Leading a Classroom Discussion: Definition, Supporting Evidence, and Measurement of the *ETS*® National Observational Teaching Examination (NOTE) Assessment Series

Margaret Witherspoon

Gary Sykes

Courtney Bell

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.

Leading a Classroom Discussion:
Definition, Supporting Evidence, and Measurement of the
***ETS*® National Observational Teaching Examination (NOTE) Assessment Series**

Margaret Witherspoon, Gary Sykes, and Courtney Bell
Educational Testing Service, Princeton, New Jersey

August 2016

Corresponding author: Margaret Witherspoon, E-mail: mwitherspoon@ets.org

Action Editor: Heather Buzick

Reviewers: Heather Howell and Michael Kane

Abstract

This paper provides a description and rationale for a performance assessment of a teaching practice—leading a classroom discussion (LCD)—included in the *ETS*® National Observational Teaching Examination (NOTE) assessment series. In this assessment, candidates interact with a small class of virtual students represented by avatars in a computer-based, simulated classroom. The five avatars are enacted by a single simulation specialist who has been trained and certified on the particular task presented, either in elementary English language arts or mathematics. The paper defines and describes the construct of LCD, then provides a review of the research and scholarly literature that supports the importance of this practice for effective teaching, and finally describes how the construct is measured in the NOTE assessment.

Key words: high-leverage teaching practices, teaching performance assessments, teacher licensure, measurement of teaching

Acknowledgments

Some of the content in this report is used in the following companion reports: *Eliciting Student Thinking: Definition, Research Support, and Measurement of the* ETS® *National Observational Teaching Examination (NOTE) Assessment Series* (RM-16-06) by Yi Qi and Gary Sykes; and *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.

Table of Contents

This paper describes and provides support for an assessment of a critical practice of teaching—leading a classroom discussion (LCD). This practice recently has risen to prominence among scholars of teaching (Resnick, Asterhan, & Clarke, 2015) but has deep roots dating back to the Socratic dialogues (Haroutunian-Gordon, 1991, 2010); forward to philosophical underpinnings provided by Buber (1970), Burbules (1993), Hawkins (1974), and others in the dialogic tradition; and in the sociocultural tradition of learning supplied by such scholars as Bakhtin (1981, 1986) and Vygotsky (1978).

Teachers promote learning in many ways that include lectures, recitation, work on projects, and others. But among instructional practices that teachers employ, discussion is an important, even critical, method. Prominent scholars today advocate discussion-based teaching because it creates opportunities for students to practice important skills such as argumentation, critical thinking, and collaboration (Gall & Gillett, 1980; Hadjioannou, 2007; Larson, 1996; Sun, Anderson, Lin, & Morris, 2015; Walshaw & Anthony, 2008). Further, discussion is a vital form of participation in the democratic way of life, as students are encouraged to find their voice, express and defend their views, function in communities of inquiry, and learn from and respond to the ideas and opinions of their classmates (Nystrand, 1997; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003; Parker, 2006; Reisman, 2015; Resnick, Michaels, & O'Connor, 2010).

A Performance Assessment

Before describing discussion-based teaching in greater detail, we provide a brief preview of the LCD assessment component of the *ETS*® National Observational Teaching Examination (NOTE) assessment series, developed by Educational Testing Service (ETS), for the reader to keep in mind. The LCD assessment engages the candidate in a simulated classroom environment that is delivered via computer. Prior to the assessment, the candidate is provided with curriculum materials and a lesson goal with time allotted for study and preparation. Then, the candidate conducts a short discussion of some 15 minutes duration with five "students" represented by avatars in the computer "classroom." The avatars are controlled by a trained and certified simulation specialist such that the candidate is engaged in an interaction with the simulated students around the lesson content that is provided. The simulation specialist can see and hear the candidate. The candidate's performance is videotaped and scored based on a rubric applied by trained and certified raters. 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

avatars—is rated on a rubric designed to assess the specialist's adherence to the standardization guidelines. Further details of this assessment are provided below.

The National Observational Teaching Examination (NOTE)

This performance assessment is one part of a new licensure examination known by its acronym, NOTE, which includes a combination of performance assessments and assessments of the common and specialized knowledge used in teaching. The assessment has been 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)

NOTE concentrates 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 (Teaching Works, 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 referenced in the section, LCD and the License to Teach). The NOTE assessments focus on practices of teaching content based on samples of tasks and items from the relevant content domain. NOTE's initial focus is on mathematics and English language arts (ELA) teaching at the elementary level.

LCD and the License to Teach

Assessment of the LCD construct 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 & Luecht, 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, APA, & NCME, 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 specific competencies are to effective practice.

Validation of test scores for a given purpose, including those used for licensure, relies on what has been termed, following Toulmin (2003), an *argument-based approach* (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 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" (Kane, 2004, p. 149), as may be derived from theory or empirical research.

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" (Kane, 2004, p. 149), and a variety of methods for standard setting have been established (see Tannenbaum & Katz, 2013).

The purpose of this report is to provide backing for the warrants associated with the use of the LCD performance assessment for teacher licensure. In addition to the backing derived from the scholarly and research literature on this construct, ETS also is surveying practitioners on the importance of LCD and other critical practices assessed in the NOTE assessment series. In separate studies, ETS is conducting standard setting for LCD, which will not be taken up in this report.

Support for the importance of this construct begins with implications derived from standards for learning, with the claim that such learning logically requires opportunities for

students to engage in discussion. For example, the *Common Core State Standards: ELA* (National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA & CCSSO], 2010a) includes attention to speaking and listening, of which one anchor standard is "to prepare for and participate effectively in a range of conversations and collaborations with diverse partners, building on others' ideas and expressing their own clearly and persuasively" (p. 22). Grade 5 standards go on to describe that students "come to discussions prepared . . . explicitly draw on that preparation and other information known about the topic to explore ideas under discussion; follow agreed-upon rules for discussions and carry out assigned roles; pose and respond to specific questions by making comments that contribute to the discussion and elaborate on the remarks of others; review key ideas expressed and draw conclusions in light of information gained from the discussions" (NGA & CCSSO, 2010a, p. 24).

Likewise, in the *Common Core State Standards: Mathematics* (NGO & CCSSO, 2010b), one of eight key mathematical practices calls for students to "construct viable arguments and critique the reasoning of others" (p. 6), continuing that "students at all grades can listen or read the arguments of others, decide whether they make sense, and ask useful questions to clarify or improve the arguments" (p. 7).

To realize learning goals of this kind, prominent teaching standards emphasize the need for discussion. The *Interstate Teacher Assessment and Support Consortium Model Teaching Standards* (CCSSO, 2013) calls on teachers to employ a variety of instructional strategies within which, for one "performance . . . the teacher asks questions to stimulate discussion that serves different purposes (e.g., probing for learner understanding, helping learners articulate their ideas and thinking processes, stimulating curiosity, and helping learners to question)" (p. 38). Then, in proposing a progression of knowledge and skill from beginning to advanced forms of practice, the document sets as a beginning competence, "The teacher develops learners' abilities to participate in respectful, constructive discussions of content in small and whole group settings. S/he establishes norms that include thoughtful listening, building on one another's ideas, and questioning for clarification" (CCSSO, 2013, p. 40).

As well, in Danielson's widely used *Framework for Teaching* (2011), "using questions and discussion techniques" (component 3-B in the instructional domain) is proposed as "the only instructional strateg[y] specifically referred to in the *Framework for Teaching*, a decision that reflects [its] central importance to teachers' practice" (p. 59). Another general observation

instrument, the Classroom Assessment Scoring System (Pianta, Hamre, & Mintz, 2012), indicates "instructional dialogue" as a key dimension characterized by such features as "cumulative, content-driven exchanges," "distributed talk," and facilitation strategies described at the high end as "the teacher and students frequently use facilitation strategies that encourage more elaborated dialogue, such as open-ended questions, repetition/extension, and active listening" (p. 89).

Subject-specific observation protocols also call for discussion. Among its 13 elements, the Protocol for Language Arts Teaching Observation (2013a) includes the following:

> The element of **Classroom Discourse** focuses on the opportunities students have for extended ELA-related talk with the teacher or among peers, and the extent to which the teacher and other students pick up on, build on, and clarify each other's ideas. At the low end, the teacher does the majority of the talking and, if student talk is present, the teacher and students do not build on previous responses; rather, the talk is disconnected. At the highest level, students engage in elaborated, coherent, and focused discussions, in which the teacher and other students build on each other's contributions and prompt each other to clarify and specify their ideas. (para. 10)

And, the Mathematical Quality of Instruction protocol emphasizes the ways that students participate in rich and productive discourse with the teacher and among themselves (Hill, Kapitula, & Umland, 2011).

The argument for LCD as an important practice for teacher licensure thus involves a two-fold justification: that discussion provides necessary opportunities to realize important learning goals while reflecting consensus standards for effective teaching in service to those goals.

We first provide a definition and rationale for the LCD construct followed by identification of key aspects or elements composing this practice. Next, we discuss how the construct has been measured in the context of a review of the research literature supporting LCD as a critical teaching practice. Then, we provide a description of ETS's approach to measurement of the construct, followed by a detailed rationale for the scoring aspects as grounded in the critical aspects of the practice. A brief conclusion offers summary reflections.

Construct Definition, Rationale, and Explication

Using TeachingWorks's (2016b) and Shaughnessy & Forzani's (2012) definition of classroom discussion coupled with our review of existing literature on the practice, we define classroom discussion as follows:

> A classroom discussion is a sustained exchange between and among teachers and their students with the purpose of developing students' capabilities or skills and/or expanding students' understanding—both shared and individual—of a specific concept or instructional goal. Classroom discussions are characterized by high quality and high quantities of student talk. Teachers must ensure that discussions are built upon and revolve around both students' contributions and the content at hand. In a discussion, the teacher's role is to question students, take up, revoice, and press students' ideas, structure and steer the conversation toward the learning goal(s), enable students to respond to one another's ideas by stepping back to listen, moderate and facilitate students' interactions, ensure that the content under discussion is represented accurately, and bring the discussion to a meaningful close.

The research literature on classroom discussion spans multiple grade levels and content areas. In what follows, we describe features of discussion that have been identified in elementary through high school classrooms.

Rationale for the Construct

While other forms of classroom-based discourse (e.g., recitation or debating) also are valuable, many scholars now call for greater use of discussion to supplement and complement other discursive practices in the classroom (see for example Resnick et al., 2015).

Considerable commentary in teaching today supports the contention that there are many positive student outcomes associated with discussion-based teaching. Hypothesized benefits of discussion include (a) higher-order thinking and reasoning skills (Cazden & Beck, 2003; Resnick et al., 2010; Sun et al., 2015; Webb, Franke, Turrou, & Ing, 2015); (b) reading comprehension (Applebee, Langer, Nystrand, & Gamoran, 2003; Brown, Pressley, Van Meter, & Schuder, 1996; Eeds & Wells, 1989; Kucan & Beck, 1997; McKeown, Beck, & Blake, 2009; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009; Nystrand, 2006; Soter et al., 2008; Wolf,

Crosson, & Resnick, 2006); (c) overall engagement (Henning, 2005; Nystrand, 1997; Resnick et al., 2010); and (d) collaboration and communication skills (Fogo, 2014; Gall & Gillett, 1980; Hadjioannou, 2007; Larson, 1996; Walshaw & Anthony, 2008).

Continuing in this vein, scholars propose other benefits as well. These include fostering students' abilities to understand and appreciate multiple points of view, analyze a wide range of issues, formulate their own positions on those issues, and make and defend arguments for those positions (Fogo, 2014; Gall & Gillett, 1980; Hadjioannou, 2007; Larson, 1996; Nystrand et al., 2003; Parker, 2006; Walshaw & Anthony, 2008). Discussions present unique opportunities for students to hone reasoning and collaboration skills while deepening their abilities to participate meaningfully in communities of disciplinary discourse and inquiry (Brown et al., 1996; Resnick et al., 2010; Sun et al., 2015).

Another form of grounding for discussion as an instructional strategy rises out of theoretical propositions in the cognitive and learning sciences. Constructivist learning principles support the importance of discussion as a means of promoting cognitive and metacognitive skills and of acquiring critical disciplinary practices. In mathematics for example, students learn how to explore mathematical ideas, make and test conjectures, and learn to use the formal language of the discipline when they are engaged in discussions (Smith, Hughes, Engle, & Stein, 2009; Walshaw & Anthony, 2008). Likewise in reading or ELA, students learn how to interpret texts, analyze arguments, explore concepts and big ideas, separate fact from opinion, and develop their own views when they undertake these practices in conjunction with other students (Applebee et al., 2003; Eeds & Wells, 1989; Nystrand, 1997).

Discussion-based teaching is also associated with an important normative argument about learning to participate in the deliberations essential to the democratic way of life (Nystrand, 1997; Reisman, 2015; Resnick et al., 2010). Citizens, this argument asserts, must become well informed and learn to use their voice to express opinions, evaluate evidence and argument, and contribute to and participate in public discourse (Fogo, 2014; Gall & Gillett, 1980; Hadjioannou, 2007; Larson, 1996; Walshaw & Anthony, 2008). Learning these practices requires that students be more than passive recipients of knowledge. Rather they must be actively engaged in creating, testing, and evaluating knowledge in the company of other students, where they also are learning how to resolve disagreements, collaborate in constructing knowledge, and sharpen their views in interchanges with the views of others. Communities of learners employ discourse to make sense

of the important matters under study, and good teachers create opportunities in the classroom for these experiences (Michaels, O'Connor, & Resnick, 2007; Resnick et al., 2010). At the same time, as we describe next, making use of discussion to promote learning is a complex practice that requires teachers to carefully structure and monitor student talk in order to be productive for ambitious learning goals.

Discussion then joins with other teaching methods aimed at helping students understand academic content. As an instructional practice, it serves as one strategy for promoting understanding of key academic content. But it also serves as an end in itself as a student learning practice that involves both general and subject-specific skills. Without opportunity for discussion and deliberation, many scholars now argue, these goals cannot be achieved through other means.

Explication of the Construct

The LCD construct involves a number of key aspects that we take up in turn:

- structuring

- questioning

- centrality of student ideas

    - uptake

    - revoicing

    - press

- student-to-student interaction

- subject matter and instruction

- discussion and classroom environment

These features are hypothesized in the research and scholarly literature as contributing to effective classroom discussions aimed at building discussion-related skills, promoting student understanding of academic content, and participating in disciplinary discourse communities.

Structuring. A meta-analysis of teaching practices suggests that structuring a lesson is an effective teaching strategy that leads to positive student outcomes (Kyriakides, Christoforou, & Charalambous, 2013). In this context, structuring refers to the teacher's role in setting up the discussion, guiding the discussion toward a set of learning goals, and concluding the discussion.

Thus, classroom discussions require teachers to have an understanding of where the discussion is headed and how to get there (Kazemi & Hintz, 2015). At the outset of a discussion, teachers should articulate a clear purpose for their students (Protocol for Language Arts Teaching Observation, 2013b; Stein, Engle, Smith, & Hughes, 2008; Stein & Smith, 2011). Once teachers have focused students on a specific learning goal or goals, they must then maintain students' attention on that goal throughout the discussion (Kyriakides et al., 2013; McKeown & Beck, 2015; Wilkinson, Murphy, & Binici, 2015). The teacher's role is to coordinate the work so that students can develop their ideas while maintaining a thematic focus (Goldenberg, 1992) that helps to organize their progress.

At the end of a discussion, it is also important to summarize the content and ideas discussed (Kyriakides et al., 2013). Research on teaching has long documented the importance of lesson closure. For example, Schoenfeld (1983) described *postmortem analysis* as a time when, after students have completed individual or group work, the students or teacher describe the process, method, or strategy just used or discussed. Reviewing content upon closing a discussion reinforces the learning taking place (Good & Brophy, 1986) and it focuses students' attention on the most important highlights from the lesson or discussion (Collins, Brown, & Newman, 1989). Lesson closure is also reflected in the professional teaching literature (e.g., Duncan & Clemons, 2012; Webster, Connolly, & Schempp, 2009; Wong, 1990) where practicing teachers often point to the importance of summarizing content covered and ideas discussed for students.

Questioning. Leading a classroom discussion requires several particular features of teacher questioning together with how they follow up on student responses. Though supported primarily in the theoretical and descriptive literature on these practices, questioning also has been identified in Kyriakides et al.'s (2013) meta-analysis, positively linking teachers' use of questioning techniques to student achievement. It is the role of the teacher to prompt students' participation, and questioning is one of the most widely used teaching strategies for doing so. Teachers assume responsibility for asking questions that elicit multiple students' thinking around a learning goal or goals (Kazemi & Hintz, 2015). Discussions, in particular, are characterized by teachers' use of authentic and open-ended questions to both ignite and sustain discussions (Cazden & Beck, 2003; Kersaint, 2015; McKeown et al., 2009; Nystrand et al., 2003; Soter et al., 2008). These are questions that have more than one possible response and, importantly, questions for which the possible responses are not known or prefigured (Applebee et al., 2003;

Hadjioannou, 2007; Nystrand & Gamoran, 1991). For further review of how teachers elicit student thinking in various formats including classroom discussion, see Qi, Sykes, and Croft (2016).

Teachers strike a balance between asking questions that allow for meaningful exploration and construction of new knowledge and not going so far outside of students' existing knowledge base that the students have nothing to say (Cazden & Beck, 2003; Kyriakides et al., 2013). Teachers make in-the-moment judgments about when to ask another question and when to allow for wait time (Cazden & Beck, 2003; Wolf et al., 2006) to give students a chance to consider what has been said and to formulate their responses. Further, when teachers are instructed in the use of wait time, research has uncovered a cascade effect, prompting other behaviors that enhance discussion, such as less teacher talk and fewer low-level questions (Russ, Sherin, & Sherin, 2016; Tobin, 1987). Kyriakides et al. (2013) argued that the length of the pause should depend on the type of question being asked. Although empirical studies have begun to document the importance of questioning for teaching practice, for the most part, small "n" case studies, observational studies, and theoretical pieces support these contentions in the context of discussion (see Dillon, 1990, for review).

Another goal of questioning is to move the discussion along (Henning, 2005), so questions should be strategic instead of formulaic (Wolf et al., 2006). Over time, and with practice engaging in strategic discussions, students can begin to ask questions of the teacher and one another that demonstrate and deepen understanding (Applebee et al., 2003; Langer, 2001; Langer & Close, 2001).

Centrality of student ideas. When leading a discussion, teachers affect a balance between structuring the discussion around targeted learning goals and allowing students to voice their opinions, ideas, and strategies around those targeted learning goals. In fact, a primary concern when leading a discussion is the degree to which teachers build the discussion on student ideas. Whereas teachers may conduct discussions with groups of varying sizes (Boerst, Sleep, Ball, & Bass, 2011; Hess & Posselt, 2002; Langer & Close, 2001; Reisman, 2015), in a discussion, teachers ensure that all student voices are heard (Kersaint, 2015; Walshaw & Anthony, 2008) and allow students to do most of the talking (Hess, 2004). This feature, perhaps most of all, distinguishes classroom discussion from the more common mode of lecture-based, or recitation-style, instruction in which teachers "deliver" content to students as opposed to

supporting students as they construct their understanding of it. Teachers must talk less (Boerst et al., 2011; Wolf et al., 2006) and be willing to cede some control of the discourse to students (Goldenberg & Patthey-Chavez, 1995; Nystrand, 1997). What lends credence to this feature are research syntheses on learning that emphasize the import of students' prior knowledge and how teachers must construct conceptual bridges from the academic content to students' initial understandings and misunderstandings (Brown et al., 1996; Sfard, 2015). Absent dialogue in which teachers probe for students' understandings, the basis for connecting content to students is seriously compromised.

Students' ideas then are the basis upon which productive discussions are built. Several researchers have asserted that discussions should be characterized by free exchanges (Applebee et al., 2003; Nystrand & Gamoran, 1991; Nystrand et al., 2003). By this, they mean that while the teacher identifies a focus and goals for the discussion, he or she does not predetermine or prescribe what will be said during the discussion (Wilkinson et al., 2015). Instead, the discussion will evolve, as discussions outside of the classroom do, as the participants respond to one another's contributions. In this way, students' ideas become the foundation for building and extending both individual and collective understanding of the topic or skill under consideration (Boerst et al., 2011; Michaels et al., 2007; Okolo, Ferretti, & MacArthur, 2007; Parker & Hess, 2001; Resnick et al., 2010).

In order to gain access to and maintain focus on student ideas, the literature describes a number of specific practices that teachers employ, particularly uptake, revoicing, and press.

Uptake. One practice along these lines is described as linking talk (Wolf et al., 2006) or uptake (Cazden & Beck, 2003; Nystrand, 1997) to maintain the emphasis on student ideas as the basis of discussion. *Uptake* refers to instances in which teachers "take up" student ideas and use them to spur further questioning or contributions to advance the discussion (McKeown et al., 2009; Nystrand & Gamoran, 1991; Reisman, 2015; Soter et al., 2008). For example, a student may ask a question or share an idea that the teacher then uses to prompt the next series of exchanges (e.g., by asking the class to respond to the student's question or idea directly). Alternatively, a student could suggest a solution that the teacher then puts on the board for other students to evaluate and debate. Teachers may also encourage students to engage in uptake of each other's ideas (Applebee et al., 2003; Boerst et al., 2011) as a way of giving them more control over the progress of the discussion.

Revoicing. Closely related to uptake is the practice of *revoicing* student contributions—that is, restating or paraphrasing what a student has said in order to "highlight significant contributions and to shape the discussion productively" (Walshaw & Anthony, 2008, p. 530). Teachers may use revoicing to encourage others to carefully consider a student's contribution (Cazden & Beck, 2003; Michaels et al., 2007; Reisman, 2015), as well as to help students articulate their ideas in discipline-specific language, making sure students understand and express disciplinary ideas, concepts, and processes accurately (Boerst et al., 2011).

Press. Experienced teachers ask questions in order to probe students' thinking, get students to elaborate on their ideas, and clarify their reasoning (Walshaw & Anthony, 2008); here the concept of *press* enters. Teachers employ strategies that press students to fully elaborate their thinking and logic (Resnick et al., 2010; Wolf et al., 2006) by asking questions or making statements designed to get students to say more about why they made a given statement or took the position that they did (Goldenberg, 1992; Lemov, 2010). Press also includes the practice of prompting students to justify their contributions by offering corroborating examples or evidence (Reisman, 2015; Walshaw & Anthony, 2008). In addition, Michaels et al. (2007) and Resnick et al. (2010) suggested that teachers press students to develop challenges and counterexamples to their ideas as a way of further developing and clarifying their thinking. Pressing students to further explicate their reasoning challenges them at a higher level of cognitive demand (McKeown & Beck, 2015; Stein, Engle, Smith, & Hughes, 2015; Walshaw & Anthony, 2008; Wilkinson et al., 2015). Students who are engaged in explaining their thinking to others are doing more thoughtful work than those who are merely called upon to recite answers from memory (Nystrand & Gamoran, 1991). Support for these instructional moves resides to a degree in the goals for discussion and the basic meaning of this discursive practice.

Student-to-student interaction. Another hallmark of discussion-based instruction identifies how teachers encourage and maintain student-to-student interaction. In fact, a key marker of classroom discussion is the degree to which students interact with one another. Once again, leading a classroom discussion involves a balancing act as teachers must ask specific, targeted questions and also provide opportunities for students to comment on and ask questions about one another's ideas. In contrast with the teacher-centered discourse pattern of initiation–response–evaluation (IRE), the teacher must help students to go beyond simply responding to teacher prompts and ensure that they interact directly with one another to successfully build on

and extend collective understanding (Applebee et al., 2003; Boerst et al., 2011; Gall & Gillett, 1980; Nystrand et al., 2003). This feature of LCD is important because interactions among students provide them with practice in engaging in the kinds of deliberative activity that constitute an essential element of democratic participation. Students learn to use their voice in such deliberations and to listen carefully and respectfully to the voices of other students (Nystrand et al., 2003; Parker, 2006; Resnick et al., 2010).

As well, when students engage with others, this process enhances the learning of all students, where such learning is associated both with the individual student and with the class as a whole (Nystrand et al., 2003). Knowledge and understanding develop through rich interactions carefully mediated by the teacher (Resnick et al., 2010). Note though that mere interaction among students absent the mediating role of the teacher is unlikely to promote learning (Franke et al., 2015). Consequently, as teachers encourage student-to-student interchanges, they draw on the other practices described here to guide the discussion, build understanding, close off unproductive or distracting sidelights, and keep the focus on the learning objectives. The studies cited here are of value in presenting examples of exemplary practices often contrasted with poor performances that fail to exemplify the desired practices. The employment of student-to-student interaction is properly regarded then as a defining feature of this construct insofar as teachers are expected to engage all students during instruction in a manner that encourages interactions among students.

Subject matter and discussion. Although key aspects of discussion are described in a generic way for the purposes of this paper, the subject matter under discussion is a feature of every classroom's context that is inherently embedded in the construct. Each discipline uses discussion in ways that draw on the subject matter together with aspects of discussion that are distinctive to disciplinary ways of knowing. Classroom discussion is proposed as a crucial means to teach students how to begin speaking and thinking like disciplinary experts (Resnick et al., 2010; Smith et al., 2009). While aspects of discussion cut across content areas and grade levels, implementation of these skills requires teachers' pedagogical understanding of the subject matter being discussed.

The purposes, tools that are used (e.g., problem-solving strategies and written texts), and structures for how these tools are used vary depending on the subject matter. While generic descriptions of critical features of classroom discussions may be identified, discussions occur in

content-specific contexts that cannot be disentangled from the practice of discussion. Effective discussion leadership relies on teachers' knowledge of the subject matter and of the content knowledge for teaching that serves as the disciplinary basis for instructional effectiveness with this practice (Ball, Hill, & Bass, 2005; Grossman, 1992; Grossman, Schoenfeld, & Lee, 2005; Wilson & Wineburg, 1988).

For example, a teacher might have a strong discussion skill set yet lack content knowledge of the subject under discussion. One can imagine that the resulting discussion would fall short of some of the important goals for the lesson. Teachers could not build a mathematical discussion toward shared understanding of the underlying mathematics without substantial content knowledge. In the discussion of a novel or of a scientific concept, the quality of discussion and its effects on learning will depend on teachers' skillful deployment of general discussion skills and strategies and on their understanding of the content knowledge that is involved.

From one angle, the point is a logical one: that teachers cannot induct students into disciplinary ways of knowing if they themselves do not possess the requisite knowledge and understanding. In practice, general and content-specific knowledge operate interdependently in contributing to the overall quality of discussion and its outcomes in student learning.

Discussion and classroom environment. Classroom discussions do not happen meaningfully unless teachers create a supportive classroom environment that is open to questioning as well as to varying perspectives. Teachers may demonstrate respectful interactions (Chapin, O'Connor, & Anderson, 2003; Wolf et al., 2006), specific participation strategies (Applebee et al., 2003; Palinscar & Brown, 1984), and the kinds of thinking and language they expect students to use (Simon, Erduran, & Osborne, 2006; Soter et al., 2008; Walshaw & Anthony, 2008). Thus, good teachers, together with their students, carefully craft and maintain classroom norms and routines that enable and encourage all students to feel comfortable in participating (Cazden & Beck, 2003; Hadjioannou, 2007; Michaels et al., 2007; Reisman, 2015). Effective teaching relies both on the deployment of specific best practices and on global factors in classrooms that include classroom management and a supportive climate along with how teachers engage students with academic content (Kyriakides et al., 2013; Lemov, 2010). Learning through discussion clearly relies on and benefits from these holistic features of instruction.

In summary, meaningful classroom discussions, the scholarly literature indicates, should be built around both students' ideas and the learning goals at hand. The teacher's role is critical for ensuring that this discussion occurs. What the teacher does to set up and support the discussion requires engagement in specific practices. Teachers press students to elaborate their ideas, take up a select few of those ideas, revoice students' contributions and encourage student-to-student interaction, guide students toward accomplishing the learning goal(s) at hand, and conclude with a summary of the discussion that took place. These practices are intertwined with the teacher's grasp of the subject matter under discussion and the content knowledge related to how to teach that subject matter to particular students.

Measurement of Discussion in the Research Literature

While richly theorized for its value in producing a range of learning outcomes, the empirical work connecting the instructional practice of discussion to such outcomes is just beginning to accumulate (Sfard, 2015; Wilkinson et al., 2015). Classroom discussion has been measured in a variety of ways in the research literature, depending on the theoretical interests and research questions pursued by investigators. The evidence to support claims is diverse. Some studies are theoretical (Cazden & Beck, 2003; Gall & Gillett, 1980) whereas others rely on observational analyses often guided by theoretical propositions and hypotheses. Most studies define discussion generally but do not examine the effects of its components on student outcomes. We next organize exposition of this literature around the outcome measures used in the studies, describing the methods and measures that have been used together with the results.

Discussion and Reasoning Skills

Several observational studies buttressed by logical claims suggest that teachers use discussion in the classroom to help students build discussion skills, such as learning to listen to others, evaluating arguments, and formulating and communicating their own views (Fogo, 2014; Gall & Gillett, 1980; Hadjioannou, 2007; Larson, 1996; Walshaw & Anthony, 2008).

Some research along these lines has demonstrated that when teachers employ discussion strategies effectively, such discourse helps students learn to reason. In one observational study (without a control group) of six classrooms using a program called Collaborative Reasoning, students who engaged in small group discussions twice each week for 5 weeks saw growth in analogical reasoning skills (Sun et al., 2015). Another observational study of three classrooms

demonstrated that students' explanations vary based on the degree to which teachers ask students to justify their thinking (Webb et al., 2015). Arguably, as students discuss, they gain insights into alternative solutions and are exposed to novel reasoning strategies generated by their peers (Resnick et al., 2010). In this way, theorists propose that discussion helps to socialize students into discourse communities that are based on repeated opportunities to engage in higher-order thinking (Cazden & Beck, 2003).

Studies of this kind tend to use qualitative methods where researchers either observe teaching in real classrooms (Hadjioannou, 2007; Webb et al., 2015; Wolf et al., 2006) or study teacher self-report data through surveys and interviews (Fogo, 2014; Larson, 1996), thereby measuring discussion via observation protocols, interviews, and surveys.

Academic Achievement

Participation in classroom discussion may positively impact students' academic achievement. In an observational study comparing high- and low-achieving middle and high school ELA classrooms in 25 schools across four states over 2 years, Langer (2001) found that schools with higher scores on standardized achievement tests tended to use more classroom discussion than lower performing schools. And in a quasi-experimental study, low-achieving second graders who regularly participated in a discussion-based intervention scored higher on standardized tests than their matched pair control groups who did not participate in the discussion program (Brown et al., 1996). In another study of 58 ELA and 57 social studies classes across 2 years in 16 midwestern schools, students in high-track classes tended to engage in discussion more often than low-track classes (Nystrand et al., 2003), suggesting that this instructional practice may be afforded inequitably across tracked classrooms. This latter finding underscores the equity implications for this practice.

Most of these studies use more correlational approaches to describe the ways in which specific discussion approaches are related to higher student achievement, most often in literacy (Applebee et al., 2003; Brown et al., 1996; Nystrand & Gamoran, 1991). Applebee et al.'s (2003) observational study of 64 classes in 19 schools used hierarchical linear models and found students who engaged in dialogic instruction and extended curricular conversations performed higher on literacy assessments. Sun et al.'s (2015) study used achievement measures to correlate observational findings with student outcomes.

These studies measure discussion in a variety of ways, as embedded in observation protocols and paper-and-pencil surveys and interviews of teacher practices correlated with outcomes on student achievement tests or characteristics of students and classrooms. For example, Nystrand et al. (2003) used the Classroom Assessment Scoring System (noted previously) to examine the dialogue between teacher and students and correlated qualitative findings from the observation protocol with socioeconomic status and type of classroom (high and low track). Brown et al. (1996) correlated students' participation in a discussion-based teaching program for reading with students' scores on a standardized reading test.

Reading Comprehension

In their extensive review of research on the teaching of literacy, Purcell-Gates, Duke, and Stouffer (2016) asserted, "There is considerable evidence supporting discussion as a means of improving reading comprehension" (p. 1,228). Several authors have documented a positive relationship between ELA classroom discussion and students' reading comprehension (Applebee et al., 2003; Brown et al., 1996; Eeds & Wells, 1989; Kucan & Beck, 1997; McKeown et al., 2009; Murphy et al., 2009; Nystrand, 2006; Soter et al., 2008; Wolf et al., 2006). In a pilot study of a reading intervention with 37 students, low-achieving middle school students in classrooms that utilized certain discussion techniques saw greater gains and more consistent growth in reading comprehension compared to randomly assigned second treatment and control classrooms (Palinscar & Brown, 1984). In this study, student learning was measured by students' responses to a series of reading comprehension questions: 10 questions for each of 13 reading passages. Based on observational studies of classroom discussions, one reason for this may be that providing students with opportunities to engage with their peers around significant content gives them more exposure to and practice in using academic language (Resnick et al., 2010).

For example, fifth-grade students engaged in a reading program, which included discussion as one component of the intervention, saw greater growth than students randomly assigned to the control group (McKeown et al., 2009). However, it is difficult to disentangle the effects of discussion from the other components of the reading intervention. In another study, Saunders and Goldenberg (1999) found that out of 116 students, fourth and fifth graders in classrooms randomly assigned to engage in discussions around particular readings had higher average reading comprehension scores than similar students in the control group that did not engage in discussion.

Disciplinary Knowledge and Practices

In a discussion, teachers help students learn to be accountable to the classroom community. Some authors argue that this focus on mutual accountability leads to content learning because it supports the development of shared understandings that are richer than those that students would arrive at on their own (Gall & Gillett, 1980; Jansen, 2006; Langer, 2001; Michaels et al., 2007; Nystrand, 1997; Okolo et al., 2007; Parker, 2006; Smith et al., 2009; Van den Branden, 2000).

A central argument in the literature on discussion-based teaching posits that this practice assists in inducting students into disciplinary communities and ways of knowing. This theme is prominent both in mathematics (Walshaw & Anthony, 2008) and in literacy (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Shanahan & Shanahan, 2008). Such induction includes the transmission of knowledge while also supplying larger conceptual structures together with ways of acquiring and evaluating knowledge in particular disciplines. Harkening back to Schwab (1978, p. 246), through discussion, teachers engage students in the *substantive* and *syntactic* structures of the disciplines, which he defined as the conceptual devices that are used for defining, bounding, and analyzing subject matter (substantive) and the different methods of verification and justification of conclusions as constituting the logical structure of a discipline (syntactic). Attending to these features of a discipline is a complex learning objective, difficult to measure, even as it is prominently represented in the theoretical literature on uses of discussion (see for example, Cazden & Beck, 2003; Fogo, 2014; Gall & Gillett, 1980).

In mathematics, there is also a tradition of case study research that provides detailed description of how a teacher builds disciplinary knowledge while enacting disciplinary community in the classroom (Ball, 1993a, 1993b; Lampert, 1986, 1990, 2001). Unlike many case studies that employ external observers, this line of work involves the scholar-teacher in studying and reporting on his or her own practice. Such studies portray what instruction of this kind looks like, contributing explorations of the demands, dilemmas, and promise that are involved in mathematical discussions in the elementary classroom. The measurement strategy involves use of videotape, postlesson reflections, examination of student work samples, and interviews with individual students and others to (re)construct the flow of mathematical discourse together with its outcomes in learning. These investigators propose that students learn how to think mathematically, to grasp the conceptual structure of the discipline, and to apply their

understanding to the solution of problems that the teacher poses, designed to elicit student solutions that may be novel as well as canonical. These investigations concentrate on explicating the practice rather than on linking the practice to outcomes of various kinds.

Engagement in Learning

Several authors have also used the concept of student engagement as an outcome, based on the hypothesis that improved engagement is likely to increase students' depth of understanding of the content under discussion (Henning, 2005; Nystrand, 1997). While the concept of engagement has been broadly defined in the literature (see Cooper, 2014; Dolezal, Welsh, Pressley, & Vincent, 2003; Marks, 2000), this line of work defines the concept more particularly in terms of the academic substance of the lessons (Henning, 2005; Nystrand, 1997; Resnick et al., 2010). For example, Nystrand and Gamoran (1991) attempted to directly link student achievement with engagement. They analyzed student and teacher questionnaires and classroom observations from 58 eighth-grade English classes for evidence of student engagement and then compared those findings with students' performance on achievement tests. They found that students in classrooms characterized by high levels of substantive engagement (e.g., participation in classroom discourse) had higher levels of achievement, as measured by scores on a literature test, than students in classrooms characterized by low levels of substantive engagement.

Discussion and Other Outcomes

As we have indicated, discussion-based instruction has been commended in relation to citizens' participation in a democratic society. Educators and researchers have long argued that classroom discussion supports students in developing the habits and skills necessary for such participation (Resnick et al., 2010). Scholars have also advanced theoretical arguments about the importance of teaching students to become better citizens, build community, and be more informed in community decision making (Harris, 2002; Hess & Posselt, 2002; Okolo et al., 2007; Parker, 2006; Tannebaum, 2014). These outcomes, it is theorized, can be achieved when students develop a voice in their classroom through opportunities to state and support arguments, debate with peers, and develop proofs and theories with others.

Sfard (2015) proposed another democratic argument in favor of classroom discussions. At a time when discussions can occur freely over social media, blogs, and online discussion

forums, and online schooling is becoming increasingly popular, she argued that "if the school is to withstand the competition, the monologic tradition of the classroom must give way to more diverse, more inclusive forms of exchange, closer in character to those that can be found in virtual meeting places" (Sfard, 2015, p. 247). Incorporating discussion into the classroom would help enable the K–12 school model to offer the interpersonal communication component of learning that online environments cannot. Together with modest, nonexperimental findings, theory-based reasons support the role of discussions as offering a practice arena for learning to participate in deliberations essential to the democratic way of life.

These broad goals also enjoy support in sociocultural theories of teaching and learning. Building on the work of Vygotsky (1978) and Bakhtin (1981, 1986), researchers argue that discussion promotes students' assimilation of cultural discourse norms (Nystrand, 1997; Reisman, 2015), helps them learn to engage with and learn from those with differing viewpoints (Nystrand et al., 2003; Parker, 2006), and mediates students' content learning through interactions with others (Eeds & Wells, 1989; Reisman, 2015).

Finally, some scholars argue that well-implemented discussion extends benefits to all students, thereby supporting equity goals. In particular, nonnative English speakers may benefit because well-orchestrated, open classroom discussions allow all students to participate in the development of shared understandings of content more fully than a traditional lecture format would (Saunders & Goldenberg, 1999). And, in their review of first language acquisition, Faltis and Valdes (2016) listed the following proposition as foundational: "Conversation is at the core of all language development. Conversation involves receptive and productive skills" (p. 569). As some studies, reviewed previously, have shown, discussion-based teaching with all of its purported advantages, may be unequally distributed to children from different backgrounds, underscoring the imperative to extend this practice broadly and equitably.

Issues in Measuring Discussion

The scholarly literature on discussion-based teaching is theoretically and normatively rich. The arguments in favor of this practice now enjoy broad acceptance, and the practice is encoded both in standards for learning and for teaching as reviewed previously. Measurement of this practice and of its hypothesized dimensions is challenging and several measurement issues stand out.

First, evidence suggests that classroom instruction can be scored reliably and accurately. In some studies, rater accuracy was confirmed by the use of high-level field note observations (e.g., Eeds & Wells, 1989; Hadjioannou, 2007), and in others, the use of specific protocols and highly trained raters promoted accurate scoring (e.g., Nystrand & Gamoran, 1991; Nystrand et al., 2003). Note that Nystrand et al. (2003) did not report rater reliabilities, and in Nystrand and Gamoran's (1991) study, raters reconciled all rating disagreements with lead raters. In 2006, Wolf et al. reported that when two raters double-scored a lesson, the rater reliabilities (Cronbach's alpha) generally ranged from .89 to .93, depending on which aspect of discussion was being measured (Wolf et al., 2006).

Second, in the majority of studies cited, measurement does not model the effects of particular components of instruction or measured, such as lesson closure or press. By and large, these studies employ more global measures of discussion, often compared to instruction that makes less use of discussion. A small set of studies does explore particular features of LCD. The Nystrand and Gamoran (1991) study, cited previously, found authentic questioning and uptake were features of discussion that increase student engagement, which has a positive impact on student achievement. But the main body of descriptive and correlational study of discussion does not tease out the effects of particular aspects or dimensions of discussion. Practices such as uptake or revoicing are theorized to be important (e.g., Cazden & Beck, 2003; Resnick et al., 2010), and studies provide descriptions of such practices (e.g., Hadjioannou, 2007), but their association with outcomes has not been modeled in ways that are generalizable or replicable.

A third point is that studies providing evidence concerning discussion simultaneously measure other factors that combine to produce learning. Discussion might well play an important role but cannot be disentangled from other aspects of classroom functioning. In both the general and subject-specific observation protocols described previously, teaching performance is scored holistically on various scales with descriptors such as instructional dialogue or classroom discourse, but these factors are combined with others to yield overall scores of teaching quality.

Finally, additional descriptive studies suggest that contextual factors shape observational scores created from samples of typical practice. Evidence here suggests that the students in the classroom and the content of instruction influence the overall observation scores a teacher earns (Qi, Bell, & Gitomer, 2014; Whitehurst, Chingos, & Lindquist, 2014). Teachers may not be equally adept at teaching students from diverse backgrounds or teaching different topics in

mathematics. These studies are not conducted with instruments that isolate discussion, so it is unclear if the findings will pertain to the measurement of discussion. But they do raise the point that when typical classroom practice is measured using standardized observation protocols, there is some risk of ascribing assessments of classroom quality to the teacher when, in fact, contextual features (i.e., students and topics) are influencing classroom quality scores. This presents important potential challenges to measuring discussion in a high stakes way in naturalistic environments.

These points concerning challenges in the measurement of discussion-based teaching serve to provide the backdrop for the measurement approach adopted for the NOTE assessment series that seeks to address some of these issues. Clearly as measures are standardized, the influence of context is reduced and this is a necessary tradeoff for a high stakes test. We turn next to how the LCD construct is being measured in NOTE.

Measurement Approach for the LCD Performance Assessment

In this section, we describe the delivery mode, content, structural features and scoring rubric used in the LCD performance assessment along with rationales for our design choices.

Delivery Mode

For the LCD performance assessment, candidates lead a discussion with five student avatars in a virtual classroom. The virtual classrooms are viewed by candidates on a computer screen, and candidates can engage with the avatars in real time. All five student avatars are controlled by one human simulation specialist. Through a computer-based interface, the simulation specialist can see and hear the candidates. Through that interface, the simulation specialist enables each student avatar to share ideas, ask questions, and respond to questions posed by the candidate and other student avatars. Candidates are able to represent written content and record ideas from the discussion using a shared electronic workspace that both the teacher and student avatars can write on. The shared workspace simulates a smart board, easel, individual or shared white board, or other tools teachers use daily in classrooms to record and represent the content of discussions. In this manner, the tasks and the simulation approximate an authentic performance to the greatest degree practicable on a high stakes test.

Simulated classroom setting. Our use of an avatar-based mode of performance assessment was chosen for several reasons. First, the simulated classroom environment promotes

standardization, which, "has been a fundamental principle for assuring that all examinees have the same opportunity to demonstrate their standing on the construct that the test is intended to measure" (AERA, APA, & NCME, 2014, p. 51). As described previously, existing classroom observational measures of discussion are conducted in real classrooms and consequently have multiple sources of measurement error partially due to uncontrolled student characteristics as well as sampling teaching over time. In the case of uncontrolled student characteristics, some candidates might have an easier time carrying out the discussion because the students are very familiar with discussion. This cannot be standardized. There are many more examples of how student characteristics might make carrying out a discussion task easier or harder. Each one introduces concerns about standardization. Standardization is necessary for fair and accurate scoring, which in turn is particularly crucial in a high stakes licensure decision.

Second, conventional assessments (e.g., multiple choice questions) can capture knowledge related to effective discussion leadership but are inauthentic. Modeling authenticity is important because the construct of discussion involves management of interactions with and among students around content. Discussion leadership is inherently interactive, calling on teachers to manage what they say in response to what students are saying in relation to important content. For licensure, it is useful to know if a candidate can produce a performance and not just evidence of knowledge related to a performance so that we can more authentically assess skills that are part of the job analysis. Using a simulated classroom environment allows teacher candidates to interact with students as if they were conducting a discussion with real students.

Third, this approach has advantages relative to other approaches to measuring the construct while paving the way for steady improvements in the technology that will achieve even greater fidelity to the construct and authenticity in the performance. We mean here that technology-enhanced assessments of this kind are a growing development in the field that will be improved steadily as the technology is tried, studied, and refined. The advantage then refers to not only its present but also its future value. The field currently lacks methodologies for measuring performances for both formative and summative purposes. The approach taken here fills this gap in our assessment arsenal.

Finally, introducing this mode of assessment may help to prompt use in preparation and so constitute a useful stimulus for the development of the requisite skills, where practice over repeated trials with feedback is a well-established method for building skills and improving

performance. We anticipate then that as the field gains familiarity with simulations of this kind, their use will increase as a training tool for summative decisions of various kinds (e.g., course and program completion, licensure, certification).

Five student avatars. All five student avatars are controlled by one human simulation specialist who is able to see and hear the candidate in real time. Simulation specialists are trained and certified to (a) understand and respond to the task and questions as would real students at that particular grade level, (b) act and speak like children at the grade level of each task, (c) ensure all candidates receive similar opportunities to demonstrate performance, and (d) use the required technology to control student avatars' movements and gestures. Certification tests must be passed before interacting with teacher candidates in a consequential assessment situation, and simulation specialists must also pass quality control checks periodically after they are certified.

The use of five student avatars is meant to approximate a classroom discussion. This choice was made given the available technology, cost of the assessment (passed to the test taker), and cognitive demands on simulation specialists. In addition, we believe five students constitutes a reasonable number of students for a candidate to work with. Our claim is that if a candidate cannot orchestrate interactions with five students, he or she is unlikely to do so with larger numbers of students. Further, teachers typically divide classes into smaller groups for a variety of purposes (Cohen & Lotan, 2014). Sometimes students pair up, sometimes students work in small groups, and sometimes the group is divided in half. The upshot is that teachers do enact small group forms of instruction where they interact with student groups of this size, often with relatively short duration, so the task reflects one common instructional format, and this is a naturally occurring unit for teachers to work with.

Discussion Content

Each discussion or task is centered around critical content from the *Common Core State Standards: ELA* (NGA & CCSSO, 2010a) and *Common Core State Standards: Mathematics* (NGA & CCSSO, 2010a) in Grades 1 through 5 that has been vetted by public and professional groups and reviewed by content experts (a sample task is included the appendix). Specific content was selected based on its importance and prevalence in ELA or mathematics across the K–6 curriculum, its centrality to students' learning at a particular grade level, appropriateness for a licensure test, and suitability for discussion in this delivery mode. For example, the type of mathematics discussion for this assessment is limited to collecting and analyzing solutions to a

problem with a focus on themes/differences across solutions and/or sequencing and building an argument or solution. Reading and writing content was selected over language, speaking and listening, and grammar for ELA tasks because it is more accessible for new teachers to learn.

The purpose of the assessment is not to collect evidence to support the claim that performance on the particular content predicts performance across the relevant universe of content. Rather, the purpose of the assessment is to gather evidence in support of the claim that if a candidate fails to perform well on this content, he or she will be less likely to perform well with similarly critical content.

Standardizing the content area ensures that the demands of the task for teacher candidates remain as consistent as possible across tasks, thereby addressing one aspect of fairness called for in the standards (AERA, APA, & NCME, 2014). Additionally it allows test administrators to hold much of the task format the same across items, reducing the time it will take to design new items.

Structural Features

Candidates have 30 minutes to prepare for the discussion and 15 minutes to facilitate and conclude the discussion. For each LCD performance task, teacher candidates are asked to complete a discussion that serves as the second half to a lesson. Each task describes the learning and activity that took place immediately preceding the discussion portion of the lesson, provides instructional materials as needed (e.g., student work samples, texts), and defines student learning goals that serve as the purpose and focus for the discussion. In order to assess candidates' discussion-leading skills, we provide information that allows candidates to assume students have had initial experience with the content such that the teacher can launch a discussion without having to provide extensive set up. The lesson, for example, is not focused on helping students to read the story or to acquire the mathematical procedures associated with a task. Rather, the scenario is focused on leading a discussion with a common floor of knowledge among the students already established.

Teachers certainly teach the basic skills that students need in order to participate meaningfully in discussion. Here, though, we assess whether a candidate, with these skills established, can then lead a discussion. These bounds are necessary for practical reasons, such that the performance is not too long, which would introduce greater complexity in scoring, place greater cognitive demands on the simulation specialists, and add to costs and administration of

the assessment. Instructions to candidates indicate what the students have already covered and already know about the task, standardizing this aspect of interaction. Then, the discussion task itself concentrates on critical skills that students are to develop, based on standards for learning.

Each task also includes instructional materials that highlight key ideas and challenges that students face when interacting with the content of the discussion, much like a lesson plan that would appear in a teacher curriculum guide. Teachers constantly make judgments about how to use those curriculum guides with their students, anticipating students' challenges and applying their understanding of pedagogy to the enactment of the lesson or discussion at hand.

Including these content-specific instructional materials not only approximates the materials teachers might receive in a real classroom setting, but it also provides a certain level of content knowledge to all candidates, thereby controlling for the content knowledge each teacher candidate has. While leading discussions requires deep and flexible knowledge about the content under discussion together with the content knowledge for teaching, the NOTE assessment provides substantial attention to this issue in its measures that address content knowledge for teaching. Here, the LCD measurement approach identifies general aspects of effective discussion leadership including how the candidate is able to represent the content during instruction. The claim of the NOTE assessment overall is that these aspects are necessary to discussion leadership, although not fully sufficient in representing all aspects of this construct as it is implemented across an array of contexts.

Scoring Criteria

These aspects map onto the features of LCD identified in the literature as described previously. The final rubric will be posted on the ETS website when completed, following pilot and field trials. Table 1 supplies a crosswalk demonstrating how the features of discussion-based teaching map onto the aspects of the rubric. An assumption underlying this rubric is that these are distinctive, recognizable dimensions of the construct that each contributes to an overall score. The rubric identifies the following subpractices:

- Aspect 1: Prompting participation from all students

- Aspect 2: Steering discussion toward the learning goal(s)

- Aspect 3: Representing the content

- Aspect 4: Concluding the discussion

As described next, each of these aspects includes specific indicators derived from the literature on discussion leadership.

Table 1. Crosswalk Between Features of Discussion-Based Teaching and Aspects of the Leading a Classroom Discussion (LCD) Rubric

Features of discussion	Aspect 1	Aspect 2	Aspect 3	Aspect 4
Structuring		X		X
Questioning & press	X	X		
Centrality of student ideas	X	X	X	
Uptake		X		
Revoicing		X	X	
Student-to-student interaction		X		
Environment[a]				
Accuracy of content			X	

[a]Environment is not included as a formal aspect of our scoring rubric; however, raters will be trained to flag performances where disrespect is observed.

Supporting Evidence for Measurement Approach

Each of the aspects identified in the scoring criteria serves as a critical feature of LCD. In certain respects, the criteria are constitutive of the construct, built into its definition. Although backing for each of these subpractices in terms of their distinctive contribution to effective implementation of the practice—and beyond that—to student learning outcomes will require continuing study, these aspects are hypothesized to be crucial elements for LCD.

Aspect 1: Prompting Participation From All Students

Two features comprise this dimension of LCD. The first involves prompting, while the second references all students. *Prompting* is associated with the empirical and theoretical literature on questioning, press, and the centrality of student ideas (e.g., Kersaint, 2015; Reisman, 2015). As stated earlier, classroom discourse is considered discussion when the dialogue takes up and works with students' ideas. Teachers lead discussion to elicit students' strategies, opinions, interpretations, and inferences by asking a variety of questions and pressing students to explain and defend their reasoning. Prompting then includes both the teacher's moves to encourage initial student contributions and to extend and explore these as the discussion unfolds.

The second feature stipulates that all students are engaged in discussion. There are several rationales for this element. First, as an equity matter, teachers are enjoined to bring all students into discussion, ensuring that learning opportunities are made available to all of the

students in the classroom. A further assumption is that discussion will be enriched to the extent that many contributions are solicited, many student ideas explored, extended, and connected to one another. When discussion orients largely around a subset of students—those who are most vocal, for example, or who are high status—then the contributions of other students are overlooked, thereby impoverishing the discussion and denying opportunity to participate to those students who are left out. This aspect also attends to the teacher's effort to promote interaction among students, calling for teacher moves to engage students with one another's responses so that all students are involved not only in interactions with the teacher but also with one another. A well-implemented discussion then involves all students in mutual and supportive interactions that promote the desired learning.

These aspects of the construct are central to its definition, derive from theoretical propositions from cognitive science, and support its basic rationale, including its normative basis in full participation by all students.

Aspect 2: Steering Discussion Toward the Learning Goal(s)

The second aspect is derived largely from observational studies and theoretical literature. Empirical support for this aspect comes from a meta-analysis of teaching practices, which suggests that both structuring a lesson around a certain learning goal and questioning students on their thinking lead to positive student outcomes (Kyriakides et al., 2013). Here, our argument in including structuring in our scoring rubric is also a logical one, positing that the teacher's role is not simply to encourage participation but also to guide the course of discussion throughout so that dialogue remains focused on the learning goal(s) and on students' emerging understanding in relation to those goals.

Here, too, student ideas appear central, but the teacher must not only solicit student thinking but also keep such thinking moving toward the goals of instruction, as these include understanding both the subject matter and the strategies for building such understanding. When teachers employ specific practices such as uptake and revoicing, they must make on-the-spot decisions about how to connect student input to the intended learning, which requires that they may engage in redirecting, challenging, paraphrasing, or linking one student's comments to another's (Stein et al., 2008). Teachers undertake these instructional moves with an eye on how the discussion is cumulating toward the learning goals involved in a particular unit, lesson, or lesson segment (e.g., Cazden & Beck, 2003; McKeown et al., 2009; Nystrand & Gamoran, 1991;

Walshaw & Anthony, 2008). This critical aspect of discussion—how teachers maintain focus on the learning goals while encouraging student participation—is featured as well in what might be termed the "best practices" literature (e.g., Kazemi & Hintz, 2015; Kersaint, 2015; Langer & Close, 2001).

As described previously, another feature of this element of LCD emphasizes the teacher's role in encouraging student-to-student interaction, so that teachers are also creating opportunities for students to engage with one another (e.g., Applebee et al., 2003; Gall & Gillett, 1980). Through encouragement and facilitation of student-to-student interaction, teachers induct students into cultural norms concerning the nature of disciplinary discourse. Students learn "how to talk" about mathematics or literature or science and, in this manner, extend their understanding of the subject matter, in its substantive and syntactic forms (e.g., Nystrand, 1997; Reisman, 2015; Resnick et al., 2010). Teachers' questions and prompts are evaluated to determine their ability to engage students in rich interactions that model participation in disciplinary discourse while maintaining focus on the substantive learning goals. Such discourse communities involve more than exchanges between students and teachers. Rather, the teacher encourages students to engage with one another around the goals and purposes of the discussion.

Certainly the descriptive and theoretical literature support these features of discussion (e.g., Cazden & Beck, 2003). For the most part, discussion-based teaching is premised on the assumption that teachers must simultaneously facilitate multiple aspects of discussion in order that it be productive. Maintaining focus on the learning goals and encouraging students to engage with one another in accord with disciplinary norms are hypothesized to be critical features of this complex practice.

Aspect 3: Representing the Content

As a logical matter, discussions must involve accurate understanding of the content that is being discussed. If teachers introduce misunderstandings or allow student misunderstandings to be accepted during discussion, this confuses the desired learning. "Content" in this sense refers not simply to facts or propositions but also to conceptual understanding of subject matter, to ways of knowing the subject, to understanding how knowledge is created and tested, and to how members of disciplinary communities engage with one another in appropriate discourse norms (Ball et al., 2005; Selling, 2015). Teachers draw on both common content knowledge and the content knowledge distinctively associated with teaching in managing this aspect of discussion.

The presumption is that teachers with deep understanding of subject matter and of how it is taught and learned are better able to implement discussion in the classroom.

Evaluating the quality of discussion then includes attention to how accurately the subject matter under discussion is represented (Hill, Rowan, & Ball, 2005; Walshaw & Anthony, 2008). In practice this does not necessarily mean that teachers quickly correct student misconceptions, errors, or misunderstandings. Teachers may hold off on such corrections to allow students to challenge one another so that accurate and complete understanding may be an emergent property of a well-conducted discussion, rather than an in-the-moment aspect of just one part of an exchange. Further, where students are encouraged to offer their opinions, predictions, hypotheses, and arguments, an important consideration may be to encourage students to support their views with evidence and argument and to contend with challenges and counterarguments from other students (Franke et al., 2015; Resnick et al., 2010). The instructional role then includes attention not simply to the contentions that students offer but also to the reasoning that supports their contentions, as such reasoning is represented in the disciplines. Finally, because discussion-based teaching aims to encourage interaction among students, teachers may choose to have students challenge, correct, or emend each other, rather than relying on the teacher as sole repository of knowledge (Kersaint, 2015). Judging whether content is fully and accurately represented, then, is a summative judgment rendered across significant samples of discussion within which teachers construct knowledge with students rather than simply conveying knowledge to students.

Aspect 4: Concluding the Discussion

A final, critical aspect involves how teachers bring discussion to a close (Schoenfeld, 1983). This phase may involve a number of moves that include providing a summary (or having one or more students do so), connecting what was discussed to the learning goals, pointing to unresolved questions or issues that require attention in subsequent lessons, highlighting main areas of agreement and disagreement as a basis for further discussion and inquiry, noting questions that emerged during the discussion that appear worthwhile for follow up, drawing attention to valuable contributions by members of the class, underscoring the main or big ideas that may have been obscured in the details, and others (Duncan & Clemons, 2012; Phillips, 1987; Wong, 1990).

The main point here ultimately ties back to the goals or purposes of the discussion coupled with the teacher's evaluation of the discussion just concluded. The teacher must make an on-the-spot judgment that takes in how the discussion has progressed and how it relates to the learning objectives, then provide bridges between the character of the discussion and its pedagogical point (Collins et al., 1989; Phillips, 1987; Wong, 1990). Concluding moves may also involve meta-commentary on how well the students engaged in the discussion, modeling the appropriate discourse norms so that concluding commentary also underscores purposes associated with helping students learn how to have discussions that comport with norms of inquiry and analysis in the various disciplines (Collins et al., 1989). Here, too, teachers' deep understanding of the subject matter enters the account.

Conversely, teachers may fail to make any kind of concluding moves because they have not allowed enough time or have not perceived what concluding moves may be called for. Concluding moves then may be of several kinds, and it may be easier to perceive the lack of closure than to distinguish among various choices that might constitute valuable closure.

Although meaningful classroom discussions are built around what students say, the teacher's role is critical. What the teacher does to set up and support the discussion requires close attention to specific practices. Teachers must take up student ideas, provide feedback on students' responses where appropriate, revoice students' contributions, and press students to elaborate their ideas in order to encourage student-to-student interaction and lead students toward the learning goal.

Limitations and boundary conditions. This measurement approach involves a number of limitations. First, avatars controlled by simulation specialists emulate a classroom but do not adequately represent all of the features associated with live classroom interactions. As well, the assessment relies on the assumption that simulation specialist performances may be standardized to minimize irrelevant variation in the scoring. As with any standardized performance assessment, training, certification, and periodic reviews are the means employed for quality control.

In real instructional settings, discussion occurs in the flow of activity such that what teachers do before and after a discussion matters for the effectiveness of the performance. In this case, the preparatory materials stipulate what the students have already learned in order to provide a common basis from which to measure discussion skills. In practice, teachers cannot

count on the fact that all students enter a discussion with the same base of learning, so in this respect, the simulation does not reflect discussion in real classrooms. Confining discussion in this way is a limitation, but the claim here is that if teachers are unable to conduct discussion under this condition, they probably will be unable to do so under the more complex circumstance of uneven entering knowledge among students.

And, as we discussed previously, the simulated classroom includes five "students" while most real classrooms include 20 or more. It is clearly true that larger classes increase complexity and demand on teachers, but we argue that managing discussion among five students meets the necessary but insufficient criterion for a licensure decision (AERA, APA, & NCME, 2014).

Finally and as already described, research tends to treat discussion-based teaching in a global fashion, rather than testing for the contribution to process and outcomes made by specific practices such as uptake or revoicing. The theoretical literature on discussion combined with both logical and normative arguments supports the aspects and dimensions of discussion as described here. Rubric elements then enjoy substantial support of this kind in the literature on discussion-based teaching.

Conclusion

Validation is a process through which evidence is accumulated and sifted, not an end state or property of a measure or test (Kane, 2004). As such, collecting validity evidence to support licensure assessments is necessarily an ongoing enterprise. Important in this process is transparency so that those affected by measurement understand and have confidence in the reasoning and the evidence that supports the definition and importance of the construct and its measurement. 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 teacher licensure.

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 assessment series continues 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.

LCD meets the standard for centrality to effective teaching practice due to its long history in models for effective teaching and its ongoing central role across many efforts to define good teaching. Such efforts have included research syntheses, observation instruments, accounts of best practice, 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 is 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 strong justification 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 Association. (2014). *Standards for educational and psychological testing* (4th ed.). Washington, DC: American Educational Research Association.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin* (M. Holquist, Ed., & C. Emerson & M. Holquist, Trans.). Austin: University of Texas Press.

Bakhtin, M. M. (1986). *Speech, genres, and other late essays* (V. W. McGee, Trans.). Austin: University of Texas Press.

Ball, D. L. (1993a). Halves, pieces, and twoths: Constructing and representing 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. *Elementary School Journal, 93*(4), 373–397.

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, 29*(1), 14–17, 20–22, 43–46.

Boerst, T. A., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record, 113*(12), 2844–2877.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology, 88*(1), 18.

Buber, M. (1970). *I and thou* (W. Kaufmann, Trans.). New York, NY: Charles Scribner's Sons.

Burbules, N. C. (1993). *Dialogue in teaching: Theory and practice*. New York, NY: Teachers College Press.

Cazden, C. B., & Beck, S. W. (2003). Classroom discourse. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 165–197). New York, NY: Routledge.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2003). *Classroom discussions: Using math talk to help students learn, grades K-6* (2nd ed.). Sausalito, CA: Math Solutions.

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

Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogeneous classroom* (3rd ed.). New York, NY: Teachers College Press.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 32–42). Hillsdale, NJ: Lawrence Earlbaum.

Cooper, K. (2014). Eliciting engagement in the high school classroom: A mixed methods examination of teaching practices. *American Educational Research Journal, 51*(2), 363–402.

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.* Retrieved from http://www.ccsso.org/Resources/Publications/InTASC_Model_Core_Teaching_Standards_and_Learning_Progressions_for_Teachers_10.html

Danielson, C. (2011). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision & Curriuclum Development.

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

Dolezal, S. E., Welsh, L. M., Pressley, M., & Vincent, M. M. (2003). How nine third-grade teachers motivate student academic engagement. *The Elementary School Journal, 103*(3), 239–267.

Duncan, C. A., & Clemons, J. M. (2012). Closure: It's more than just lining up. *Strategies, 25*(5), 30–32.

Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English*, *23*(1), 4–29.

Faltis, C. J., & Valdes, G. (2016). Preparing teachers for teaching in and advocating for linguistically diverse classrooms: A vademecum for teacher educators. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 549–592). Washington, DC: American Educational Research Association.

Fogo, B. (2014). Core practices for teaching history: The results of a Delphi panel survey. *Theory & Research in Social Education, 42*(2), 151–196.

Franke, M. L., Turrou, A. C., Webb, N. M., Ing, M., Wong, J., Shin, N., & Fernandez, C. (2015). Student engagement with others' mathematical ideas: The role of teacher invitation and support moves. *The Elementary School Journal, 116*(1), 126–148.

Gall, M. D., & Gillett, M. (1980). The discussion method in classroom teaching. *Theory into practice, 19*(2), 98–103.

Goldenberg, C. (1992). Instructional conversations: Promoting comprehension through discussion. *The Reading Teacher, 46*(4), 316–326.

Goldenberg, C., & Patthey-Chavez, G. (1995). Discourse processes in instructional conversations: Interactions between teacher and transition readers. *Discourse Processes, 19*(1), 57–73.

Good, T. L., & Brophy, J. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–775). New York, NY: Macmillan.

Greenleaf, C., Schoenbach, R., Cziko, C., & Mueller, F. L. (2001). Apprenticing adolescent readers to academic literacy. *Harvard Educational Review, 71*(1), 79–129.

Grossman, P. L. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research, 62*(2), 171–179.

Grossman, P. L., Schoenfeld, A., & Lee, C. (2005). Teaching subject matter. In L. B. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness, & H. Duffy (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 201–231). San Francisco, CA: Jossey-Bass.

Hadjioannou, X. (2007). Bringing the background to the foreground: What do classroom environments that support authentic discussions look like? *American Educational Research Journal, 44*(2), 370–399.

Haroutunian-Gordon, S. (1991). *Turning the soul: Teaching through conversation in the high school*. New Haven, CT: Yale University Press.

Haroutunian-Gordon, S. (2010). *Learning to teach through discussion: The art of turning the soul*. New Haven, CT: Yale University Press.

Harris, D. E. (2002). Classroom assessment of civic discourse. In W. C. Parker (Ed.), *Education for democracy: Contexts, curricula, assessments* (pp. 211–233). Charlotte, NC: Information Age.

Hawkins, D. (1974). *The informed vision: Essays on learning and human nature*. New York, NY: Agathon.

Henning, J. E. (2005). Leading discussion: Opening up the conversation. *College Teaching, 53*(3), 90–94.

Hess, D. E. (2004). Discussion in social studies: Is it worth it? *Social Education, 68*(2), 151–157.

Hess, D. E., & Posselt, J. (2002). How high school students experience and learn from the discussion of controversial public issues. *Journal of Curriculum and Supervision, 17*(4), 283–314.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794–831.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Jansen, A. (2006). Seventh graders' motivations for participating in two discussion-oriented mathematics classrooms. *The Elementary School Journal, 106*(5), 409–428.

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

Kazemi, E., & Hintz, A. (2015). *Intentional talk: How to structure and lead productive mathematical discussions*. Portland, ME: Stenhouse.

Kersaint, G. (2015). *Orchestrating mathematical discourse to enhance student learning*. Retrieved from http://casamples.com/downloads/WP_OrchestratingDiscourse.pdf

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research, 67*(3), 271–299.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 143. doi:10.1016/j.tate.2013.07.010

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. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

Langer, J. A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal, 38*(4), 837–880.

Langer, J. A., & Close, E. (2001). *Improving literary understanding through classroom conversation*. Albany, NY: National Research Center on English Learning & Achievement.

Larson, B. E. (1996, November). *Social studies teachers' conception of discussion: A grounded theory study*. Paper presented at the annual meeting of the National Council for the Social Studies, Washington, DC.

Lemov, D. (2010). *Teach like a champion: 49 techniques that put students on the path to college (K–12)*. San Francisco, CA: John Wiley & Sons.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.

McKeown, M. G., & Beck, I. L. (2015). Effective classroom talk is reading comprehension. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44*(3), 218–253.

Michaels, S., O'Connor, C., & Resnick, L. B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education, 27*(4), 283–297.

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), 740–264.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010a). *Comon Core State Standards: English Language Arts.* Retrieved from http://www.corestandards.org/ELA-Literacy/

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010b). *Comon Core State Standards: Mathematics.* Retrieved from http://www.corestandards.org/Math/

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, NY: Teachers College Press.

Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English, 40*, 392–412.

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English, 25*(3), 261–290.

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35*(2), 135–198.

Okolo, C. M., Ferretti, R. P., & MacArthur, C. A. (2007). Talking about history: Discussions in a middle school inclusive classroom. *Journal of Learning Disabilities, 40*(2), 154–165.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction, 1*(2), 117–175.

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

Parker, W. C. (2006). Public discourses in schools: Purposes, problems, possibilities. *Educational Researcher, 35*(8), 11–18.

Parker, W. C., & Hess, D. E. (2001). Teaching with and for discussion. *Teaching and Teacher Education, 17*(3), 273–289.

Phillips, L. V. (1987). Closure: The fine art of making learning stick. *Instructor, 97*(3), 36–38.

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom assessment scoring system: Secondary manual*. Charlottesville, VA: Teachstone.

Protocol for Language Arts Teaching Observation. (2013b). *Description of the thirteen elements.* Retrieved from http://platorubric.stanford.edu/Elements.html#dis

Protocol for Language Arts Teaching Observation. (2013a). *The protocol for language art teaching observation.* Retrieved from http://platorubric.stanford.edu

Purcell-Gates, V., Duke, N., & Stouffer, J. (2016). Teaching literacy: Reading. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1217–1268). Washington DC: American Educational Research Association.

Qi, Y., Bell, C. A., & Gitomer, D. H. (2014, April). *The role of topic and activity structure in teacher observation scores*. Paper presented at the American Educational Research Association, Philadelphia, PA.

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.

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. Testing and assessment in school psychology and education* (pp. 391–414). Washington DC: American Psychological Association.

Reisman, A. (2015). Entering the historical problem space: Whole-class text-based discussion in history class. *Teachers College Record, 117*(2).

Resnick, L. B., Asterhan, C. S., & Clarke, S. N. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well structured) talk builds the mind. In D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: perspectives on learning, teaching, and human development* (pp. 163–194). New York, NY: Springer.

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

Saunders, W. M., & Goldenberg, C. (1999). The effects of instructional conversations and literature logs on the story comprehension and thematic understanding of English proficient and limited English proficient students. *Elementary School Journal, 99*(4), 277–301.

Schoenfeld, A. H. (1983). *Problem solving in the mathematics curriculum: A report, recommendations, and an annotated bibliography*. Washington, DC: Mathematical Association of America.

Schwab, J. (1978). Education and the structure of the disciplines. In I. Westbury & N. Wilkof (Eds.), *Joseph Schwab. Science, curriculum and liberal education* (pp. 229–272). Chicago, IL: University of Chicago Press.

Selling, S. K. (2015, April). *Making mathematical practices explicit in discourse: Experienced and beginning instruction*. Paper presented at the National Council of Teachers of Mathematics, Boston, MA.

Sfard, A. (2015). Why all this talk about talking classrooms? Theorizing the relation between talking and learning. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content area literacy. *Harvard Educational Review, 78*(1), 40–59.

Shaughnessy, M., & Forzani, F. (2012, Month). *High-leverage teaching practices in teacher education and assessment*. Paper presented at the Indiana Association of Colleges for Teacher Education, Muncie, IN.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education, 28*(2–3), 235–260. doi:10.1080/09500690500336957

Smith, M. S., Hughes, E. K., Engle, R. A., & Stein, M. K. (2009). Orchestrating discussions. *Mathematics Teaching in the Middle School, 14*(9), 548–556.

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.

Stein, M. K., Engle, R. A., Smith, M., & Hughes, E. K. (2015). Orhcestrating productive mathematical discussions: Helping teachers learn to better incorporate student thinking. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

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

Stein, M. K., & Smith, M. (2011). *5 practices for orchestrating productive mathematics discussions*. Reston, VA: National Council of Teachers of Mathematics.

Sun, J., Anderson, R. C., Lin, T.-J., & Morris, J. (2015). Social and cognitive development during collaborative reasoning. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Tannebaum, R. (2014). Preservice social studies teachers' conceptions of and experiences with discussion as a pedagogical tool. *Journal of Social Studies Research*, 135–140.

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 Research in Education, 57*(1), 69–95.

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

Van den Branden, K. (2000). Does negotiation of meaning promote reading comprehension? A study of multilingual primary school classes. *Reading Research Quarterly, 35*(3), 426–443.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental psychological processes*. Cambridge, MA: Harvard University Press.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research, 78*(3), 516–551.

Webb, N. M., Franke, M. L., Turrou, A. C., & Ing, M. (2015). Exploration of teacher practices in relation to profiles of small-group dialogue. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Webster, C. A., Connolly, G., & Schempp, P. G. (2009). The finishing touch: Anatomy of expert lesson closures. *Physical Education and Sport Pedagogy, 14*(1), 73–87.

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations: Lessons learned in four districts*. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/Evaluating-Teachers-with-Classroom-Observations.pdf

Wilkinson, I. A., Murphy, P. K., & Binici, S. (2015). Dialogue-intensive pedagogies for promoting reading comprehension: What we know, what we need to know. In C. S. Asterhan, S. N. Clarke, & L. B. Resnick (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington DC: American Educational Research Association.

Wilson, S. M., & Wineburg, S. S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record, 89*(4), 525–539.

Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2006). *Accountable talk in reading comprehension instruction* (CSE Technical Report 670). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wong, Y. H. (1990). Lesson closure: The way to make learning stick. *Teaching and Learning, 11*(1), 15–22.

Appendix. Leading Classroom Discussion (LCD) Sample Task

Leading Classroom Discussion
Pet Show!—Grade 1

Introduction

This assessment measures your ability to lead a small-group discussion with students. The purposes of a discussion are to (1) support students in building understanding in relation to specific learning goals and (2) encourage students to practice listening, speaking, and interpreting ideas. In a discussion, the teacher and students use one another's ideas as resources as they work together on specific content. The teacher's role is to elicit student ideas, help students see similarities and differences among their own and others' ideas, and use those ideas to make progress toward a specific learning goal or goals. The teacher and all students contribute orally, listen actively, and respond to and learn from each other's contributions. The teacher summarizes and articulates one or more conclusions that are focused on the key concepts and learning goals.

You will facilitate a small-group discussion with five student avatars who will appear on a computer screen in front of you in a virtual, interactive classroom environment. You should interact with them as you would interact with students in a typical classroom.

Preparation and Timing

You will have a maximum of 30 minutes of preparation time; however, you do not need to use all of it. Following your preparation, you will have 15 minutes to conduct your discussion in a simulated classroom with the student avatars. At the end of 15 minutes, the proctor will close the simulator, and your interaction will be finished. A timer will start when you begin your discussion. You should monitor the timer to ensure that you have enough time to carry out the discussion, summarize it, and bring the discussion to a conclusion.

The discussion you will lead is designed to be the last 15 minutes of a longer lesson. Assume that you have already taught the first part of the lesson to this group of students. Below is a synopsis of the first part of the lesson and instructions for how to launch into the discussion. Also included are learning goals for the discussion which will be based on the story *Pet Show!*

Materials

You will have a shared electronic workspace and scratch paper to use as you prepare for and lead your discussion. You may use these materials and any notes you have as you lead the discussion.

Scoring Information

You will be assessed on the following aspects of your performance.

- Eliciting and probing for ideas from all students
- Using students' ideas to steer the discussion toward the learning goals
- Representing the content accurately, verbally and/or visually
- Summarizing and concluding the discussion

Lesson and Discussion Outline

Synopsis of the first part of the lesson	In the first part of the lesson, you led these first-grade students through a reading of a text called *Pet Show!* by Ezra Jack Keats. You ensured that all students understood the text (i.e., what happened in the story), but you did not begin any further discussion about the text.
	To prepare students for the work they are about to do, you have already introduced the process of making inferences and explained that making inferences involves using evidence from the text and background knowledge to understand ideas not explicitly stated in the text.
Plan for this part of the lesson	In this next part of the lesson, you will lead a discussion aimed at helping students develop their ability to make and support inferences. During this part of the lesson, students have copies of the text.
	You have asked the students to think on their own about **what the personality of the main character (Archie) is like and what evidence there is in the text (e.g., key events, characters' actions) to support different claims about his personality.** Now you will begin a discussion about the students' ideas.
	The learning goal for the discussion is listed below.
	• In order to more deeply comprehend the text, make inferences about Archie's personality and support those inferences with evidence from the text.
	Once your session begins, you should immediately launch into the discussion as if you had already carried out the first part of the lesson as described above.
	When you are ready to launch the virtual classroom, you will say, "Begin interaction." Once you say this, the discussion will begin, and you will have 15 minutes to carry out the rest of the discussion and bring it to a close. To signal to the student avatars that you are ready, begin the discussion by saying, **"You all have done a good job reading and understanding the text Pet Show!. Now I want us to think more carefully about Archie's personality—about what he is like as a person. What words would you use to describe Archie?"**
	During the discussion, you may use the shared electronic workspace to record students' ideas.

Additional Information

The following materials are designed to help you understand the English language arts content and the ways in which students at this grade level would be likely to interact with this text, including difficulties students might face around this content. The following can be used as a resource for you when planning the discussion, but this content is not designed for use with students and should not be used as a lesson plan.

Notes on the text	The story *Pet Show!* is about a boy who is creative. The fact that Archie says that he has brought a germ shows Archie is creative and inventive because he was able to think of a way to participate even though germs are an unusual choice for a pet show. Other personality traits that are supported by the text include, but are not limited to, the following. • Kind • Good friend • Smart • Generous • Determined
Common student challenges	Students may struggle with the following. • Making inferences • Identifying appropriate evidence • Logically connecting and/or explaining evidence and inferences (e.g., explaining how bringing a germ shows that he is creative) • Distinguishing facts about Archie from personality traits (e.g., he seems to have lots of friends versus he is friendly and kind) • Vocabulary (e.g., students default to referring to Archie as "nice")
Content knowledge for teaching	When readers make inferences they use clues in the text and their own background knowledge to come to a logical conclusion about something that is not explicitly stated. In this task, you will ask students to make inferences about personality traits. **Personality traits are qualities of a person's nature that become evident in various ways that include, but are not limited to, what the person says and does.**

Notes

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