Research Report

Using Performance Tasks to Provide Feedback and Assess Progress in Teacher Preparation

ETS RR–23-08

Geoffrey Phelps
Devon Kinsey
Thomas Florek
Nathan Jones

December 2023

ETS®

ETS Research Report 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 Report 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 Report 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.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Using Performance Tasks to Provide Feedback and Assess Progress in Teacher Preparation

Geoffrey Phelps, Devon Kinsey, Thomas Florek, & Nathan Jones

ETS, Princeton, NJ

This report presents results from a survey of 64 elementary mathematics and reading language arts teacher educators providing feedback on a new type of short performance task. The performance tasks each present a brief teaching scenario and then require a short performance as if teaching actual students. Teacher educators participating in the study first reviewed six performance tasks, followed by a more in-depth review of two of the tasks. After reviewing the tasks, teacher educators completed an online survey providing input on the value of the tasks and on potential uses to support teacher preparation. The survey responses were positive with the majority of teacher educators supporting a variety of different uses of the performance tasks to support teacher preparation. The report concludes by proposing a larger theory for how the performance tasks can be used as both formative assessment tools to support teacher learning and summative assessments to guide decisions about candidates' readiness for the classroom.

Keywords performance assessment; formative assessment; teacher education; core-practices; mathematics; reading language arts

doi:10.1002/ets2.12371

In this report, we explore the potential for using a new type of short performance task to explore uses both as part of teacher preparation and for assessing key competencies needed for effective teaching. The report begins by discussing arguments to organize teacher development and associated assessment tools around competencies used in the work of teaching school subjects. We trace evolving ideas about the knowledge and skills that need to be the focus of teachers' professional learning, emphasizing recent interest in the use of teaching simulations to provide novices with opportunities to learn core teaching practices. The main body of the report presents results from a survey of 64 elementary mathematics and reading language arts teacher educators who were asked to evaluate both the value and potential uses of the newly developed short performance tasks. We present detailed descriptions of the performance tasks and their key design features, study instruments and methods, and survey responses. We conclude by expanding our discussion to consider a general theory for how simulated tasks, in particular short performance tasks, can be used to support both professional learning and ultimately teaching quality.

New Directions in Teacher Preparation

> The field of teacher education is in the midst of a major shift—a turn away from an intense focus on the knowledge needed for teaching to a focus on the use of that knowledge in practice. The fundamental goal undergirding this turn is to better support teachers in learning to enact teaching practices skillfully and knowledgeably in ways that support student learning. (Grossman et al., 2018, p. 3)

From the very start of formal teacher education, there has been a focus on what teachers need to learn to teach effectively. While ideas about what constitutes good teaching have evolved, educators have long focused on providing novice teachers with opportunities to learn how to teach. For example, as part of the normal school movement of the 1800s, novice teachers were required to prepare and deliver rehearsal lectures in front of their professors and fellow novice teachers as a way to practice and demonstrate their teaching skill (Forzani, 2014). In the 1960s and 1970s, a line of research often referred to as "process-product" set out to identify effective teaching behaviors. This research led to several general recommendations such as leaving a certain amount of "wait time" after a question to allow students an opportunity to

Corresponding author: Geoffrey Phelps, E-mail: gphelps@ets.org

respond (Gage & Needels, 1989). Building on the findings from the process-product research, “microteaching” programs were developed to train novice teachers to carry out short enactments of these teaching behaviors (Macleod, 1987).

Starting in the 1980s, the process-product research and associated interventions such as microteaching fell out of favor as scholars pointed out that behaviors such as wait time are highly depended on the context of instruction, including the particulars of an activity, content focus, and students’ learning needs (Shulman, 1986a). Attention shifted from the more discrete behaviors of the process-product research to a focus on the moment-to-moment judgments and the associated knowledge about content, students, and instructional goals used in the work of teaching (see, for example, Ball et al., 2008; Lampert, 1985; Shulman, 1986b, 1987). Teacher educators increasingly considered ways to organize teacher preparation and novice teachers’ development of understanding around the knowledge and decisions used in teaching.

More recently, Grossman, Hammerness, and McDonald (2009) argued for renewed attention to practice as the organizing feature of teacher preparation:

> The view of teaching over the past several decades has evolved from an emphasis on teacher characteristics to a focus on teachers’ behavior to more recent cognitive views of teachers as decision-makers and reflective practitioners … Teacher education should move away from a curriculum focused on what teachers need to know to a curriculum organized around core practices, in which knowledge, skill and professional identity are developed in the process of learning to practice. (p. 274)

In the sections that follow, we provide a brief overview of this new turn toward practice. We start by summarizing arguments for practice-based teacher preparation. Next, we provide an overview of the shift toward practice-based conceptualizations of teacher content knowledge. Then we focus on the most recent calls to organize teacher preparation around a set of core practices. Finally, we highlight the importance of developing standardized opportunities for teachers to engage in approximations of practice.

Practice-Based Teacher Preparation

Interest among educational researchers and teacher educators in organizing professional learning around the work that teachers do in the moment-to-moment and day-to-day work of teaching school subjects has generally been referred to as “practice-based” teacher education. The term practice-based points to a primary focus on the knowledge and skills that are used to carry out some aspect of the work of teaching. For example, Ball and Cohen (1999) called for redesigning teacher education so that teaching is placed at the very center of professional learning. They argued that “one way to do this lies within the course of teachers’ everyday work—in the regular tasks of planning, enactment, reflection, and assessment” (p. 20). As Ball and Cohen contended, organizing professional preparation around practice means something both different and more than field placement or other opportunities to be immersed in actual K-12 classrooms:

> Centering professional education in practice is not … about a physical local or some stereotypical professional work, [but rather] about the terrain of an action and analysis that is defined first by identifying the central activities of teaching practice and, second, by selecting or creating materials that usefully depict that work and could be selected, represented, or otherwise modified to create opportunities for novice and experienced practitioners to learn. (p. 13)

Practice-based approaches to teacher learning call for (re)conceptualizing teacher preparation in ways that organize around what teachers do in ways that integrate practice into all aspects of the professional curriculum.

These new arguments for practice-based professional preparation challenge the status quo where typical teacher education coursework has traditionally attended to developing content knowledge, theories of teaching and student development, and learning about various teaching methods. In this traditional breakdown of teacher competence, the more practical aspects of actual teaching were largely the province of the school-based practicum and student teaching. Mary Kennedy (1999) noted that this divide between coursework and the K-12 classroom creates “a problem of enactment,” where novice teachers are left largely to their own devices to bridge the gulf between what they have learned about in their teacher preparation coursework and what they need to learn to do as they teach in the classroom. A focus on practice-based approaches in teacher education can be seen as one way to bridge this gulf and address the problem of enactment.

Practice-Based Models for Understanding Teacher Content Knowledge

At least initially, practice-based approaches to teacher preparation focused primarily on drawing a more direct connection between teaching practice and attention to closely related forms of knowledge and reasoning. One important example of knowledge that is closely linked to practice is the concept of pedagogical content knowledge or PCK. Shulman (1986b) defined PCK as:

> The most useful forms of representation … the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the most useful ways of representing and formulating the subject that make it comprehensible to others … Pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult. (p. 7)

The move toward practice-based professional learning recognized that PCK is directly linked to, and arguably defined by, the content work of teaching (Ball & Bass, 2003).

These ideas had a major impact on scholarship focused on the types of knowledge and professional competencies that need to be taught and learned as part of professional preparation (Ball et al., 2008). General interest in identifying the knowledge used in teaching led to efforts across school subjects to both identify and assess these forms of professional knowledge (see, for example, Gitomer et al., 2015; Herbst & Kosko, 2014; Hill et al., 2004; Kersting, 2008; Krauss et al., 2008; Mikeska et al., 2018; Phelps, Gitomer, et al., 2020; Phelps & Schilling, 2004; Sadler et al., 2013; Smith & Banilower, 2015; Tatto et al., 2008).

This shift in the focus of teacher preparation also has implications for practice-based assessment. In discussing the assessment of the content knowledge used in teaching, Phelps, Gitomer, et al. (2020) argued that the conceptualizing of tasks of teaching that have focused on knowledge could be expanded to also include performance competencies. For example, tasks of teaching such as "explaining to support student understanding, could be extended from the current focuses on cognitive competencies, such as evaluating or selecting explanations, to also assess the content teaching skills used to generate and deliver an explanation" (p. 108). In this view, the work of teaching is composed of a wide range of practices, some focused directly on thought processes and decisions while others are focused on teaching behaviors and performances.

The Core-Practice Movement

The current turn to focus on core practices should not be interpreted as rejecting or downplaying the importance of teachers developing knowledge about content, students, teaching goals, and so on. Instead, the shift is better understood as taking an expanded view that incorporates knowledge and decision making into the work of teaching. Grossman, Hammerness, and McDonald (2009) described this turn to focus on core practices as a move to

> … help novices develop professional knowledge and skill, as well as an emerging professional identity around these practices. The practices of teaching provide the warp threads of the professional curriculum, while the knowledge and skill required to enact these practices constitute the weft. (p. 277)

Ball and Forzani (2009) also emphasized the critical role of knowledge in learning core practices:

> To make practice the core of the curriculum of teacher education requires a shift from a focus on what teachers know and believe to a greater focus on what teachers do. This does not mean that knowledge and beliefs do not matter but, rather, that the knowledge that counts for practice is that entailed by the work. (p. 503)

This turn toward core practices, therefore, is not a turn away from knowledge but rather a move to see knowledge and beliefs as integral to the behaviors and complex decisions that make up the work of teaching school subjects.

Identifying Core Practices

To make progress identifying the core practices that could serve as the basis for a new type of teacher education curriculum, Grossman, Hammerness, and McDonald (2009) suggested an initial set of criteria. Core practices are ones that

> … occur with high frequency in teaching; novices can enact in classrooms across different curricula or instructional approaches; novices can actually begin to master; allow novices to learn more about students and about teaching; preserve the integrity and complexity of teaching; and are research-based and have the potential to improve student achievement. (p. 277)

Teacher educators working to define core practices have started to generate frameworks. Some efforts have focused on identifying the core practices that are used to teach a particular subject such as mathematics (Franke et al., 2007), science (Kloser, 2014; Windschitl et al., 2012), English language arts (Grossman et al., 2013), or history (Fogo, 2014) or a particular population such as students with learning disabilities (McLeskey & Brownell, 2015). Other approaches have sought to identify core practices that are relevant regardless of grade level or subject (Ball & Forzani, 2011). Some core practices, such as leading group discussions or modeling and explaining concepts, are almost ubiquitous across these frameworks, while other practices such as "providing focused and intensive instruction for homogenous groups with special learning needs" (McLeskey & Brownell, 2015) or "linking science concepts to phenomena" (Kloser, 2014) are uniquely tied to a particular group of students or a particular subject. Each of these efforts, however, share a focus on identifying a bounded set of teaching practices that can be used to structure learning opportunities for novice teachers.

Pedagogies of Practice

In their seminal cross-professional study, Grossman, Compton, et al. (2009) investigated the methods used to teach teachers to teach, what they refer to as pedagogies of practice. Their analysis represents an important step toward developing a theory for professional learning by proposing three central features that characterize the teaching and learning of professional practice: decomposition, approximation, and representation.

- *Decomposition* refers to the breaking down of complex professional work into its constituent core practices for the purposes of supporting professional learning. When complex work like teaching is decomposed, the resulting component core practices are small and more manageable for facilitating professional learning. These practices need to be meaningful in the sense that they capture key features of the more complex work of teaching effectively. But they also need be manageable in that they are useable by teacher educators and learnable for novices.
- *Approximation* refers to opportunities to engage in a core practice in ways that are related to the actual practices that make up the work of teaching. Approximations provide opportunities for teachers to try out a practice and receive feedback.
- *Representation* refers to how a core practice is designed to foreground and focus on a particular feature of the practice that is important for teachers to notice, consider, and learn.

These ideas suggest that a set of core practices can become the foundation for a curriculum of professional preparation, and the activities used to approximate and represent teaching can be viewed as an initial step toward understanding the teaching and learning interactions that are used to enact this professional curriculum. Novice teachers are given the opportunity to learn about and try out a core practice, receive feedback, and then learn more about the practice before trying it out again. Professional knowledge, beliefs, and identity are developed around these core practices and directly tied to the work of classroom teaching.

Implicit in this cycle of professional learning are standards of quality that specify what teachers need to learn to do, what constitutes a successful or high-quality performance, and the criteria that guide feedback (Moss, 2011). Learning how to *do* core practices to a clearly defined standard of quality arguably provides a set of foundational skills that provide the basis for enacting more complex practices in actual classroom instruction.

Simulating Practice

Teacher educators across disciplines have begun to develop approximations of practice that are aligned with the core practices suggested by Grossman, Compton, et al. (2009). These approximations of practice, also referred to as rehearsals or simulations, are designed to scaffold novices' learning opportunities, often through a cycle of planning, enactment, and debriefing (Lampert & Graziani, 2009). Most examples of rehearsal appear to be focused on the core practice of classroom discussion (see, for example, Alston et al., 2018; Amador, 2017; Davis et al., 2017; Kavanagh et al., 2020; Kazemi et al., 2015; Kloser et al., 2019; Shaughnessy et al., 2019).

While rehearsals can vary, they typically are designed around a role play simulation in which a novice teacher plays the role of the K-12 teacher and other novice teachers play the role of K-12 students. Simulations may require substantial preparation before the actual rehearsal to provide background on the goals and the various roles involved (Schutz et al., 2018). For example, novice teachers playing the role of K-12 students can spend substantial time considering the types of insights, challenges, and behaviors that K-12 students might exhibit during the rehearsal. Preparation of this sort, while involved, can provide deep opportunities to learn about practice. During the live simulation, teacher educators can pause a simulation at various points to provide an opportunity for the teacher educator, or other novice teachers participating in or observing the simulation, to share insights and explore questions or ideas (Averill et al., 2016; Campbell et al., 2020; Davis et al., 2017). Simulations can also be rerun to provide novices with opportunities to retry aspects of the focal practices.

Virtual reality environments can also be used to provide opportunities to approximate or simulate practice. For example, digital human-in-the loop technologies (Dieker et al., 2014) allow a novice teacher to interact with virtual students who are voiced and animated in real time by a human behind the scenes. Using the computer audio and camera, the simulation specialist can respond in real time to the novice teacher as the simulation unfolds. The student avatars can be enacted with distinct personalities (e.g., shy, outgoing), behavior profiles (e.g., disruptive, compliant), or learning profiles. The student avatars can also be assigned different racial characteristics, genders, and ages, allowing for a simulation encounter where the virtual students more closely approximate the characteristics of a particular group of K-12 students.

Whether simulations make use of role play or immersive digital environments, they can be designed to provide opportunities for novice teachers to learn about and try out a wide range of practices such as classroom management (Delamarre et al., 2021), culturally responsive teaching (Hardin & Freeman-Green, 2015; Self & Stengel, 2020), conducting parent-teacher conferences (Kelley & Wenzel, 2018), and providing feedback to support novice teacher learning (Bardach et al., 2021).

Standardizing Simulations of Practice

Development and associated research on approximations of practice have largely been conducted and limited to a local context. They are most often developed by a single or small group of teacher educators, designed to address the needs of a particular methods course, and rarely developed so that teacher educators other than the developers can implement the simulation. The actual task, material, and purpose of the simulation are typically selected and shaped by the methods instructor based on a local need, interest, or learning goal. Often the standards used to evaluate performances are also developed and implemented locally.

While there are advantages to local development, including the potential to tailor approximations and simulations to the focus of a particular methods course or the learning needs of a particular group of novice teachers, there are also costs and limitations. Local development places substantial demands on teacher educators, including identifying the focal content, designing the simulation activity itself, identifying criteria for evaluating the quality of performances, and providing feedback that will support novice teachers in developing the desired knowledge and skill. Local development also limits the potential for establishing activities that can be used across methods classes. This sharply constrains the potential for teacher educators to share their work, challenges, and insights with their colleagues. Furthermore, when all teacher educators are essentially using different tasks and materials, there is little basis for developing a common curriculum.

In contrast, standardized simulations can be administered with some level of consistency across settings by incorporating one, or all, of the following features. The first feature is explicit specification of what might be referred to as the "content" of the simulation (i.e., what the simulation is about, its components and objectives, and what constitutes having completed the simulation). The second feature is consistent guidelines for enactment of the simulation. In the cases where the approximation involves interaction, guidelines for how live actors or simulation specialists in the role of K-12 students initiate or respond during the interaction impact both the actual content and challenge of the simulation experience. The third feature is specification of what constitutes a successful performance. Well-specified standards of quality provide the basis for providing consistent and interpretable feedback on the performances (Moss, 2011). To illustrate how these features can vary, we describe four different approaches to standardizing simulations, as follows:

1. *Simulated encounters focused on differences in power and privilege (i.e., SHIFT)*. Each SHIFT encounter "simulates a situation that is common in teaching and that foregrounds identity, positionality, and systems of oppression in an attempt to make them more visible" (Self & Stengel, 2020, p. 3). The SHIFT simulation enactment cycle has five steps. In the first step, the novice teacher reviews a specified critical incident and prepares for the interaction. The second step is a 10–12-minute simulated encounter with a live actor. The live actors receive training that includes information on what novice teachers should learn from the particular SHIFT encounter, how to start the simulation, the essential actions they must take with every participating candidate, anticipated responses from teacher candidates, and how to end the encounter. Live actors are expected to provide a comparable performance across each SHIFT interaction. The final three steps of the SHIFT cycle involve processing the interaction with other novice teachers who each participated in the same encounter, reviewing a video of the interaction, and then discussing the simulation experience. Participants are expected to come to a collective interpretation of the meaning of the encounter and of the associated standards for judging both productive and less-productive teaching.
2. *Argumentation-focused discussion*. Mikeska and Howell (2020) developed virtual reality simulations using human-in-the loop immersive digital technology (Dieker et al., 2014). The goal for the novice teacher is to lead a group of five K-12 student avatars in an argumentation-focused discussion where the students interact with each other's ideas, reconcile incompatible claims, and work toward consensus. Before engaging in the simulation, novice teachers review supporting materials that describe the lesson context, student learning goals, information about students' background, information about previous activities that students engaged in, and a set of specific directions for what they need to do to successfully complete the simulation. In the 20-minute interaction, the novice teachers engage the virtual students in an argumentation-focused discussion. For each discussion, simulation specialists have been trained on five distinct student profiles that include specific conceptual ideas and anticipated responses for each student. The novices' discussions are evaluated across the following dimensions: "(a) attending to students' ideas, (b) facilitating a coherent and connected discussion, (c) encouraging student-to-student interactions, (d) developing students' conceptual understanding, and (e) engaging students in argumentation" (Mikeska & Howell, 2020, p. 10). Human raters code each of these dimensions using a rubric, and results from this coding are used to provide structured feedback to teacher educators and novice teachers.
3. *Eliciting and understanding student thinking*. Shaughnessy and colleagues (Shaughnessy et al., 2019; Shaughnessy & Boerst, 2018) developed a live actor simulation designed to assess the knowledge and skills used in eliciting and understanding student thinking. Before beginning the simulation, the novice teacher is provided 10 minutes to review an example of student work and consider questions and other teaching moves they might make when interacting with the simulated student. A live actor plays the role of the student and is trained on general response orientations (e.g., give the least amount of information that is responsive to the preservice teachers' question) and specific responses to provide for the mathematical questions that the novice teacher asks. The interaction lasts up to 5 minutes and is ended at any point when the novice teacher is satisfied that she or he understands the student's thinking. Human raters are trained to score performances on four dimensions: "(a) eliciting the student's process, (b) probing the student's understanding of key mathematical ideas, (c) attending to the student's ideas, and (d) deploying other moves that support learning about student thinking" (Shaughnessy & Boerst, 2018, p. 45).
4. *Discussion to establish classroom norms*. Cohen et al. (2020) designed a human-in-the-loop, virtual reality simulation focused on facilitating classroom norm setting discussions. During each 5-minute simulation, the novice teachers use redirections to address off-task student behaviors (e.g., humming, taking calls, singing). A redirection is considered effective when it is timely, succinct, and calm based on the Responsive Classroom guidelines for behavior management (Responsive Classroom, 2014). Human raters use a scoring rubric to assess the timeliness of redirections, the proportion of specific redirections, the succinctness of redirections, and overall quality (i.e., effective use of calm, warm and supportive redirection) to provide an overall score for the performance.

These four examples illustrate that a move toward standardization is not an either-or proposition: Not only can a simulation standardize different components, but each of these components can be standardized to varying degrees. For example, in respect to content of the simulation, the SHIFT encounter includes extensive preassignment reading and a detailed questionnaire to prepare both the novice teacher and the live actor for the encounter, while the eliciting and understanding student thinking task presents a single example of students' work immediately before the simulation begins with the expectation that the interaction will focus on this work sample. In respect to standardizing actor or simulation specialist behaviors, the SHIFT encounter allows actors to adjust the behaviors based on how the novice teacher understands the encounter, while the classroom norm simulation provides explicit rules designed to ensure, for example, "that each candidate had the same number of opportunities to respond to similar off-task behaviors" (Cohen et al., 2020, p. 7). In respect to standards for evaluating the performance, the novice teacher participating in the SHIFT encounter develops their own standards for evaluating the performance, while for each of the other simulations trained raters systematically apply a set of rules with the goal of achieving agreement among raters.

Each of the standardized simulations also varies in other basic design features such as the time required to plan for, administer, evaluate, and report on the simulation. For example, the SHIFT simulation involves a multi-hour, multi-stage learning encounter and lengthy discussions to process the encounter. Likewise, the argumentation-focused discussion simulation incorporates extensive preparation materials, a 20-minute simulated performance, and involved scoring and reporting procedures. In contrast, the eliciting and understanding student thinking simulation is relatively short and bounded, allowing only 10 minutes to prepare, and is focused on a single student work sample with a brief performance that can be stopped as soon as the participant is satisfied that they understand the student's thinking. Likewise, the classroom norms simulation focuses on a small set of teaching moves that can be used to redirect students, can be completed in under 5 minutes, and employs a limited number of highly specified scoring rules.

It follows that these simulations are also more or less suitable for different uses depending, for example, on the amount of time available for the simulation, the degree of standardization required, whether simulation will be used as an assessment, if the simulation can be conducted in person, the specific competencies of interest, various costs associated with use of the simulation, and so on. Given the complexity of teaching and the associated professional learning needs for novice teachers, there is arguably a need for a potentially large set of practice-based learning resources, including different types of standardized simulations. In the next section of the report, we describe a new type of standardized simulation task and report on how teacher educators view a range of potential uses.

Teacher Educator Study

The current teacher educator study is designed to gather initial insight into the value and potential uses of a new type of short, standardized performance task. Like other standardized simulations, the newly developed performance tasks simulate core practices in consistent ways across administrations. A focus on short tasks that can be completed in 3 – 5 minutes has several additional affordances including a bounded and targeted focus on a given competency, rapid administration time, straightforward rubrics for evaluating the performance, and suitability for providing concise feedback. These features support a range of uses in teacher education methods coursework that are not possible when using longer and more involved simulations. For example, the short performance tasks can be used for homework assignments that can feasibly be completed by all novice teachers in a methods class, in-class assignments where all novice teachers complete a task and receive feedback from peers or a teacher educator, or assignments that involve completing multiple tasks or retrying the same task multiple times. The short form and support for rigorous, rapid scoring also supports other uses such as including multiple different performance tasks as part of more conventional summative assessments such as teacher licensure tests.

Furthermore, the short performance tasks, unlike other standardized simulations tools, do not involve real-time interaction with a virtual or live actor simulating student responses. Instead, the tasks focus on independent teaching moves and behaviors such as explaining procedures, modeling concepts, responding to student work and questions that do not require a simulated student audience. This focus on teaching moves greatly simplifies the demands and costs associated with developing and administering the short performance tasks because there is no need to involve a human interactor as part of the task administration.

While these design features have many potential advantages that are largely enabled by the short and bounded nature of the tasks, they also introduce a few potential limitations. For example, the high level of decomposition, with a focus on discrete teaching behaviors, runs the risk of breaking up very complex teaching practices in ways that oversimplify or fail to represent the complexity of teaching. Furthermore, novices still need to learn how to knit together the various skills represented into a coherent whole as they learn to carry out longer and more involved practices such as leading discussions.

The purpose of the current study is to gather feedback from teacher educators on how they view the potential affordances and limitations of these new tasks. The study provides insight into the following six primary research questions:

1. Do teacher educators view the performance tasks as relevant and authentic?
2. What are recommended uses of the performance tasks for preservice teachers?
3. What are recommended uses of the performance tasks for teacher educators?
4. What resources would support the use of performance tasks in methods courses?
5. Are teacher educators likely to use performance tasks in their methods courses?
6. Which tasks do teacher educators think are the most and least useful?

To provide insight into how teacher educators with a focus on different subjects and student populations view the performance tasks, the results are also broken out by groups of teacher educators based on their focus on elementary reading language arts (RLA), elementary mathematics, general education, and special education.

Methods

Short Performance Tasks

The short performance tasks used in this study build on and extend frameworks that have been used to develop assessments of content knowledge for teaching (CKT) (Phelps, Gitomer, et al., 2020). However, while CKT assessments identify knowledge used in tasks such as evaluating student work or selecting material to meet particular learning goals, the new performance tasks focus on teaching skills such as explaining content procedures or modeling concepts (Phelps, Bridgeman, et al., 2020). Extending the CKT frameworks provides a useful basis for integrating practice-based knowledge and skill (Phelps & Bridgeman, 2022; Phelps & Sykes, 2020).

The new performance tasks were developed, piloted, and revised as part of two previous studies. In the first study, the tasks were administered to a sample of 59 novice teachers in the last year of their program. As reported in Phelps, Bridgeman, et al. (2020), administration times were within the desired time parameters and human raters had an acceptable level of agreement for an initial pilot. Results from the pilot administration and scoring were reviewed to guide revision of eight RLA and eight mathematics tasks. The next study employed a group comparison administering the newly revised tasks to a sample of 100 novice teachers in the last year of their program and 100 experienced elementary teachers. Unfortunately, this study was interrupted by COVID-19 quarantine restrictions after only 24 administrations to the novice teachers. Nonetheless, one important product of this partially completed study was a revised pool of assessment tasks and a set of sample performances available for use in the current teacher educator study.

The performance tasks each follow a similar design logic presenting information first on the directions screen and next on the performance screen. The directions screen provides a very brief teaching context, information on student and teaching background, grade level, goals for the teaching, and in some cases examples of student work or short video clips of K-12 students interacting or discussing their work. If the performance involves making use of student work or their interactions, this information is provided as part of the directions screen. The directions screen also includes explicit directions for the performance. Four minutes is allowed for the novice teacher to prepare for the performance. As soon as the directions screen is activated, a timer begins a countdown from 4 minutes. At 4 minutes, the performance screen is automatically launched. However, participants can opt to start their performance whenever they are ready. The components of the directions screen are presented in Figure 1.

The performance screen presents a virtual whiteboard. There are a standard set of tools that participants can use to write, mark up, or draw/diagram on the virtual whiteboard. If the task makes use of manipulatives, these will also be included in the toolbar. The virtual whiteboard may be populated by manipulatives, text, student work samples, or other similar material that participants need to mark up or interact with as part of their performance. Both the spoken performance and performance on the virtual whiteboard are captured in real time. Participants can review the task directions at any point by clicking on the task direction bar. This will pull up the directions form the directions screen and will not stop

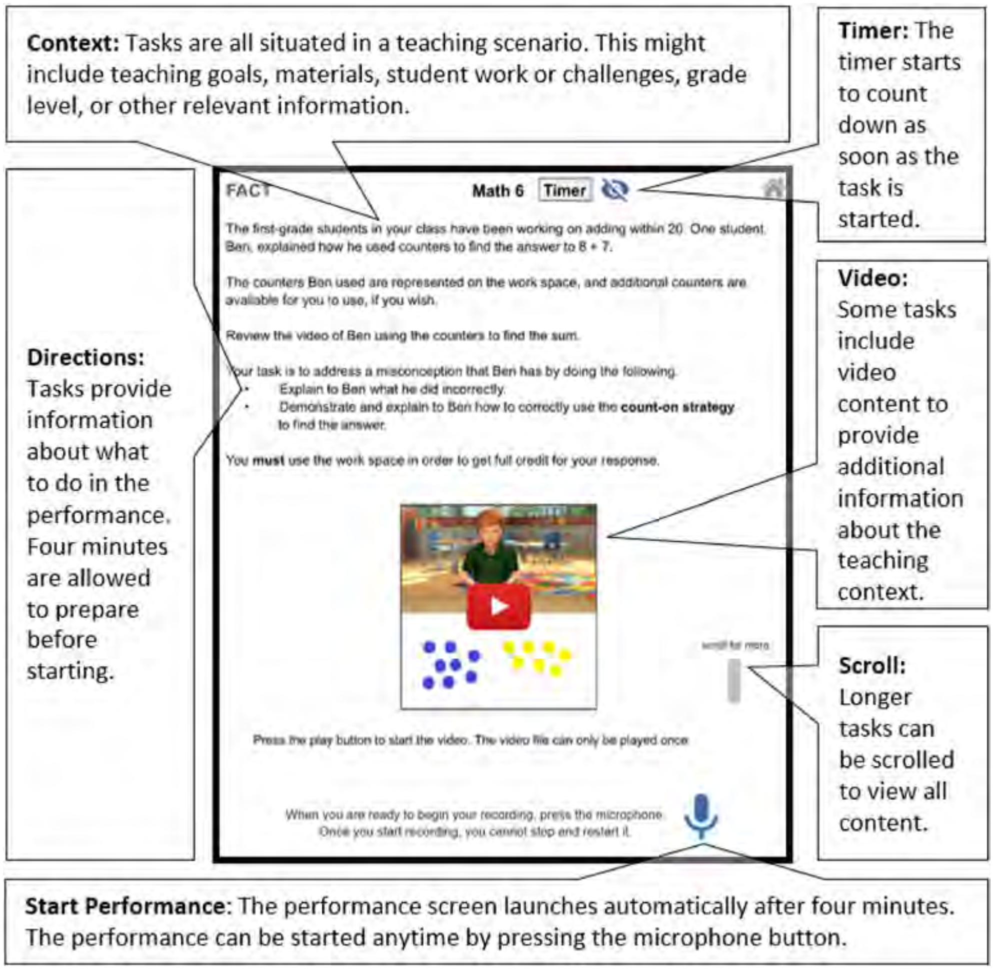


Figure 1 Directions screen.

or pause the performance. The performance screen also includes a timer that begins to count down from 2 and a half minutes as soon as the performance is launched. The performance ends automatically at 2 and half minutes. However, a participant can end the performance at any point prior to that by clicking on the microphone button. Performance screen features are illustrated in Figure 2.

The short performance tasks are designed for administration on tablet devices such as an Apple iPad. Task administration is online using a web interface with all performance data captured on a secure server. Performances are then formatted as movies that are available for viewing for the purposes of providing feedback or more formal scoring through a web interface.

Instruments

Two interconnected instruments were used for the current teacher educator study. The first instrument was a website designed to provide teacher educators with an opportunity to review a set of six subject specific performance tasks (one set for RLA and another set for mathematics). The second instrument was an online survey that presented questions about the performance tasks and collected background information about the teacher educators.

Task Review Websites

The task review websites were designed to provide teacher educators with an opportunity to try out and review the performance tasks. The websites for RLA and mathematics followed the same format and included the same content; the only

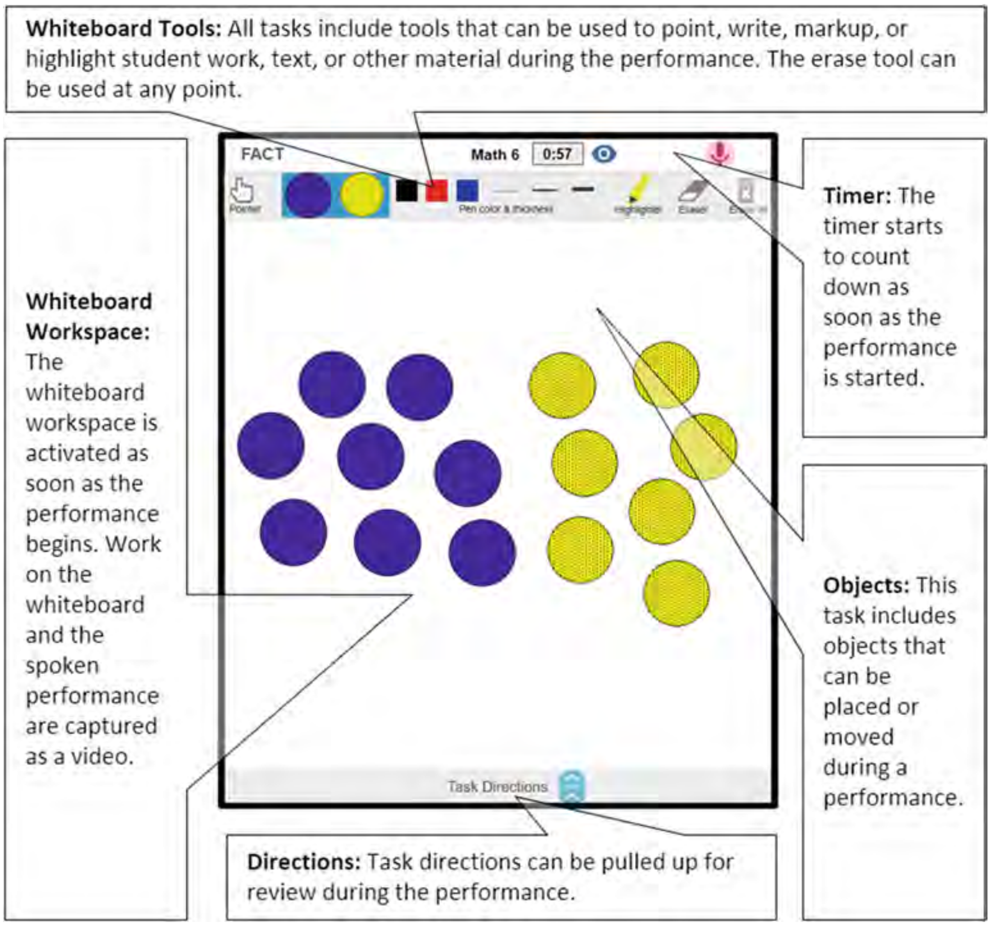


Figure 2 Performance screen.

difference between the two was that the RLA website presented RLA tasks that focused on RLA content and the mathematics website presented mathematics tasks that focused on mathematics content.[1] Each website allowed participants to navigate across four pages of content that included a page for general directions, a page that provided an opportunity to try out each of the six tasks, a page that provided additional in-depth information on two of the tasks, and finally a page that provided a link to the survey. Each of these pages is described in more detail below.

The general directions page provided a brief overview to orient teacher educators to the general logic for using these tasks as part of teacher preparation, the longer-term vision for the types of resources that would support their use by both teacher educators and preservice teachers, and information on the website review process and the type of feedback they would be asked to provide on the survey instrument. The task tryout page provided an opportunity to try out the tasks by accessing video stimulus, using the whiteboard tools and experiencing the actual time allowed to prepare and deliver a performance. Participants were also provided with an overview of the design and functionality for each task. Teacher educators were encouraged to try out each task both by carefully considering the task directions and providing an actual performance. The subject matter foci for the tasks are described in Table 1 for RLA and Table 2 for mathematics. The direction screen for each task is presented in Appendix A.

The in-depth explanations page provided additional information for two of the tasks included in the previous try-out section. Each of these tasks could be accessed by clicking on the task image. Clicking on the image led to a page with an in-depth task description and examples of actual performances from preservice teachers. The in-depth description included a short explanation of the teaching competencies that were required by the task, why these competencies matter, how the performance is evaluated, and the types of challenges that typically characterize preservice teachers' responses. Four examples of actual preservice teacher performances were also provided. Viewing the performances provided a real-time illustration of different learning needs that teacher educators are likely to encounter when using tasks like these as part

Table 1 RLA Task Characteristics

Task	Task of teaching	Content topic	Grade
RLA 1	Explain process/concept using student work	Persuasive writing and use of examples	4
RLA 2	Model or explain reading process or concept	Grapho-phonemic segmenting	K
RLA 3	Explain process/concept using student work	Poetry and evidence for theme	6
RLA 4	Model how to improve writing	Narrative writing sensory details	4
RLA 5	Explain process/concept using student work	Reading fluency	3
RLA 6	Model or explain reading process or concept	Word roots and affixes	4

Note. K = kindergarten; RLA = reading language arts.

Table 2 Mathematics Task Characteristics

Task	Task of teaching	Content topic	Grade
MATH 1	Explain student error and model correct method	Two-digit subtraction with regrouping	2
MATH 2	Explain and model mathematical concepts	Line of symmetry	4
MATH 3	Represent concepts using mathematical model	Partitive and measurement models of division	3
MATH 4	Compare student methods	Two-digit addition with regrouping	2
MATH 5	Explain problems using a representation	Fraction word problems and number line	4
MATH 6	Explain student error and model correct method	One-digit addition	1

Note. MATH = mathematics.

of their teacher preparation coursework. Each of these performances is linked to a summary of the key features of the performance. The task descriptions and the performance details for each of the two tasks in this section are provided in Appendix B. Once teacher educators completed their task review, they navigated to the final page of the website, where they accessed a link to the survey.

Survey

Teacher educators accessed and completed the survey online. The survey included two primary sections. The first section asked teacher educators questions about the tasks they reviewed and potential uses for the tasks. Most questions presented a statement about the performance tasks or possible uses and then presented a set of Likert scale response choices (e.g., Agree, Somewhat Agree, Somewhat Disagree, Disagree). All Likert survey questions are included in full in the tables presented in the results section below. The second section of the survey asked participants to provide information about themselves, their educational background, their K-12 teaching experience, their role as a teacher educator, and their current institution.

Teacher Educator Sample

Sixty-four teacher educators participated in the study. The sample was predominantly female (88%) and White (92%). A large majority of the sample had previously taught in a K-12 setting for 4 or more years (80%), with more than half of the participants having taught the K-12 subject areas and grade levels that were the focus of the performance tasks they reviewed for this study (Table 3).

As shown in Table 4, over 90% of the teacher educators had 4 or more years of experience working with preservice teachers. Half of the sample reported teaching mathematics methods and over half taught RLA methods. A quarter of the sample focused on preparing preservice teachers to be special educators. Over 90% of participants taught methods courses, with the majority also conducting research on teacher education or providing instruction in content for teachers. Roughly half of the teacher educators taught at educator preparation programs (EPPs) they described as Research 1 or 2 and half of the sample primarily focused on teacher preparation. Nearly 10% of teacher educators taught at EPPs described as minority serving institutions.

Fifty-five unique educator preparation programs were represented by the study sample. While five of the teacher educators came from a single program, no other institution included more than two educators. The 55 teacher educator

Table 3 Demographic and K-12 Teaching Experience ($n = 64$)

Characteristic	%
Gender	
Female	88%
Male	13%
Race/ethnicity[a]	
Asian or Asian American	6%
Black or African American	2%
Hispanic/Latino	6%
White	92%
Degree obtained	
Bachelor's	100%
Master's	98%
Doctoral	92%
Previously K-12 classroom teacher	
Yes	94%
No	6%
Years spent teaching in K-12 classroom	
0–3 years	14%
4–10 years	55%
More than 10 years	25%
K-12 grades taught[a]	
K–2	61%
3–5	66%
6–8	47%
9–12	31%
K-12 subjects taught[a]	
Reading or language arts	72%
Mathematics	84%
Social studies	67%
Science	67%
Special education	27%
Other	16%

[a] Participants could mark more than one choice.

programs represented in the study were geographically diverse with 21 (38%) located in the Midwest, 14 (25%) located in the Northeast, 12 (22%) located in the South, and eight (15%) located in the West.

Results

The presentation of results follows a similar pattern for each of the question groupings administered on the survey. First, the proportion of teacher educators who chose each of the Likert options for each of the individual questions is summarized and reported in an accompanying table. To facilitate review, the questions in each set are ordered from the ones receiving the highest level of endorsement to the ones receiving lower levels of endorsement.[2] Next, responses for RLA educators ($n = 33$) and mathematics educators ($n = 31$) are compared and responses for general educators ($n = 47$) and special educators ($n = 17$) are compared. For the group comparisons, the proportion selecting the strongest endorsement is provided (i.e., the proportions selecting Agree, Very Useful, or Very Likely).

Do Teacher Educators View the Performance Tasks as Relevant and Authentic?

A large majority of teacher educators strongly agreed that the teaching competencies that are the focus of the performance tasks are important, feel like actual teaching, and are a focus of their teacher preparation program. While still a strong majority, the level of endorsement was somewhat lower for using the performance tasks as part of teacher licensure (Table 5).

In the group comparisons, RLA educators were notably more likely than mathematics educators to view the performance tasks as important for effective teaching, a focus of their teacher education coursework, and important to use as

Table 4 Teacher Education Experience and Institution Information ($n = 64$)

Characteristic	%
Institution type[a]	
Research 1 or 2	48%
Primary emphasis is professional preparation	50%
Minority-serving institution	9%
Institution level	
Baccalaureate college	20%
Master's college/university	30%
Doctoral university	48%
Other	2%
Role at institution[a]	
Teacher education director, program chair, or program lead	30%
Instructor in teaching methods	94%
Instructor in content for teachers	52%
Practicum or field supervisor	41%
Research on teacher education or teacher learning	69%
Other	6%
Subjects taught to preservice teachers[a]	
Reading or language arts	58%
Mathematics	50%
Special education	25%
Other	16%
Years spent teaching preservice teachers	
0–3 years	9%
4–10 years	48%
More than 10 years	43%

[a] Participants could select more than one answer.

Table 5 Relevance and Authenticity: Proportion of Teacher Educators Selecting Each Response Category

	Educators selecting each response category ($n = 64$)			
The teaching knowledge, skills, and abilities required by the performance tasks:	Agree	Somewhat agree	Somewhat disagree	Disagree
are important for beginning teachers.	92%	8%	0%	0%
are important for effective teaching.	89%	9%	2%	0%
felt like ones used in actual teaching.	81%	13%	6%	0%
are a focus of my teacher preparation coursework.	73%	25%	2%	0%
should be assessed as part of teacher licensure.	59%	34%	6%	0%

Table 6 Relevance and Authenticity: Proportion of Each Educator Group Selecting Agree

	Proportion of each educator group selecting agree			
The teaching knowledge, skills, and abilities required by the performance tasks:	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
are important for beginning teachers.	94%	90%	89%	100%
are important for effective teaching.	94%	84%	85%	100%
felt like ones used in actual teaching.	82%	81%	77%	94%
are a focus of my teacher preparation coursework.	79%	68%	66%	94%
should be assessed as part of teacher licensure.	67%	52%	51%	82%

Note. RLA = reading language arts.

Table 7 Recommended Use for Preservice Teachers: Proportion Selecting Each Response Category for All Educators

	Educators selecting each response category ($n = 64$)			
Preservice teachers would benefit from:	Agree	Somewhat agree	Somewhat disagree	Disagree
receiving feedback on their own performance.	97%	3%	0%	0%
discussing performances with other preservice teachers.	88%	11%	2%	0%
completing performance tasks like these as homework assignments.	86%	13%	2%	0%
reflecting on their own performance.	86%	14%	0%	0%
scoring their own performance using a provided rubric.	67%	31%	0%	2%
scoring each other's performances using a provided rubric.	55%	30%	8%	8%

Table 8 Recommended Use for Preservice Teachers: Proportion Selecting Agree for Each Educator Group

	Educators selecting agree			
Preservice teachers would benefit from:	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
receiving feedback on their own performance.	94%	100%	96%	100%
discussing performances with other preservice teachers.	82%	94%	87%	88%
completing performance tasks like these as homework assignments.	88%	84%	83%	94%
reflecting on their own performance.	85%	87%	83%	94%
scoring their own performance using a provided rubric.	82%	52%	62%	82%
scoring each other's performances using a provided rubric.	58%	52%	47%	76%

Note. RLA = reading language arts.

part of licensure. Special educators were consistently more likely than general educators to view the performance tasks as assessing skills that are relevant and authentic, with particularly large differences in their use of these tasks in their teacher education coursework and their views on using these tasks as part of teacher licensure assessments (Table 6).

What Are Recommended Uses of the Performance Tasks for Preservice Teachers?

Almost every teacher educator agreed that preservice teachers would benefit from receiving feedback on their own performance. While endorsement was still strong, it fell slightly for other uses including discussing performances, completing the performances as homework, and reflecting on their own performance. A slim majority of teacher educators agreed that preservice teachers would benefit from using a rubric to score their own performance or the performances or other preservice teachers (Table 7).

Mathematics and RLA educators had notably different levels of agreement for two of the questions. Just over 10% more mathematics educators than RLA educators felt that preservice teachers would benefit from discussing the tasks with other preservice teachers. However, roughly a third more RLA educators felt that preservice teachers would benefit from scoring their own performances using a rubric. For the most part, special educators agreed more often than general educators that preservice teachers would benefit from using these tasks. Roughly 10% more special educators felt that preservice teachers would benefit from using the performance tasks as part of their homework and for personal reflection. Special educators were 20% more likely to agree that preservice teachers would benefit from using a rubric to score their own performance, and nearly a third more special educators felt that preservice teachers would benefit from scoring each other's performances (Table 8).

What Are Recommended Uses of the Performance Tasks for Teacher Educators?

The teacher educator responses on recommended uses of these tasks were spilt. The endorsement was very strong for formative assessment uses, including providing feedback. However, summative uses as an end-of-course evaluation and to assess preservice teachers' readiness for the classroom was notably lower, with roughly only half of teacher educators providing a strong endorsement (Table 9).

Table 9 Recommended Uses for Teacher Educators: Proportion Selecting Each Response Category for All Participants

	Educators selecting each response category ($n = 64$)			
Teacher educators could benefit from using performance tasks like these:	Agree	Somewhat agree	Somewhat disagree	Disagree
to provide feedback to preservice teachers.	92%	8%	0%	0%
as formative assessments to inform understanding of what preservice teachers know and still need to learn.	89%	9%	2%	0%
as an end-of-course evaluation.	52%	34%	5%	9%
to assess whether preservice teachers are ready to teach actual students.	47%	38%	9%	6%

Table 10 Recommended Uses for Teacher Educators: Proportion Selecting Agree for Each Educator Group

	Educators selecting agree			
Teacher educators could benefit from using performance tasks like these:	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
to provide feedback to preservice teachers.	94%	90%	91%	94%
as formative assessments to inform understanding of what preservice teachers know and still need to learn.	91%	87%	85%	100%
as an end-of-course evaluation.	52%	52%	45%	71%
to assess whether preservice teachers are ready to teach actual students.	45%	48%	38%	71%

Note. RLA = reading language arts.

There was no noteworthy difference in the proportion of RLA and mathematics educators agreeing that teacher educators could benefit from using the performance tasks. However, there was a substantial difference between general and special educators. Fifteen percent more special educators saw value in using the tasks as formative assessment tools. Roughly a third more special educators saw value in using these tasks as summative assessments for end-of-course evaluation and for determining preservice teachers' readiness or entering the classroom (Table 10).

What Resources Would Support the Use of Performance Tasks in Methods Courses?

Teacher educators strongly endorsed most of the listed supporting resources, including rubrics for use by teacher educators and preservice teachers in scoring performances, sample performances that illustrate levels of the rubric, an online delivery platform, and provision of additional performance tasks. About 60% of teacher educators felt that diagnostic and summative assessments would be very useful. However, only about half the teacher educators felt that explanations for why the tasks are important for effective teaching would be very useful (Table 11).

The RLA teacher educators were about 10% more likely than mathematics teacher educators to feel that sample performances would be very useful. Mathematics teacher educators were about 10% more likely than RLA teacher educators to say that additional performance tasks tailored to their coursework would be very useful. Special educators were 10 to 15% more likely than general educators to feel that rubrics to support teacher educators scoring performances, an online platform for administering and viewing performances, and both diagnostic and summative assessments would be very useful. Nearly a third more special educators felt that explanations of why the tasks are important for effective teaching would be very useful (Table 12).

Are Teacher Educators Likely to Use Performance Simulations in their Methods Coursework?

Teacher educators were asked two sets of questions about their use of performance tasks in their methods coursework. The first set of questions (Tables 13 and 14) asked about whether they currently use any form of simulated performance activity. The second set of questions asked about the likelihood that either they or their colleagues might use the new type of short performance task reviewed for the current survey (Tables 15 and 16).

In respect to their current use of simulated performance activities, approximately half of teacher educators reported that they often engaged preservice teachers in peer teaching, microteaching, rehearsals, or other types of approximations

Table 11 Usefulness of Supporting Resources: Proportion Selecting Each Response Category for All Participants

	Educators selecting each response category ($n = 64$)		
How useful would each of the following resources be for supporting the use of the performance tasks in a methods course?	Very useful	Somewhat useful	Not useful
Sample preservice teacher performances to illustrate levels of the rubric	89%	11%	0%
Additional performance tasks focused on content specifically tailored to the coursework	84%	11%	5%
Rubrics and explanations designed for preservice teachers to use in scoring performances	81%	17%	2%
Rubrics and explanations designed for teacher educators to use in scoring performances	80%	20%	0%
Online platform for assigning performance tasks and reviewing preservice teacher performances	78%	20%	2%
Summative assessments that include performance tasks like these for use after unit or course completion	64%	25%	11%
Diagnostic assessments that include performance tasks like these to determine what preservice teachers know and need to learn	61%	30%	9%
Descriptions of why performance tasks are important for effective teaching	55%	38%	8%

Table 12 Usefulness of Supporting Resources: Proportion Selecting Very Useful for Each Educator Group

	Educators selecting very useful			
How useful would each of the following resources be for supporting the use of the performance tasks in a methods course?	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
Rubrics and explanations designed for teacher educators to use in scoring performances	79%	81%	77%	88%
Rubrics and explanations designed for preservice teachers to use in scoring performances	85%	77%	79%	88%
Sample preservice teacher performances to illustrate levels of the rubric	94%	84%	87%	94%
Online platform for assigning performance tasks and reviewing preservice teacher performances	82%	74%	74%	88%
Descriptions of why performance tasks are important for effective teaching	55%	55%	47%	76%
Diagnostic assessments that include performance tasks like these to determine what preservice teachers know and need to learn	64%	58%	57%	71%
Summative assessments that include performance tasks like these for use after unit or course completion	61%	68%	60%	76%
Additional performance tasks focused on content specifically tailored to the coursework	79%	90%	83%	88%

Note. RLA = reading language arts.

that can be enacted through live actor simulation or forms of role play. However, only 5% reported that they often used some form of simulated performance technology (Table 13).

There was no notable difference between how often RLA and mathematics educators used either live actor/role play approximations or simulated classroom environments. Special educators, however, were over 30% more likely than general educators to report that they often engaged preservice teachers in live actor/role play approximations and 10% more likely to engage preservice teachers in simulated classroom approximations (Table 14).

In respect to the short performance tasks, nearly three quarters of teacher educators said they would be very likely to use tasks like these in their methods coursework. However, just over a third reported that their colleagues in their current program would be very likely to use these performance tasks (Table 15).

While the RLA teacher educators and mathematics teacher educators showed little difference in their own likelihood of using the performance tasks, the RLA teacher educators were about 15% more likely than mathematics teacher educators to report that their colleagues would be likely to use the performance tasks. Special educators were about 15% more likely than general educators to say they were very likely to use these tasks. Over a third more special educators said that their colleagues would be very likely to use these performance tasks (Table 16).

Which Tasks Do Reading Language Arts Teacher Educators Think Are the Most and Least Useful in Teacher Preparation?

When asked which performance tasks they thought were most useful, roughly 20 to 30% of RLA educators chose persuasive writing and use of examples, grapho-phonemic segmenting, or narrative writing sensory details. Interestingly,

Table 13 Use of Approximations of Practice in Methods Course

	Proportion selecting each response category ($n = 64$)			
	Often	Sometimes	Rarely	Never
How often do you engage your preservice teachers in peer teaching, microteaching, teaching rehearsals, or other similar approximations of practice?	47%	42%	9%	2%
How often do you have your preservice teachers use a simulated classroom (i.e., Mursion simulation or similar environment)?	5%	2%	17%	77%

Table 14 Use of Approximations of Practice in Methods Course: Proportion Selecting Often for Each Educator Group

	Educators selecting often			
	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
How often do you engage your preservice teachers in peer teaching, microteaching, teaching rehearsals, or other similar approximations of practice?	48%	45%	38%	71%
How often do you have your preservice teachers use a simulated classroom (i.e., Mursion simulation or similar environment)?	3%	6%	2%	12%

Note. RLA = reading language arts.

roughly the same proportion choose grapho-phonemic segmenting as least useful, suggesting that this performance task was viewed very differently by the teacher educators in this study (Table 17). Further breakdowns are provided in Table 18 for RLA educators who identified as general education and special education. However, given the small number of educators who identified as RLA special education, these results should be viewed as at best suggestive.

Which Tasks Do Mathematics Teacher Educators Think Are the Most and Least Useful in Teacher Preparation?

Nearly a third of the mathematics teacher educators identified fraction word problems and number line as the most useful task. The remaining mathematics educators were relatively evenly divided in choosing two-digit subtraction with regrouping, partitive and measurement models of division, two-digit addition with regrouping, or one-digit addition as most useful. None of the mathematics teacher educators choose line of symmetry as most useful (Table 19). Further breakdowns are provided in Table 20 for mathematics educators who identified as general education and special education. However, given the small number of teacher educators who identified as mathematics special education, these results should be viewed as at best suggestive.

Discussion

The most significant finding from the teacher educator survey was the overall very positive response to both the performance tasks and their value for a range of different uses. Ninety percent or more of the teacher educators felt that the competencies assessed by these types of performance tasks are important for beginning teachers and for effective teaching, that preservice teachers would benefit from discussing their performances with other preservice teachers and receiving feedback, and that teacher educators would benefit from using these types of tasks to support formative assessment processes and as tools for providing feedback.

More than three quarters of teacher educators agreed that these types of tasks felt like actual teaching and are currently a focus of their teacher preparation and that preservice teachers would benefit from reflecting on their own performances. They felt that rubrics associated with these types of tasks would be very useful for both teacher educators and preservice teachers and that they would benefit from having additional tasks like these to use in their courses. This level of endorsement was even higher when considering teacher educators who somewhat agreed with statements about the value or potential uses for these types of tasks.

Table 15 Likelihood of Including Tasks in Methods Courses: Proportion Selecting Each Response Category for All Participants

	Educators selecting each response category ($n = 64$)		
	Very likely	Somewhat likely	Not likely
How likely are you to incorporate performance tasks like the ones you reviewed into your own methods coursework?	72%	27%	2%
How likely are your colleagues in your current program to incorporate performance tasks like the ones you reviewed into their methods coursework?	39%	56%	5%

Table 16 Likelihood of Including Tasks in Methods Courses: Proportion Selecting Very Likely for Each Educator Group

	Educators selecting very likely			
	RLA ($n = 33$)	Math ($n = 31$)	General ($n = 47$)	Special ($n = 17$)
How likely are you to incorporate performance tasks like the ones you reviewed into your own methods coursework?	73%	71%	68%	82%
How likely are your colleagues in your current program to incorporate performance tasks like the ones you reviewed into their methods coursework?	45%	32%	30%	65%

Note. RLA = reading language arts.

However, not all uses for the performance tasks were as highly endorsed by the teacher educators. For example, lower proportions of teacher educators agreed that these tasks should be used in teacher licensure, as end-of-course evaluations, to assess whether preservice teachers are ready to teach actual students, or for purposes of diagnostic assessment. Teacher educators were also less supportive of having preservice teachers use a rubric to score their own performance or score the performances of other preservice teachers. It is important to recognize, however, that the lower rates of endorsement could be due to how these questions were phrased. For example, these questions might have received a different level of endorsement if the questions had specified that the purpose for scoring was to support self-reflection or peer feedback.

Differences between RLA and mathematics educators were observed on a relatively small number of specific questions. Neither RLA nor mathematics educators stood out as providing stronger endorsement across the full set of questions. However, there was notably stronger general endorsement by special educators compared to general educators. For example, special educators were much more likely than general educators to see value in using these types of tasks in licensure testing, as end-of-course evaluations, and to assess whether preservice teachers are ready to teach actual students. Special educators were also much more likely to see value in preservice teachers scoring their own performances and the performances of other preservice teachers using a rubric. Special educators were also much more likely than general educators to see value in using these types of tasks in their own methods coursework and to say that their special education colleagues would also support the use of these types of tasks.

Limitations of Survey Results

There are a few limitations to this study that need to be considered when interpreting these results. First, the quality of the survey responses almost certainly was dependent on the teacher educators carefully reviewing the example performance tasks presented on the website. While the teacher educators were directed to carefully read the descriptions on the website about intended uses for the tasks, try out the six performance tasks as if they were a preservice teacher in their methods class, and then carefully review the two in-depth task descriptions and four sample performances, it is possible that some number of teacher educators did not invest the time and energy to carefully review all of these materials and responded to the survey based on an incomplete or shallow understanding of each of the performance tasks.

In addition, the task review process relied on teacher educators imagining or thinking their way into how these tasks might be used. While the expanded in-depth task explanations (see Appendix B) provided some insight into the task design and potential preservice teacher responses, the teacher educators were still left to imagine particular uses and

Table 17 RLA Teacher Educators' Selection of Most and Least Useful Tasks for Use in Teacher Preparation ($n = 33$)

Which task do you think would be the most and least useful in teacher preparation?	Most useful	Least useful
Persuasive writing and use of examples (RLA 1)	27%	3%
Grapho-phonemic segmenting (RLA 2)	24%	21%
Narrative writing sensory details (RLA 4)	21%	12%
Poetry and evidence for theme (RLA 3)	12%	24%
Reading fluency (RLA 5)	9%	30%
Word roots and affixes (RLA 6)	6%	9%

Note. RLA = reading language arts.

Table 18 RLA Teacher Educators' Selection of Most and Least Useful Tasks for Use in Teacher Preparation for General and Special Educators ($n = 33$)

	General educators ($n = 21$)		Special educators ($n = 12$)	
Which task do you think would be the most and least useful in teacher preparation?	Most useful	Least useful	Most useful	Least useful
Persuasive writing and use of examples (RLA 1)	29%	5%	25%	0%
Grapho-phonemic segmenting (RLA 2)	29%	19%	17%	25%
Narrative writing sensory details (RLA 4)	24%	10%	17%	17%
Poetry and evidence for theme (RLA 3)	10%	19%	17%	33%
Reading fluency (RLA 5)	10%	38%	8%	17%
Word roots and affixes (RLA 6)	0%	4%	17%	8%

Note. RLA = reading language arts.

responses. Future work will seek to address this limitation through pilots where teacher educators are provided curriculum plans and resources for using these tasks in their course.

A second limitation is the relatively small convenience sample of teacher educators. The goal of this study was to gather evidence on how teacher educators in general might view the value of these performance tasks. With this goal in mind, efforts were made to recruit teacher educators from different types of universities and colleges that served students from a range of backgrounds. We also sought out teacher educators who themselves represented different racial and cultural groups, academic backgrounds, K-12 teaching experience, and experience as a teacher educator. While we were at least moderately successful in these recruiting goals, the results were almost certainly influenced by the particular teacher educators in the sample and should not be taken to be representative of teacher educators in general or teacher educators who work in particular contexts of interest.

Finally, the sample is relatively small for the purposes of reporting descriptive survey statistics. The proportion of teacher educators endorsing any particular response could be influenced by outlier perspectives. This threat to interpreting the results is particularly important to consider in the group comparisons.

While these limitations need to be considered carefully, the generally high level of endorsement indicates that teacher educators value the new performance tasks and see an important role for them as part of teacher preparation. These findings provide strong backing to support continued development and research on the use of short performance tasks to support teacher educators in providing preservice teachers with opportunities to develop knowledge and skill used in core teaching practices.

Considering a Wider Range of Assessment Uses for Performance Tasks

The focus of this study was on gathering input from teacher educators on the potential for using short performance tasks as part of their teacher preparation methods coursework. However, the potential uses for these performance tasks go substantially beyond providing opportunities to learn core practices as part of teacher preparation and also include uses such as diagnosing professional learning needs; delivering end-of-course evaluations; providing initial licensure; and studying, evaluating, and promoting teaching quality (Phelps, Bridgeman et al., 2020; Phelps & Sykes, 2020). Considering a more comprehensive set of uses could provide additional context for understanding the value of these tasks as tools

Table 19 Mathematics Teacher Educators' Selection of Most and Least Useful Tasks for Use in Teacher Preparation ($n = 31$)

Which task do you think would be the most and least useful in teacher preparation?	Most useful	Least useful
Fraction word problems and number line (MATH 5)	29%	6%
Two-digit subtraction with regrouping (MATH 1)	19%	16%
Partitive and measurement models of division (MATH 3)	19%	0%
Two-digit addition with regrouping (MATH 4)	19%	0%
One-digit addition (MATH 6)	13%	3%
Line of symmetry (MATH 2)	0%	74%

Note. MATH = mathematics.

Table 20 Mathematics Teacher Educators' Selection of Most and Least Useful Tasks for Use in Teacher Preparation for General and Special Educators ($n = 31$)

	General educators ($n = 26$)		Special educators ($n = 5$)	
Which task do you think would be the most and least useful in teacher preparation?	Most useful	Least useful	Most useful	Least useful
Fraction word problems and number line (MATH 5)	27%	8%	40%	0%
Two-digit subtraction with regrouping (MATH 1)	15%	19%	40%	0%
Partitive and measurement models of division (MATH 3)	23%	0%	0%	0%
Two-digit addition with regrouping (MATH 4)	23%	0%	0%	0%
One-digit addition (MATH 6)	12%	4%	20%	0%
Line of symmetry (MATH 2)	0%	69%	0%	100%

Note. MATH = mathematics.

to support professional learning. To provide additional context, we briefly discuss a logic model that lays out how an integrated set of use cases for performance tasks might support teacher learning and improvement in teaching quality. While the logic model can be applied in general to all performance tasks, particular components of the model are best addressed by short, standardized performance tasks.

The logic model presented in Figure 3 incorporates the main features of the Grossman, Compton, et al. (2009) model of professional preparation. The first panel represents a cycle of learning around core practices. Preservice teachers have opportunities to learn about and study a practice, including how the practice is related to actual instructional interaction with students. Teachers try out the practice, receive feedback, and refine both their understanding and skill. The second panel of the logic model focuses on the importance of ensuring that teachers have learned to *do* these practices successfully. These evaluations provide one source of evidence to guide teacher educators in making decisions about whether prospective teachers are ready for actual classroom teaching or need additional opportunities to practice and learn. The third and fourth panels of the logic model indicate key outcomes of practice-based professional learning. Developing competence in critical core practices provides teachers with a strong practice-based foundation for engaging in effective instruction that ultimately supports student learning. The logic model also illustrates how teacher learning, effective teaching, and student learning outcomes all organize around the core practices that make up the work of teaching.

Assessment plays a critical role at each stage of this model. The performance tasks function as a type of professional formative assessment, providing opportunities to try out a practice and receive feedback and as needed structure for additional learning opportunities. These same tasks can be used as part of diagnostic assessments to determine what teachers can already do or still need to learn. Based on these assessment results, teachers can receive additional training or move to apply these skills in the more complex work of actual classroom instruction. The assessment tasks can also be used as outcome measures to evaluate professional preparation and identify areas where programs need improvement. Finally, performance assessments can be used as one of multiple measures of teaching quality to understand better how foundational skills contribute to both teaching quality and student learning.

Performance assessment also provides an important empirical basis for supporting the argument that particular core practices and the associated tasks used to simulate or approximate teaching are indeed component elements of effective instruction. This type of validity evidence provides strong support for focusing on particular core practices as part of practice-based professional preparation and as outcome measures to diagnose teacher learning and evaluate teacher education and professional development programs.

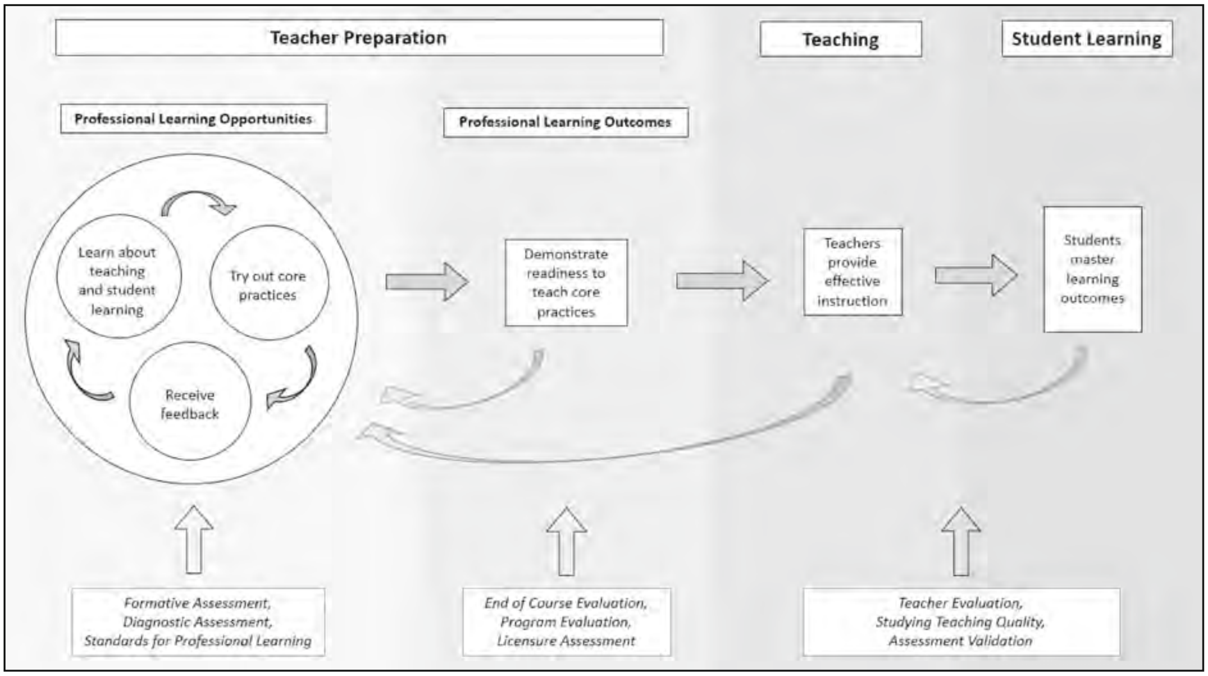


Figure 3 Theory of action for practice-based teacher learning.

It is also important to recognize that many, and perhaps most, current approaches to teacher preparation are missing several components specified in this logic model, in large part due to a lack of suitable assessments of teachers' performance (Phelps & Bridgeman, 2022; Phelps & Sykes, 2020). In many instances, teachers receive few opportunities to try out core practices as part of professional learning. Instead, most professional learning is organized around opportunities for teachers to develop *knowledge*: knowledge of students' cognitive development, knowledge of effective instructional approaches, or foundational knowledge about subject matter itself. While building knowledge is clearly important, a knowledge-only approach provides limited scaffolding for learning and developing foundational teaching skills. Without structured opportunities to practice and receive feedback, teachers are left to learn and refine their teaching skills in the context of their own classroom, arguably at the expense of their students (Ball & Forzani, 2009, 2011). A primary focus on knowledge coupled with a general lack of attention to core practices is likely to lead to less effective instruction and ultimately lower levels of student achievement.

The lack of valid and reliable assessments of teaching performance also has an impact across this model. Teacher educators are essentially left to invent their own ad hoc simulations to approximate practice (e.g., asking teachers to microteach an instructional skill to their peers). These locally developed simulations are unlikely to be standardized in ways that allow sharing and comparison across programs and even instructors, limiting potential for more systematic improvement in preparation. Further, the standards of quality that are used to evaluate performances are unlikely to have the same rigor as the carefully developed rubrics used to evaluate and score standardized performance assessment tasks. A general absence of valid and reliable performance assessments means that there are few, if any, tools for ensuring that teachers have actually learned to carry out or do critical core practices.

Looking Forward

We view the research and development activities discussed in this report as part of a larger agenda to support professional learning and improve teaching quality. To make progress on these goals, it is important to conceptualize and develop performance tasks that are valid approximations of teaching and to avoid the potential risks associated with decomposing teaching into practices that are either not defensible or trivialize the complexity of teaching. Mary Kennedy (2016), for example, raised the concern that the core practice movement could focus on teaching behaviors in ways that are problematic and counterproductive:

> Throughout our history, we have tried to define the practice of teaching in terms of lists of specific bodies of knowledge or lists of specific behaviors rather than in terms of what those behaviors are intended to achieve … We have misplaced our focus on the *actions* we see; when what is needed is a focus on the purposes those actions serve. (p. 9)

Ken Zeichner (2012) raised a related concern when he pointed to the pitfall of treating teaching as mere technical work:

> There is a danger of narrowing the role of teachers to that of technicians who are able to implement a particular set of teaching strategies, but who do not develop the broad professional vision (deep knowledge of their students and of the cultural contexts in which their work is situated), and the relational skills they need to be successful in the complex institutional settings in which they will work. (p. 379)

These and other similar critiques of the core-practice movement need to be carefully considered. Making progress will require advancing our understanding of key questions related to the teaching and learning of professional practice. What makes a practice core to teaching and teacher learning? What is the appropriate size of a core practice? How can we approximate core practices in ways that are true to the work of teaching *and* manageable for use in supporting teacher learning? What methods can be used to provide evidence of high-quality enactment of these core practices? How should core practices be sequenced in a curriculum of professional preparation? What tools, including assessments, do teacher educators need to support novice teachers' professional learning? We see the current line of work as one important step toward making a start on these questions and toward better understanding how performance tasks can be used to support teacher learning and ultimately teaching quality.

Notes

1 Mathematics and RLA websites can be access at the following URLs: https://researchtech1.ets.org/toeflsvr/tpServerRoot/wb/FactWeb2021/apwWebPage.html?M (mathematics) https://researchtech1.ets.org/toeflsvr/tpServerRoot/wb/FactWeb2021/apwWebPage.html?R (RLA)

2 Readers interested in reviewing the survey with questions in the order of administration can contact the corresponding author (gphelps@ets.org).

References

Alston, C. L., Danielson, K. A., Dutro, E., & Cartun, A. (2018). Does a discussion by any other name sound the same? Teaching discussion in three ELA methods courses. *Journal of Teacher Education, 69*(3), 225–238. https://doi.org/10.1177/0022487117715227

Amador, J. M. (2017). Preservice teachers' video simulations and subsequent noticing: A practice-based method to prepare mathematics teachers. *Research in Mathematics Education, 19*(3), 217–235. https://doi.org/10.1080/14794802.2017.1315317

Averill, R., Drake, M., Anderson, D., & Anthony, G. (2016). The use of questions within in-the-moment coaching in initial mathematics teacher education: Enhancing participation, reflection, and co-construction in rehearsals of practice. *Asia-Pacific Journal of Teacher Education, 44*(5), 486–503. https://doi.org/10.1080/1359866X.2016.1169503

Ball, D. L., & Bass, H. (2003). Toward a practice-based theory of mathematical knowledge for teaching. In E. Simmt & B. Davis (Eds.), *Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group* (pp. 3–14). CMESG/GCEDM.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Towards a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). Jossey-Bass.

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

Ball, D. L., & Forzani, F. M. (2011). Building a common core for learning to teach: And connecting professional learning to practice. *American Educator, 35*(2), 17–21.

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

Bardach, L., Klassen, R. M., Durksen, T. L., Rushby, J. K., Bostwick, K. C. P., & Sheridan, L. (2021). The power of feedback and reflection: Testing an online scenario-based learning intervention for student teachers, *Computers & Education*, *169*, Article 104194. https://doi.org/10.1016/j.compedu.2021.104194

Campbell, M. P., Baldinger, E. E., & Graif, F. (2020). Representing student voice in an approximation of practice: Using planted errors in coached rehearsals to support teacher candidate learning. *Mathematics Teacher Educator, 9*(1), 23–49. https://doi.org/10.5951/MTE.2020.0005

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis, 42*(2), 208–231. https://doi.org/10.3102%2F0162373720906217

Davis, E. A., Kloser, M., Wells, A., Windschitl, M., Carlson, J., & Marino, J.-C. (2017). Teaching the practice of leading sense-making discussions in science: Science teacher educators using rehearsals. *Journal of Science Teacher Education, 28*(3), 275–293. https://doi.org/10.1080/1046560X.2017.1302729

Delamarre, A., Shernoff, E., Buche, C., Frazier, S., Gabbard, J., & Lisetti, C. (2021). The interactive virtual training for teachers (IVT-T) to practice classroom behavior management. *International Journal of Human-Computer Studies*, *152*, Article 102646. https://doi.org/10.1016/j.ijhcs.2021.102646

Dieker, L. A., Rodriguez, J. A., Lignugaris/Kraft, B., Hynes, M. C., & Hughes, C. E. (2014). The potential of simulated environments in teacher education: Current and future possibilities. *Teacher Education and Special Education*, *37*(1), 21–33. https://doi.org/10.1177/0888406413512683

Fogo, B. (2014). Core practices for teaching history: The results of a delphi panel survey. *Theory & Research in Social Education, 42*(2), 151–196. https://doi.org/10.1080/00933104.2014.902781

Forzani, F. M. (2014). Understanding "core practices" and "practice-based" teacher education: Learning from the past. *Journal of Teacher Education*, *65*(4), 357–368. https://doi.org/10.1177/0022487114533800

Franke, M. L., Kazemi, E., & Battey, D. (2007). Understanding teaching and classroom practice in mathematics, In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). McMillian.

Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *The Elementary School Journal, 89*(3), 253–300. https://doi.org/10.1086/461577

Gitomer, D. H., Phelps, G., Weren, B. H., Howell, H., & Croft, A. J. (2015). Evidence on the validity of content knowledge for teaching assessments. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 493–528). Wiley. https://doi.org/10.1002/9781119210856.ch15

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111*(9), 2055–2100. https://doi.org/10.1177/016146810911100905

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching, 15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education, 119*(3), 445–470. https://doi.org/10.1086/669901

Grossman. P., Schneider Kavanagh, S., & Pupik Dean, C. G. (2018). The turn towards practice-based teacher education: Introduction to the work of the core practice consortium. In P. Grossman (Ed.), *Teaching core practices in teacher education* (pp. 1–14). Harvard Education Press.

Hardin, S., & Freeman-Green, S. (2015). Infusing culturally responsive strategies in STEM instruction for special education teachers. In T. Bousfield, M. Hynes, C. Hughes, C. Straub, L. Dieker, & K. Ingrahm (Eds.), *Proceedings of 3rd National TLE TeachLivE™ Conference 2015: Dissecting education* (pp. 19–22). University of Central Florida.

Herbst, P., & Kosko, K. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J.-J. Lo, K. R. Leatham, & L. R. Van Zoest (Eds.), *Research trends in mathematics teacher education* (pp. 23–45). Springer. https://doi.org/10.1007/978-3-319-02562-9_2

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *Elementary School Journal*, *105*(1), 11–30. https://doi.org/10.1086/428763

Kavanagh, S. S., Metz, M., Hauser, M., Fogo, B., Taylor, M. W., & Carlson, J. (2020). Practicing responsiveness: Using approximations of teaching to develop teachers' responsiveness to students' ideas. *Journal of Teacher Education, 71*(1), 94–107. https://doi.org/10.1177/0022487119841884

Kazemi, E., Ghousseini, H., Cunard, A., & Turrou, A. C. (2015). Getting inside rehearsals: Insights from teacher educators to support work on complex practice. *Journal of Teacher Education, 67*(1), 18–31. https://doi.org/10.1177/0022487115615191

Kelley, M. J., & Wenzel, T. (2018). A parent-teacher reading conference project: Using a virtual environment (TeachLivE™) to improve elementary pre-service teachers' conferencing skills. *The Reading Professor*, *41*(1), 48–59.

Kennedy, M. M. (1999). The role of preservice teacher education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of teaching and policy* (pp. 54–86). Jossey Bass.

Kennedy, M. (2016). Parsing the practice of teaching. *Journal of Teacher Education, 67*(1), 6–17. https://doi.org/10.1177/0022487115614617

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*(5), 845–861. https://doi.org/10.1177/0013164407313369

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. *Journal of Research in Science Teaching, 51*(9), 1185–1217. https://doi.org/10.1002/tea.21171

Kloser, M., Wilsey, M., Madkins, T. C., & Windschitl, M. (2019). Connecting the dots: Secondary science teacher candidates' uptake of the core practice of facilitating sensemaking discussions from teacher education experiences. *Teaching and Teacher Education, 80,* 115–127. https://doi.org/10.1016/j.tate.2019.01.006

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM Mathematics Education, 40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9

Lampert, M. (1985). How do teachers manage to teach? Perspectives on problems in practice. *Harvard Educational Review, 55*(2), 178–194. https://doi.org/10.17763/haer.55.2.56142234616x4352

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *Elementary School Journal, 109*(5), 491–509. https://doi.org/10.1086/596998

Macleod, G. (1987). Microteaching: End of a research era? *International Journal of Educational Research*, *11*(5), 531–541. https://doi.org/10.1016/0883-0355(87)90013-9

McLeskey, J., & Brownell, M. (2015). *High-leverage practices and teacher preparation in special education* (Document No. PR-1). University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center. https://ceedar.education.ufl.edu/wp-content/uploads/2016/05/High-Leverage-Practices-and-Teacher-Preparation-in-Special-Education.pdf

Mikeska, J. N., & Howell, H. (2020). Simulations as practice-based spaces to support elementary teachers in learning how to facilitate argumentation-focused science discussions. *Journal of Research in Science Teaching, 57*(9), 1356–1399. https://doi.org/10.1002/tea.21659

Mikeska, J. N., Kurzum, C., Steinberg, J. H., & Xu, J. (2018). *Assessing elementary teachers' content knowledge for teaching science for the ETS® Educator Series: Pilot results* (Research Report No. RR-18-20). ETS. https://doi.org/10.1002/ets2.12207

Moss, P. A. (2011). Analyzing the teaching of professional practice. *Teachers College Record, 113*(12), 2878–2896. https://doi.org/10.1177/016146811111301202

Phelps, G., & Bridgeman, B. (2022). From knowing to doing: Assessing the skills used to teach reading and writing. *Reading and Writing, 35*, 2023–2048. https://doi.org/10.1007/s11145-022-10278-2

Phelps, G., Bridgeman, B., Yan, F., Steinberg, J., Weren, B., & Zhou, J. (2020). *Preliminary evidence on measurement characteristics for the foundational assessment of competencies for teaching performance tasks* (Research Report No. RR-20-27). https://doi.org/10.1002/ets2.12310

Phelps, G., Gitomer, D. H., Iaconangelo, C. J., Etkina, E., Seeley, L., & Vokos, S. (2020). Developing assessments of content knowledge for teaching using evidence-centered design. *Educational Assessment*, *25*(2), 91–111. https://doi.org/10.1080/10627197.2020.1756256

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal, 105*(1), 31–48. https://doi.org/10.1086/428764

Phelps, G., & Sykes, G. (2020). The practice of licensure, the licensure of practice. *Phi Delta Kappan*, *101*(6), 19–23. https://doi.org/10.1177/0031721720909582

Responsive Classroom. (2014). *The responsive classroom approach: Good teaching changes the future*. https://www.responsiveclassroom.org/sites/default/files/pdf_files/RC_approach_White_paper.pdf

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal, 50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Schutz, K. M., Grossman, P., & Shaughnessy, M. (2018). Approximations of practice in teacher education. In P. Grossman (Ed.), *Teaching core practices in teacher education* (pp. 57–84). Harvard Education Press.

Self, E. A., & Stengel, B.S. (2020). *Toward anti-oppressive teaching: Designing and using simulated encounters*. Harvard Education Press.

Shaughnessy, M., & Boerst, T. A. (2018). Uncovering the skills that preservice teachers bring to teacher education: The practice of eliciting a student's thinking. *Journal of Teacher Education, 69*(1), 40–55. https://doi.org/10.1177/0022487117702574

Shaughnessy, M., Ghousseini, H., Kazemi, E., Franke, M., Kelley-Petersen, M., & Hartmann, E. S. (2019). An investigation of supporting teacher learning in the context of a common decomposition for leading mathematics discussions. *Teaching and Teacher Education, 80,* 167–179. https://doi.org/10.1016/j.tate.2019.01.008

Shulman, L. S. (1986a). Paradigms and research programs for the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–36). Macmillan.

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

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Smith, P. S., & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 88–104). Routledge. https://doi.org/10.4324/9781315735665-9

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. Michigan State University, College of Education, Teacher Education and Development International Study Center.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903. https://doi.org/10.1002/sce.21027

Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of Teacher Education, 63*(5), 376–382. https://doi.org/10.1177/0022487112445789

Appendix A

Task Screenshots

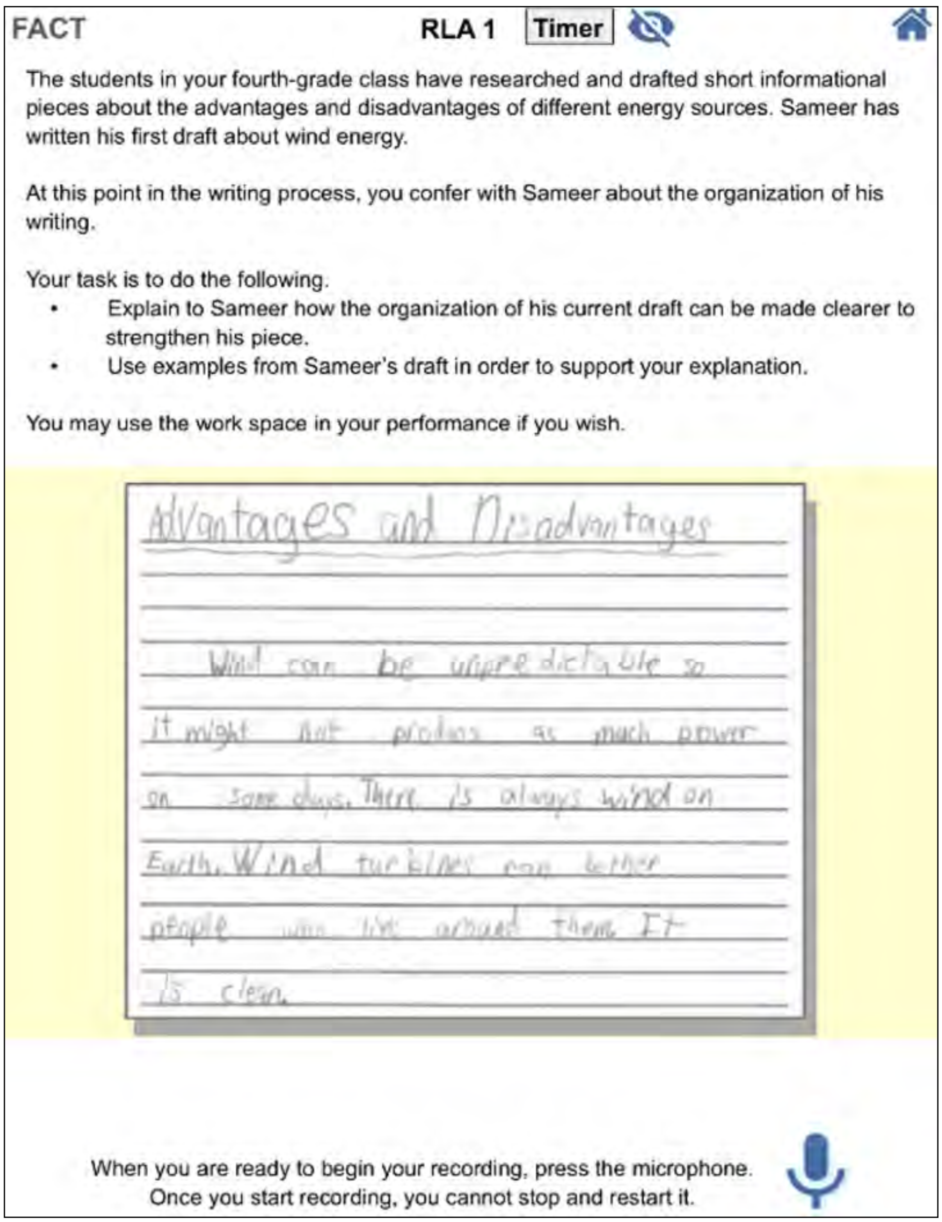


Figure A1 RLA 1: Persuasive writing and use of examples (explain process or concept using student work).

Figure A2 RLA 2: Grapho-phonemic segmenting (model or explain reading process or concept).

Video Script:

Alex: I think the only place in the poem where the author shows the idea that nothing good lasts forever is right here in the last line. *(Alex underlines the final line of the poem.)*

Jordan: Why do you think that?

Alex: Because it says "nothing gold can stay."

Jordan: Is that the only time he says that?

Alex: I think so. The other lines talk about leaves and flowers and other stuff, and the last line is the only one where the author says anything about how long things stay.

Jordan: OK. I couldn't find any other lines where the author talks about things staying around.

Figure A3 RLA 3: Poetry and evidence for theme (explain process or concept using student work).

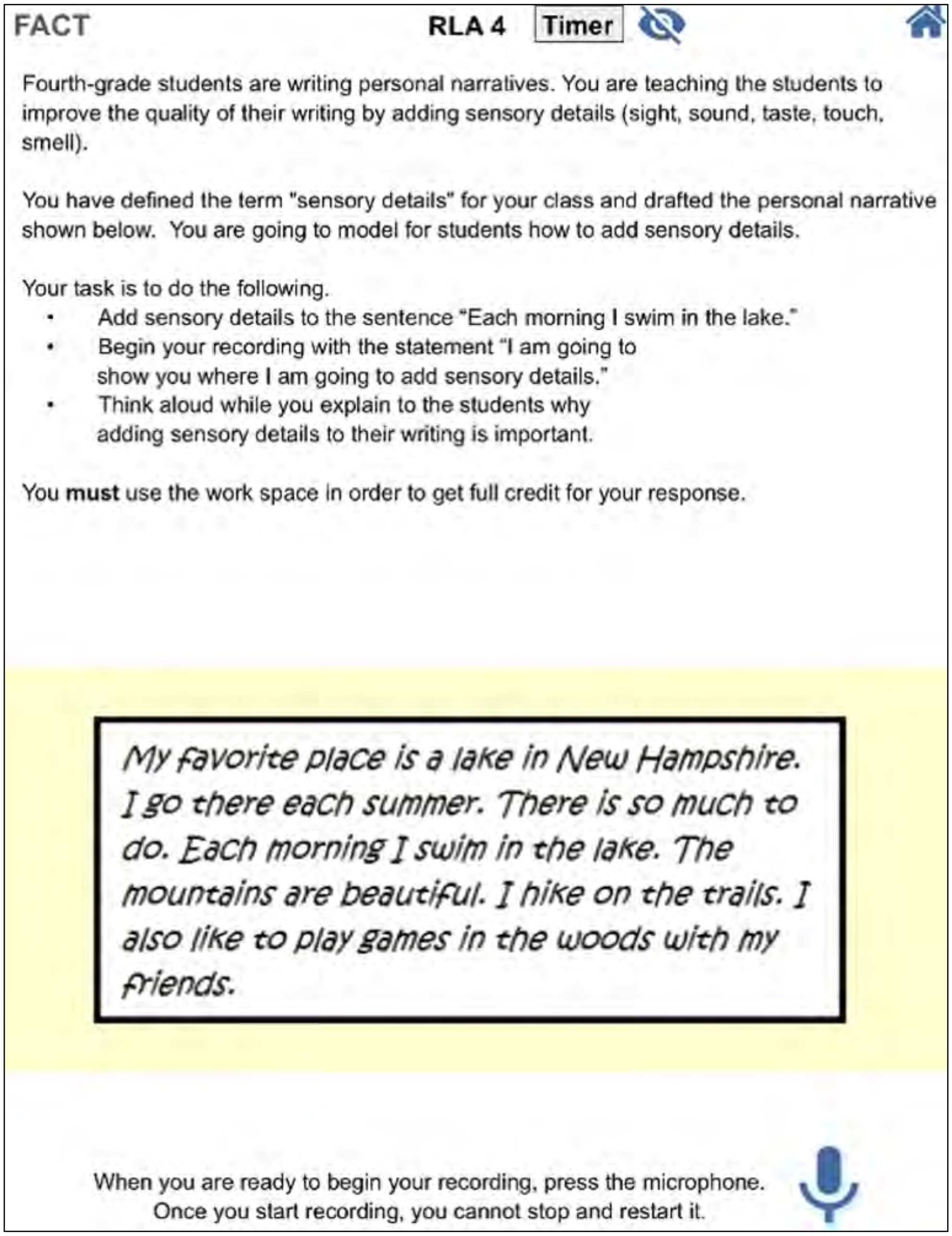


Figure A4 RLA 4: Narrative writing sensory details (model how to improve writing).

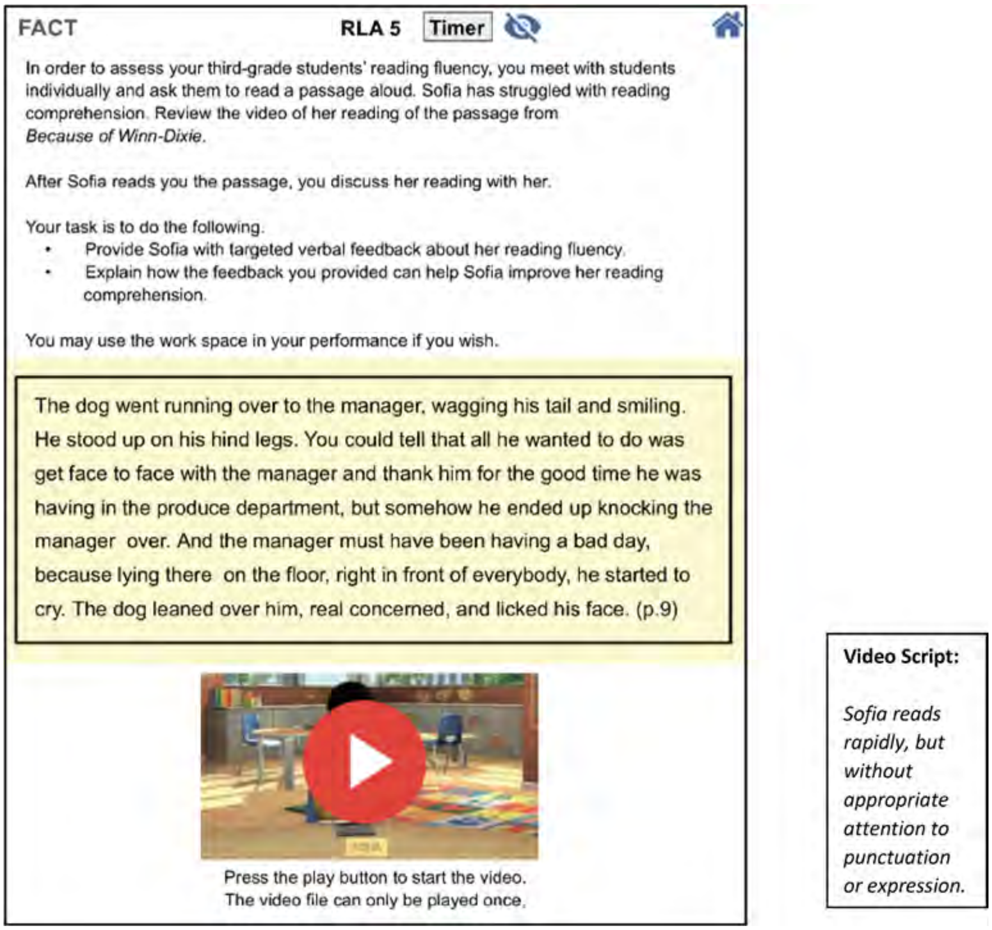


Figure A5 RLA 5: Reading fluency (explain process or concept using student work).

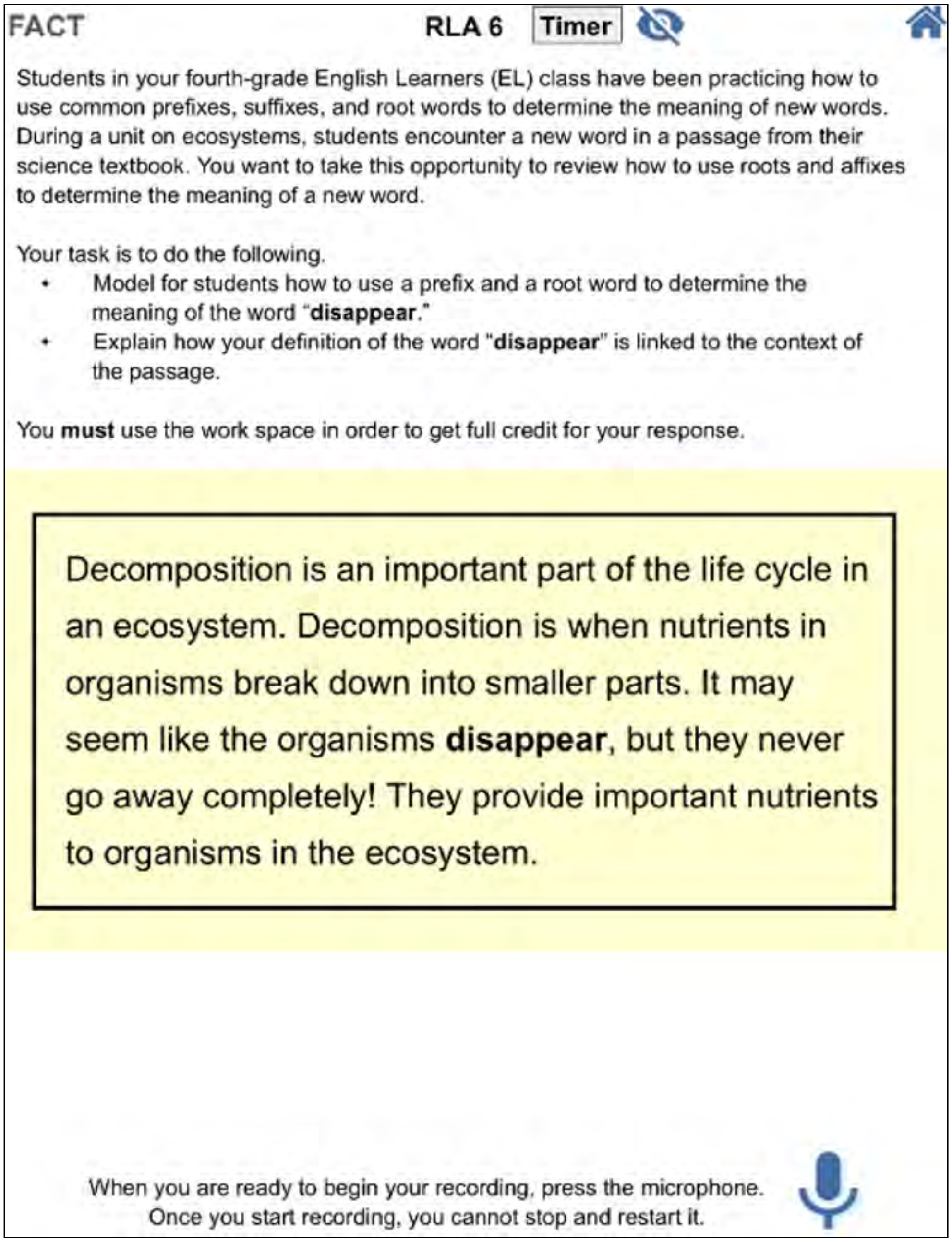
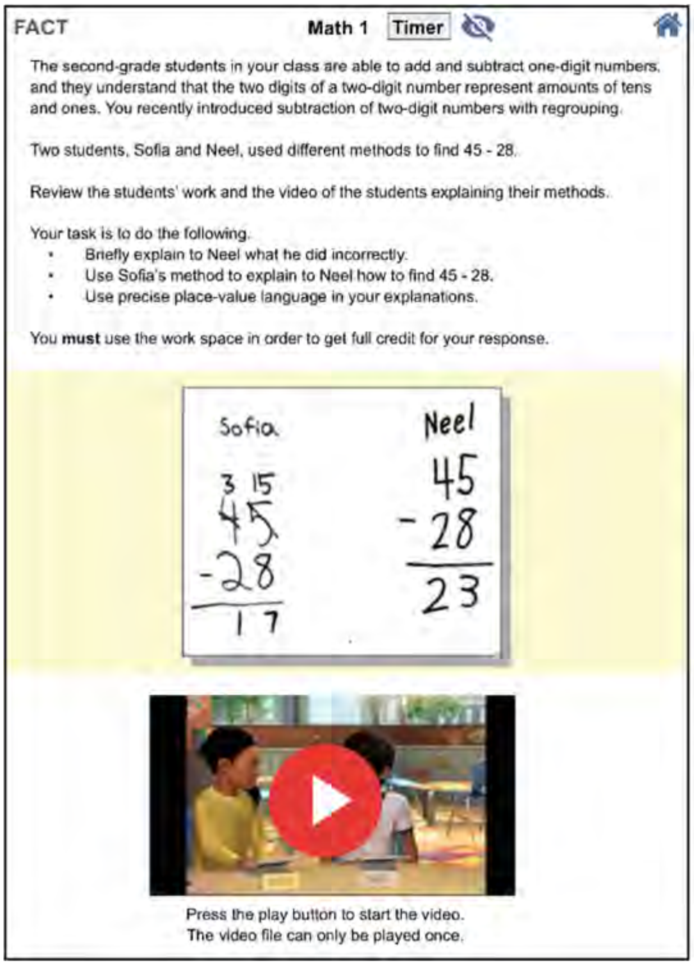


Figure A6 RLA 6: Word roots and affixes (model or explain reading process or concept).



Video Script:

Sofia: How did you subtract the numbers, Neel?

Neel: First I took 5 away from 8. I got 3.

Neel: Next I took 2 away from 4. I got 2, so the answer is 23. (short pause) How did you get 17?

Sofia: I took a ten from 40 and traded it for 10 ones, so I had 3 tens and 15 ones.

Sofia: Then I subtracted. 15 minus 8 is 7, so I wrote 7 in the ones place. In the tens place, I have 3 minus 2 is 1.

Sofia: So my answer is 17.

Figure A7 MATH 1: Two-digit subtraction with regrouping (explain student error and model correct method).

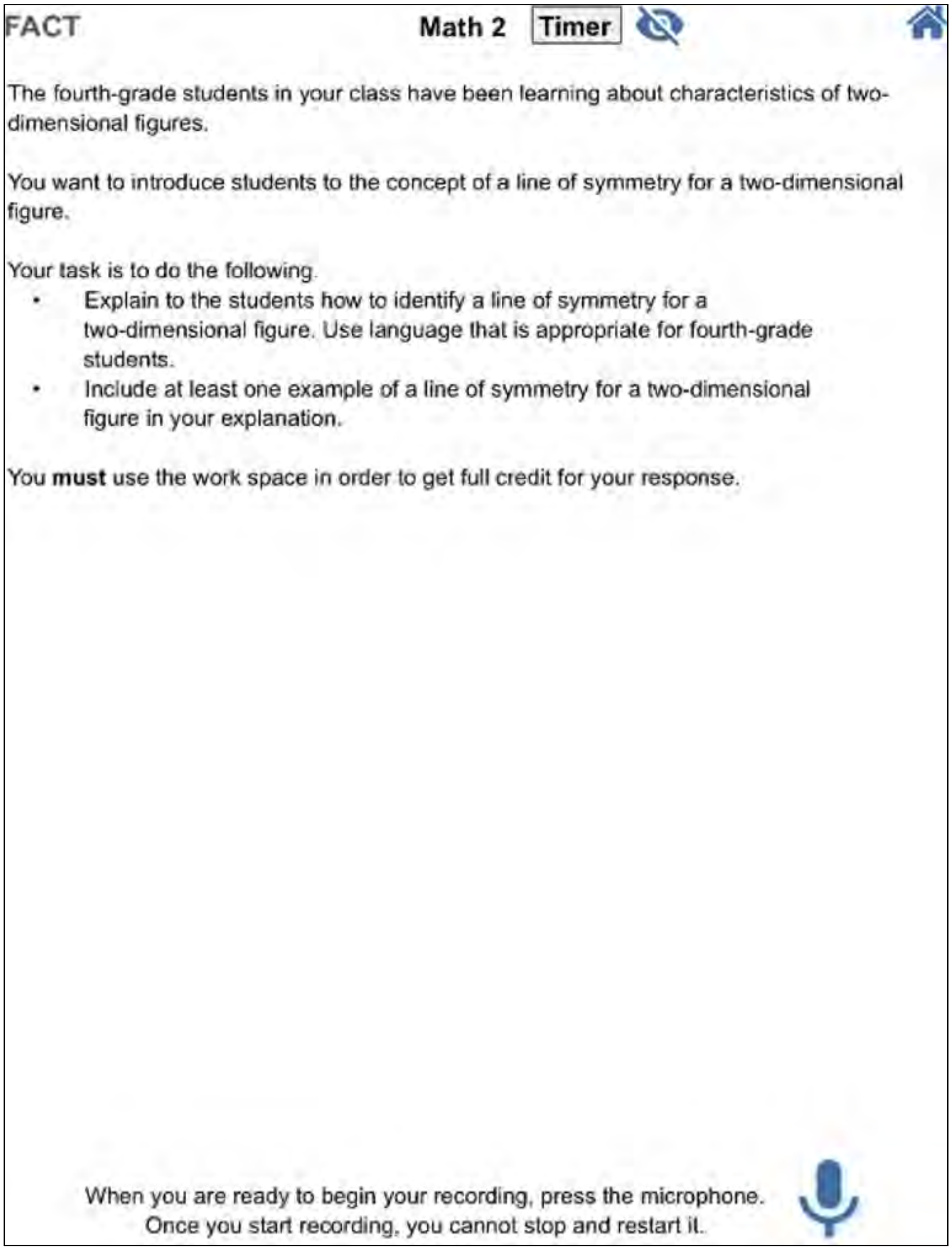
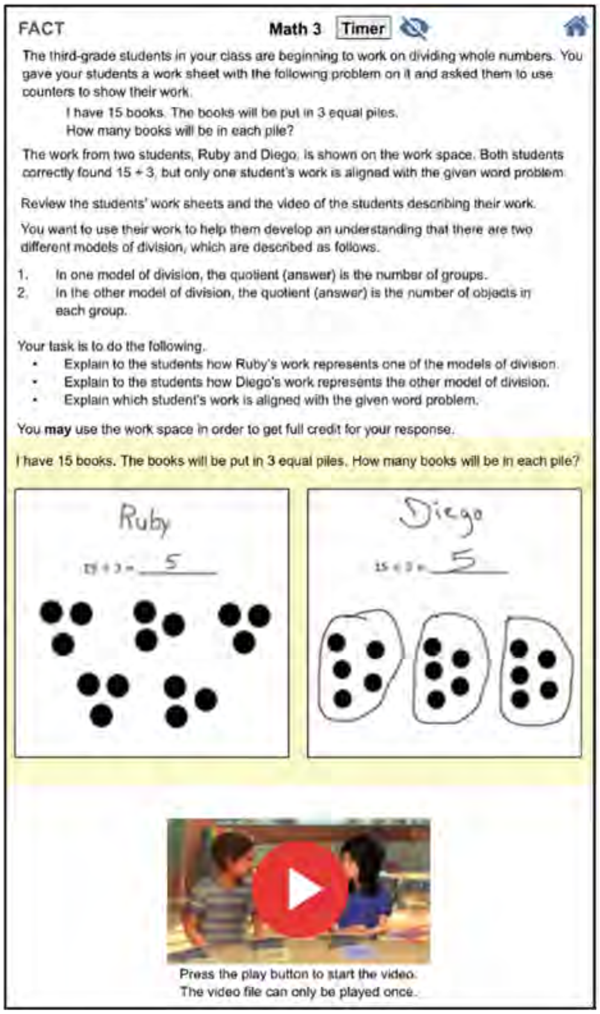


Figure A8 MATH 2: Line of symmetry (explain and model mathematical concepts).



Video Script:

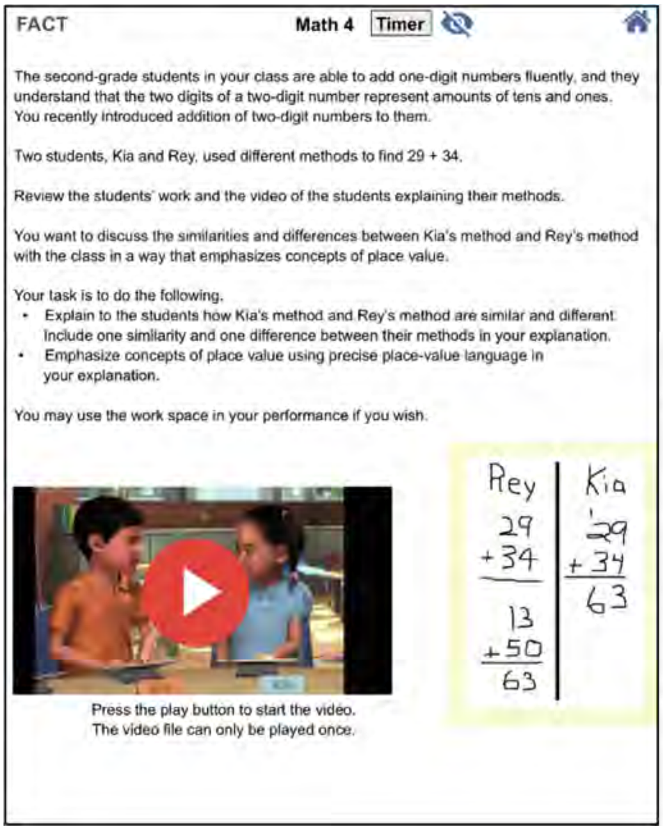
Diego: What did you do to answer the problem, Ruby?

Ruby: Well, I started with 15 counters for the books. I put 3 together, then another 3, then another 3, and another 3, and then one last 3. So I have 3 five times, so the answer is 5. Did you do the same thing?

Diego: Mine looks different. I started with 15 counters, too. But then I drew 3 circles for the piles. I put 1 counter in each circle. Then I put another counter in each circle. I kept going. When I ran out of counters, I had 5 counters in each circle. So there are 5 books in each pile.

Ruby: Hmm. We both got 5, but our pictures look really different.

Figure A9 MATH 3: Partitive and measurement models of division (represent concepts using a mathematical model).



FACT Math 4 Timer

The second-grade students in your class are able to add one-digit numbers fluently, and they understand that the two digits of a two-digit number represent amounts of tens and ones. You recently introduced addition of two-digit numbers to them.

Two students, Kia and Rey, used different methods to find 29 + 34.

Review the students' work and the video of the students explaining their methods.

You want to discuss the similarities and differences between Kia's method and Rey's method with the class in a way that emphasizes concepts of place value.

Your task is to do the following.

- Explain to the students how Kia's method and Rey's method are similar and different. Include one similarity and one difference between their methods in your explanation.
- Emphasize concepts of place value using precise place-value language in your explanation.

You may use the work space in your performance if you wish.

Press the play button to start the video.
The video file can only be played once.

Video Script:

Rey: How did you get your answer?

Kia: First I added the ones. 9 plus 4 is 13. I put down the 3 in the ones place and I wrote the 1 above the tens. Then in the tens, 1 plus 2 is 3, and 3 plus 3 is 6, so I put down 6 in the tens place. The answer is 63. What did you do, Rey?

Rey: 9 plus 4 is 13 and 20 plus 30 is 50. Then 13 more than 50 is 63, so the answer is 63.

Figure A10 MATH 4: Two-digit addition with regrouping (compare student methods).

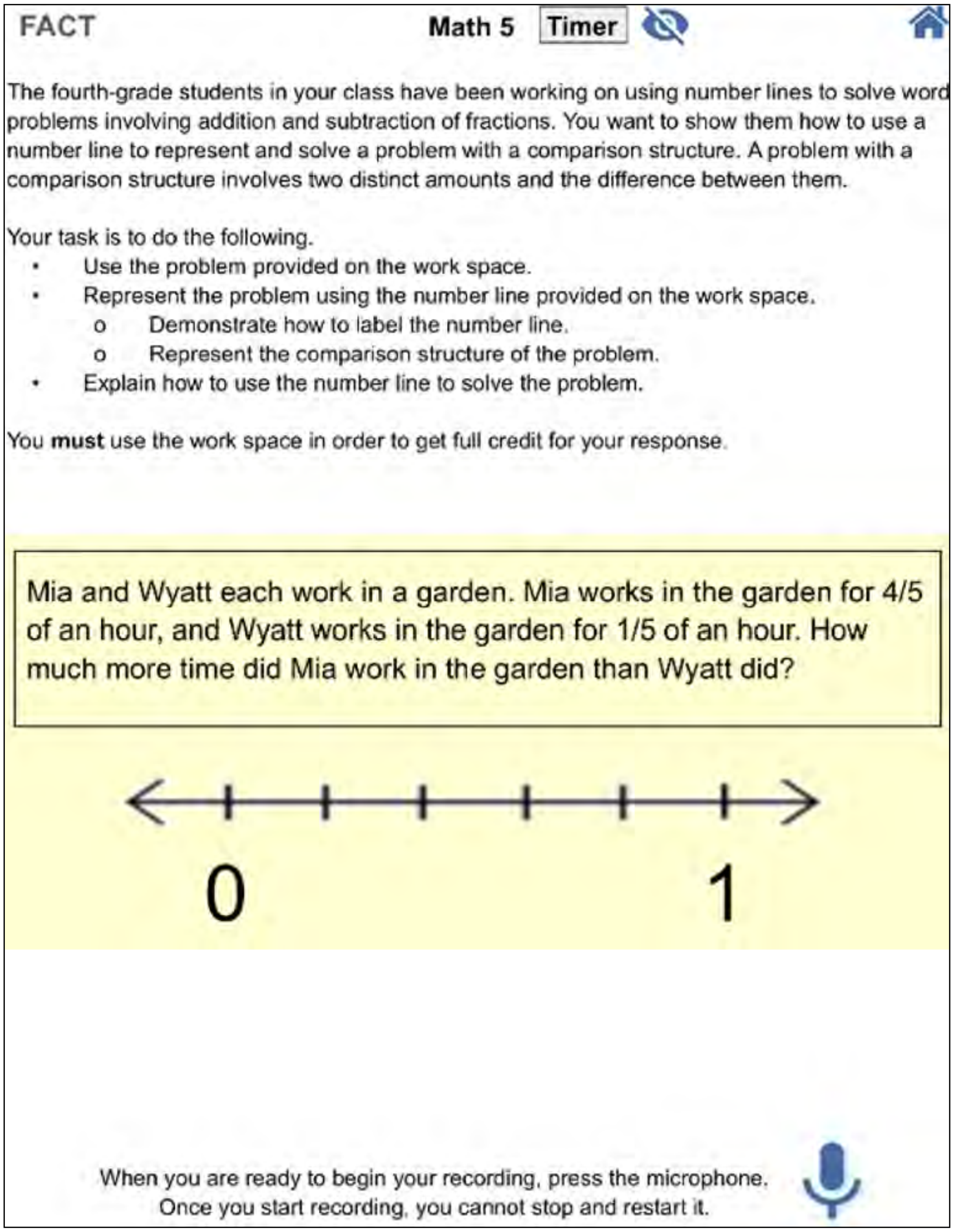


Figure A11 MATH 5: Fraction word problems and number line (explain problems using a representation).

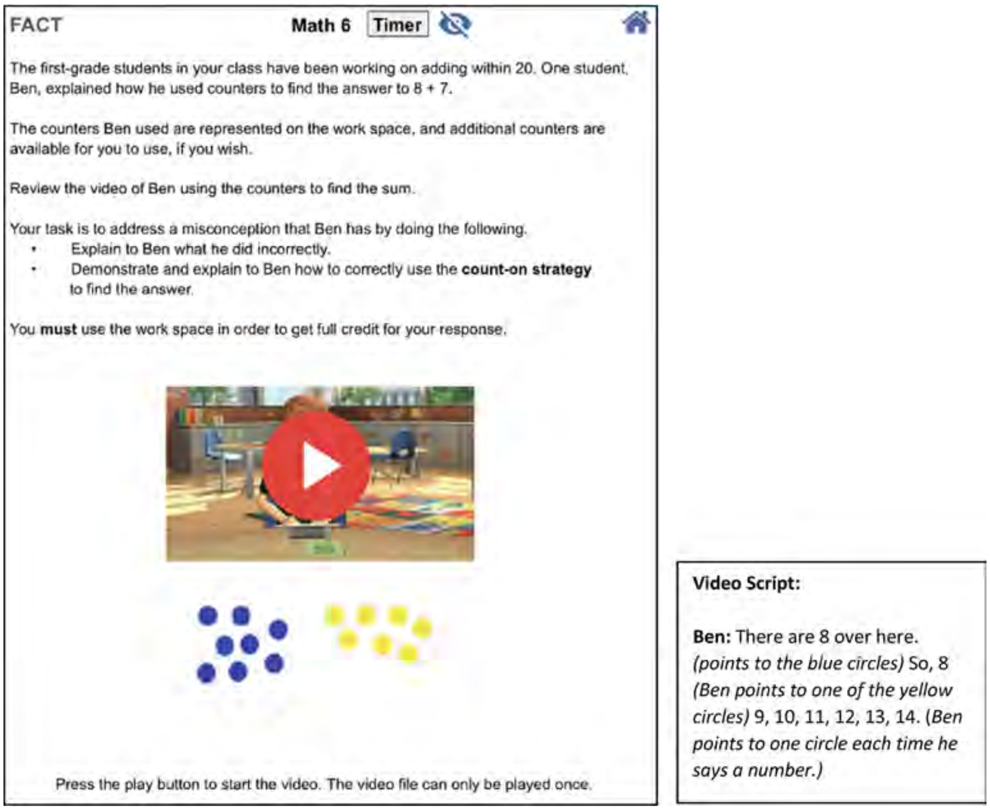


Figure A12 MATH 6: One-digit addition (explain student error and model correct method).

Appendix B

In-Depth Task Descriptions

What teaching competencies are required? Successfully completing this task requires drawing on knowledge about phonemic awareness, specifically how to segment a word into its individual phonemes. In order to do this, a teacher must be able to identify and isolate individual phonemes. This also includes modeling to illustrate the process of phonemic segmenting and guiding the students through the steps taken when encountering a new word. A strong performance involves modeling for students how to segment ALL of the words (i.e., back, chin, shop) both verbally and visually on the workspace. Inherent in this is the teacher's ability to define both "phoneme" and "digraph" as concepts. The teacher then must explain in a well-organized, concise, and appropriate manner for kindergarteners how hearing the sounds in words will help students as readers and writers.

Why does this task matter? The research supporting the science of reading names phonemic awareness as one of five fundamental areas on which to focus for effective early reading instruction. A teacher must be able to identify the relationship between letters and sounds so that the students can solve new words and become a more fluent reader. There are 44 phonemes in the English language and beginning readers need multiple opportunities to identity and manipulate phonemes and the associated letters as they learn to read . If students are able to identify and isolate all of the unique sounds, they can more easily decode more complex words where two letters make one sound, or a digraph. Modeling thought processes that go into decoding new words is a necessary teaching tool for explicitly instructing students how to approach new words. This type of direct and targeted modeling is appropriate for beginning or struggling English readers at any age therefore is important to incorporate into professional preparation

How can you evaluate a performance? The performance can be evaluated along multiple dimensions. Kindergarten students need repetition in the act of isolating phonemes, but they also need to do it in multiple contexts. Therefore, the content can be evaluated by how clearly and accurately each sound is isolated both verbally and visually, indicated with marks such as slashes, dashes, etc. The whole performance can be evaluated for the extent to which the language is grade-appropriate and the presentation is logically organized, clear, and concise. Finally, it is important to explain to students how practicing phonemic segmentation can help them learn to read words.

What challenges characterize preservice teachers' responses? Common difficulties that are observed for this performance task include the following.

- Segmenting the word ways that do not accurately represent phoneme segmentation (e.g., onset/rime, in half)
- Focusing discussion on the names of letters instead of the sounds associated with a letter or letter combination
- Imprecise pronunciation of phonemes (e.g.: ck = "kuh"; n = "nuh")
- Introducing additional example words that do not fit or confuse the concept
- Verbal narration not matching visual segmentation on the whiteboard
- Not framing the explanation of phoneme segmentation in the larger process of learning to read words

Figure B1 RLA 2: Phoneme segmentation (task description).

Performance Example #1. In this performance, the preservice teacher candidate:

- Attempts segmenting only the first word and focuses solely on the digraphs for the second two examples
- Is imprecise with language
  - "'S' and 'h' combine to make one sound, 'shuh.' So it would be 'sh-op.'"
  - "'B' and 'a' make the 'buh' sounds."
- Chooses words inaccurately that do not match what is being modeled in writing
  - "So now you have 'buh' and 'ack.'" (whiteboard: ba- ck)

Performance Example #2. In this performance the preservice teacher candidate:

- Begins by giving examples of digraphs, but one example (st) is a blend. "St" is a blend and therefore stands out as inaccurate among the other examples
- Directly identifies the digraphs in each example but incorrectly segments the word. By segmenting the word WITHIN the two letters that make up the digraph, the candidate is dividing one sound into two
- Focuses only on the digraph and does not segment the entire word

Performance Example #3. In this performance, the preservice teacher candidate:

- Sounds out the word accurately, writing each letter/digraph as she writes, instead of modeling how to divide a full word into phonemes
- Phoneme segmentation involves dividing a whole word into sound parts to be manipulated. Without showing the division of the whole word, it is not necessarily clear how the parts fit together
- Accurately writes out each phoneme from the words provided
- Adequately explains digraphs as well as the larger role the sounds play in reading

Performance Example #4. In this performance, the preservice teacher candidate:

- Links to previous concepts by beginning with an example of a simple 1-to-1 letter-sound correspondence (bat)
- "Taps out" the word "back" when writing it
- Visually models by underlining each phoneme in a different color with extra attention paid to "ck" in the first word
- Creates a list next to the segmented word of the number of sounds vs. letters
- Segments the second and third words accurately while accentuating each sound and pointing out the digraphs
- Links loosely to the larger process of reading and writing

Figure B2 RLA 2: Phoneme segmentation (performance examples).

What teaching competencies are required? Successfully completing this task requires drawing on knowledge about personal narrative and the role that sensory details play in drawing the reader into the writer's experience. This task also requires understanding what constitutes modeling revision of a piece of writing. This includes how to use modeling to illuminate both how and why writers make particular decisions about where to add sensory details, how to select appropriate sensory details, and how these details engage readers. A strong performance involves activating and coordinating this knowledge in a performance that is well-organized, concise, and appropriate for fourth-grade students.

Why does this task matter? Personal narratives are stories that relate a personal experience, convey something that is unique or important to the writer, and are typically told from the first person. They provide young writers with opportunities to draw on their own life experience, identify what is most important to them, and consider how to share this with their audience. Sensory details are important in personal narratives because they help the audience experience a story from the perspective of the author. Modeling the thought processes that go into selecting sensory details is a useful teaching tool for making these writing and editing decisions explicit to students. Because this type of targeted, instructional modeling is not something that adult writers do as they think through their own writing, it is important to incorporate it into professional preparation. Preservice teachers need opportunities to practice modeling that is appropriate for a particular writing sample, genre, and student level.

How can you evaluate a performance? The performance can be evaluated along multiple dimensions. The particular revisions can be evaluated for whether they draw on the sensory experience of swimming in the lake, are appropriate for fourth-grade students, and so on. The modeling can be evaluated for whether the writer's decisions are clearly and appropriately explained (i.e., making explicit the thinking behind how a sensory detail engages the audience and brings them into the writer's experience). The explanation for the importance of adding sensory details to personal narratives can also be evaluated for its relevance to improving personal narratives and its appropriateness for fourth-grade learners. Finally, the whole performance can be evaluated for the extent to which it is logically organized, clear, and concise.

What challenges characterize preservice teachers' responses? Common difficulties that are observed for this performance task include the following:

- Choosing descriptive details that are vague, don't invoke a strong sensory response, or are not sensory details. For example adding the descriptive detail "beautiful" so the sentence reads "I swim in the beautiful lake." Or adding the words "like to" so the sentence reads, "I like to swim in the lake."
- Listing or talking through a range of ideas without making a specific revision to the text
- Choosing without sufficient elaboration or explanation complex details that many fourth-grade students may not understand
- Adding details without modeling, or "thinking aloud," to explain the decisions that guide the revision
- Explaining the reason for adding details in ways that are vague, off the mark, incorrect, or not at an appropriate level for fourth-grade students

Figure B3 RLA 4: Adding sensory details to narrative writing (task description).

Performance example #1. In this performance, the preservice teacher candidate:

- Talks through the revisions without making any written edits to the text
- Describes an action related to touch (e.g., "I like to touch the water with my feet") instead of providing a detail that helps the reader to experience the sensation of touching the water
- Explains that adding the detail about liking to touch the water makes the sentence more "relatable and enjoyable," gives more "depth," and makes the narrative "richer," without any reference to bringing the audience into the writer's experience

Performance example #2. In this performance, the preservice teacher candidate:

- Talks through revisions without making any written edits to the text
- Moves quickly through each sentence in the passage either providing actual examples of details (e.g., adding "crystal blue" to "lake") or pointing to places in the passage to think about adding details (e.g., What makes the mountains beautiful? What color trees do they have?)
- Ends by saying that adding these details makes the passage more descriptive, without any reference to bringing the audience into the writer's experience

Performance example #3. In this performance, the preservice teacher candidate:

- Starts by noting that the sentence "Each morning I swim in the lake" is boring and she wants to add to the writing with sensory details related to sight, sound, taste, touch, and smell
- Talks through her thinking about adding details and writes each detail on the whiteboard indicating where it goes in the sentence
  - some of these are sensory details (e.g., <u>sunny</u> morning; <u>warm, glistening</u> lake)
  - some of these are not sensory details (e.g., <u>happily</u> swim)
- Ends by reading the initial sentence and then the revised sentence, noting that she thinks the revision is more interesting, without any reference to bringing the audience into the writer's experience

Performance example #4. In this performance, the preservice teacher candidate:

- Starts by noting that adding sensory details to "Each morning I swim in the lake" involves sight, sound, taste, touch, and smell
- Eliminates sound, taste, and smell and focuses just on details related to sight and touch
- Talks through her thinking as she adds details related to sight and touch (i.e., writes out "Each <u>bright</u> morning I swim in the <u>cold</u> lake")
- Ends by explaining that writers need to think about adding sensory details that describe what is going on in their story and make the reader feel as if they are actually right there in the story

Figure B4 RLA 4: Adding sensory details to narrative writing (performance examples).

What teaching competencies are required? Successfully completing this task requires explaining the process of using a number line to represent and solve a subtraction word problem. Specifically, this requires correctly and appropriately presenting a fraction comparison problem by showing two quantities on the number line and then using the number line to find the difference between the quantities. A strong performance involves activating and coordinating this knowledge in a performance that is well-organized, concise, and appropriate for fourth-grade students.

Why does this task matter? Proficiency with fractions is an important foundation for learning more advanced mathematics. However, many students struggle to reconcile their existing understanding of whole numbers with their nascent understanding of fractions. Number lines are an important tool for teaching and learning many fraction concepts, including part-to-whole relationships and comparing the magnitude of different fractions. Preservice teachers need opportunities to expand their understanding of different mathematical models and how to use these models in appropriate ways to support student learning. This involves learning how to use appropriate mathematical language for a given model and how to use a given model to solve problems in a way that helps students develop a strong conceptual foundation.

How can you evaluate a performance? The performance can be evaluated for whether it uses a number line to represent how to use a comparison strategy to subtract fractions and how likely it is that the explanation would help fourth-grade students understand how to use this strategy for other similar problems. The demonstration of how to correctly use the number line to solve the problem can be evaluated for how clearly and accurately the problem is represented and its appropriateness for fourth-grade students. Finally, the whole performance can be evaluated for the extent to which it is logically organized, clear, and concise.

What challenges characterize preservice teachers' responses? Common difficulties that are observed for this performance task include the following:

- Labeling the number line incorrectly (e.g., using whole numbers; representing fifths unequally) or neglecting to label the number line before using it to represent the problem
- Using imprecise fraction language (e.g., naming fractions as "one over five;" treating the fractions on the number line as whole numbers by counting "1, 2, 3…")
- Counting on the number line without using a pointer to indicate what is being counted.
- Incorrectly communicating the problem context (e.g., four hours and one hour rather than four-fifths of an hour and one-fifth of an hour; saying that Mia worked three-fifths more than Wyatt did)
- Using a traditional algorithm or another model (e.g., fraction bars; cubes) rather than using the number line to represent the problem
- Demonstrating a takeaway strategy instead of a comparison strategy to solve the problem

Figure B5 MATH 5: Representing fraction word problems on a number line (task description).

Performance example #1. In this performance, the preservice teacher candidate:

- Labels the number line to represent time spent working in the garden
- Explains how to solve the problem using a traditional algorithm rather than by demonstrating using a number line
- Incorrectly communicates the problem context (e.g., “Mia worked for 3/5 more time in the garden than Wyatt did”)

Performance example #2. In this performance, the preservice teacher candidate:

- Incorrectly represents the fractions on the number line (e.g., the time Mia spent working in the garden is placed between 3/5 and 4/5 instead of at 4/5)
- Counts the spaces between Mia and Wyatt on the number line but does not make this visible (because the pointer is not used)
- Uses imprecise fraction language (e.g., “Wyatt worked 1/5, so we’ll go 2, 3, 4 to get to Mia.”)
- Incorrectly communicates the problem context (e.g., “3/5 is how much more she worked than Wyatt did.”)

Performance example #3. In this performance, the preservice teacher candidate:

- Demonstrates how to label the number line using fifths
- Accurately marks the number line to represent the times at 4/5 of an hour and 1/5 of an hour
- Emphasizes counting the spaces between 1/5 and 4/5 and represents this with hops on the number line
- Uses precise fraction language (e.g., “If I count the space in between, I have 1/5 in between, 2/5 in between, 3/5 in between.”)

Performance example #4. In this performance, the preservice teacher candidate:

- Emphasizes that the number line will be labeled with fifths because there are five equal spaces, but counts the spaces without indicating what is being counted
- Correctly places labels for Mia’s and Wyatt’s times on the number line
- Uses imprecise fraction language (e.g., “five over five”; “If we see our labels for Mia and Wyatt, we can see that they’re spaced 1, 2, 3/5 apart.”)
- Explains how to find the answer by representing the distance between the times with hops on the number line
- Concludes by accurately explaining the answer in context

Figure B6 MATH 5: Representing fraction word problems on a number line (performance examples).

What teaching competencies are required? Successfully completing this task requires noticing that the student correctly counted the counters in the first group but made an error when counting on from the first counter in the second group. This task also requires familiarity with the count-on strategy in order to clearly explain how to use this strategy to find a sum. This task requires clearly and explicitly indicating the counters on the whiteboard workspace to show counting on. A strong performance involves activating and coordinating these competencies in a performance that is well-organized, concise, and appropriate for first-grade students.

Why does this task matter? The count-on strategy is an early introduction to mental math and helps young students build conceptual understanding about both numeracy and addition. For example, using the count-on strategy helps students learn that they do not need to recount from one every time when adding two numbers together. In addition, preservice teachers need to learn to pay close attention to what students say and do in order to provide an appropriate teaching response. While the mathematics in this task may seem straightforward, learning how to respond to students in ways that appropriately address specific errors or areas of confusion can be very challenging.

How can you evaluate a performance? The performance can be evaluated along multiple dimensions. The explanation of the student's error can be evaluated for whether it addresses the student's error and how likely it is that the explanation would help first-grade students understand the error. The demonstration of how to correctly use the count-on strategy can be evaluated for whether the strategy is used, how clearly the strategy is presented, and the demonstration's appropriateness for first-grade students. Finally, the whole performance can be evaluated for the extent to which it is logically organized, clear, and concise.

What challenges characterize preservice teachers' responses? Common difficulties that are observed for this performance task include the following:

- Neglecting to explain what error was made or providing an unclear explanation
- Counting without indicating which counters are being counted
- Communicating in a way that would likely be difficult for first-grade students to follow
- Demonstrating a count-all strategy rather than a count-on strategy

Figure B7 MATH 6: Addressing a student's misconception about how to use the count-on strategy to add one-digit numbers (task description).

Performance example #1. In this performance, the preservice teacher candidate:

- Rearranges the counters so that they are neatly organized on the whiteboard workspace
- Does not explain to the student what error was made
- Finds the correct sum using a traditional algorithm rather than by demonstrating the count-on strategy

Performance example #2. In this performance, the preservice teacher candidate:

- Says the student needs to remember to "count everything" (suggesting that the student missed one counter or lost track when counting) and does not further explain the student's error
- Demonstrates how to count on from the eighth counter to find the correct sum, rather than demonstrating how to count on from the group of 8 counters
- Says that "writing the number inside each counter is really important to ensure that each counter is counted"

Performance example #3. In this performance, the preservice teacher candidate:

- Briefly explains that "you already know there are 8 blue circles"
- Clearly demonstrates how to count on from the group of 8 counters to find the correct sum at a pace that is appropriate for first-grade students
- Provides an additional example in which the candidate more thoroughly addresses the error the student made and explains how to count on correctly
- Reviews how to use the count-on strategy to find the correct sum in the original problem

Performance example #4. In this performance, the preservice teacher candidate:

- Thoroughly explains to the student that the error was counting 8 two times
- Uses a number line as a different representation to show that 8 comes after 9
- Demonstrates how to count-on from 8 without clearly indicating on the whiteboard workspace that is she starting from the group of 8 counters

Figure B8 MATH 6: Addressing a student's misconception about how to use the count-on strategy to add one-digit numbers (performance examples).

Suggested citation:

Phelps, G., Kinsey, D., Florek, T., & Jones, N. (2023). *Using performance tasks to provide feedback and assess progress in teacher preparation* (Research Report No. RR-23-08). ETS. https://doi.org/10.1002/ets2.12371

Action Editor: Jamie Mikeska

Reviewers: Yi Song and Leslie Nabors Olah