand.' "

She wrinkled her nose. "I hate it when people say that. Peppy will be a smart young amphibian, and he'll know who not to listen to." Violet tried to whisper *ribbit* but it came out more like a burp. Her giggle made me blush. "If you want," she said suddenly, "I could send you some links to Amazon rain-forest photography. Some of the frogs are pretty amazing."

"Sure," I said.

Suddenly, I heard a noise in the next aisle and there was Mrs. G's face, her eyes bulbous and glassy, looking at us sideways.

Figure B1. Student text as it will be displayed on the shared workspace.

Part 2: Interaction
(10 minutes maximum)

Do you have any questions before you start the interaction? If so, please ask the proctor now. If not, tell the proctor that you are ready to start the interaction.

The proctor will show you where to sit for your interaction and will provide you with your shared workspace device (iPad).

You will have ten minutes for the interaction. A timer will display the time remaining. You do not need to use all of the time allowed. When you feel that you have sufficiently elicited the student's thinking so that you can interpret it with the specific goals in mind, you may end the interaction.

We will be recording your performance so that we can evaluate it later.

Reminders:

- You will use the evidence you collect from the interaction to support your responses to the questions in Part 3 of this task.
- You are free to use any notes you took during the preparation period and the interaction.
- You should not teach; your goal is to elicit and interpret the student's thinking and understanding.
- You do not need to "break the ice" to start the conversation.

When you are ready to begin the interaction, please hold your participant ID card in front of you and say, "Begin interaction." (You can put the ID card down once the interaction starts.)

When you are finished, please hold your participant ID card in front of you again and say, "End interaction."

After the interaction, the proctor will show you where to enter your responses to the post-performance questions.

Notes

[1] NOTE involves three performance assessments: Leading a Classroom Discussion, Eliciting Student Thinking, and Modeling and Explaining Content.

[2] In this paper we refer to what teachers do, intending this to mean effective teachers. The reader should just be aware of this important point.