Research Report

PRAXIS® Content Knowledge for Teaching: Initial Reliability and Validity Results for Elementary Reading Language Arts and Mathematics

ETS RR–20-15

Geoffrey Phelps
Jonathan Steinberg
Dawn Leusner
Jennifer Minsky
Karen Castellano
Laura McCulla

December 2020

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

PRAXIS® Content Knowledge for Teaching: Initial Reliability and Validity Results for Elementary Reading Language Arts and Mathematics

Geoffrey Phelps, Jonathan Steinberg, Dawn Leusner, Jennifer Minsky, Karen Castellano, & Laura McCulla

Educational Testing Service, Princeton, NJ

The primary purpose of this report is to provide preliminary evidence on the measurement properties for newly designed assessments of content knowledge for teaching (CKT) in elementary reading language arts (RLA) and mathematics. The goal is to offer the CKT tests through the *PRAXIS*® assessment. Additional analyses were conducted to provide initial evidence on the validity of the CKT assessments. One set of analyses investigated whether the test scores were sensitive to differences in participants' educational backgrounds that might be associated with opportunities to develop CKT. A second set of analyses involved examining score differences by the race/ethnicity of the participating candidates to provide evidence on whether the sample of participants in this study show group score differences that are comparable to what is typically observed on licensure exams. Finally, participant performance on the CKT tests was compared with performance on comparable PRAXIS assessments to examine potential differences in the difficulty of the items.

Keywords Teacher licensure assessment; content knowledge for teaching; pedagogical content knowledge; reading language arts; mathematics

doi:10.1002/ets2.12295

Teacher content knowledge has long been assessed as part of initial licensure. Herein, we briefly review how ideas about the types of content knowledge that teachers need to teach a school subject have evolved, how these ideas have led to a new generation of assessments designed to provide evidence of the types of content proficiency specialized to the work of teaching, and current arguments about how to apply evidence-centered design (ECD) to develop assessments of content knowledge for teaching (CKT).

From Content Knowledge to Content Knowledge for Teaching

Content knowledge has long been viewed as one of the primary competencies needed to teach a school subject (Ball & McDiarmid, 1990). Historically, in both licensure and other settings, content tests for teachers have focused on the content that students are expected to master as part of their K–12 education (Gitomer & Zisk, 2015). However, in the mid-1980s, new ideas about the content knowledge that mattered for teaching began to emerge. As part of the effort to lay a conceptual foundation for a new set of teacher exams that would eventually become the National Board of Professional Teaching Standards, Lee Shulman and his colleagues developed a framework that defined a seven-part knowledge base for teaching (Shulman, 1987). However, the knowledge domain that caught the attention of the education field, eventually leading to major changes in both the conceptualization and assessment of teacher content knowledge, was referred to as pedagogical content knowledge, or simply PCK. Shulman (1987) defined PCK as a distinct domain of content knowledge that represented "the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners and presented for instruction" (p. 8). One reason PCK caught the attention of the field was because it provided support for arguments that teaching should be considered a profession with an associated professional knowledge base.

Interest in PCK soon led to new efforts to both conceptualize and assess the full range of content competencies needed for teaching. These contemporary theories all start with the argument that strong content background, while necessary,

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

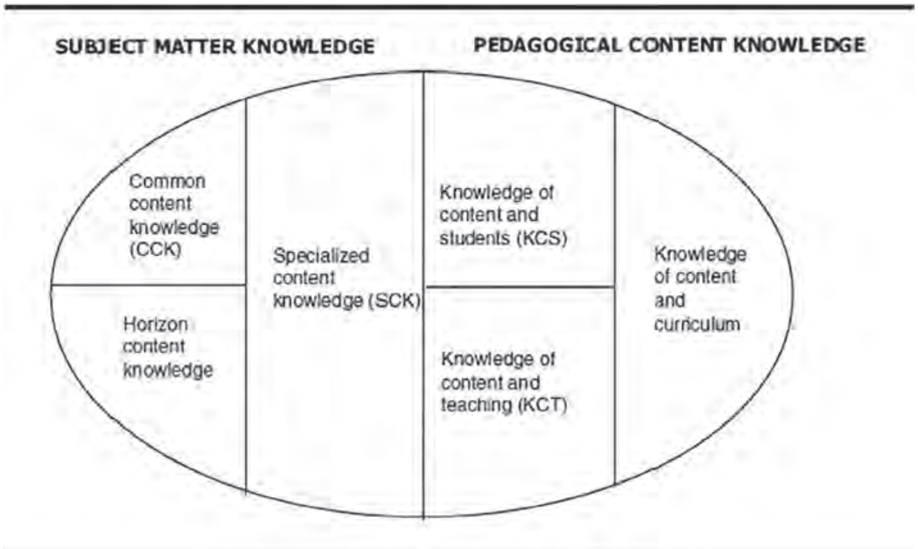

Figure 1 Content knowledge for teaching domain map. Adapted from Ball et al. (2008). Copyright 2008 by SAGE.

is insufficient for teaching a subject. Teaching requires additional types of content knowledge distinctive to the work of teaching (see, e.g., Gitomer et al., 2014; Hill et al., 2004; Kersting, 2008; Krauss et al., 2008; McCrory et al., 2012; Phelps & Schilling, 2004; Rowland et al., 2005; Sadler et al., 2013; Shulman, 1986).

Arguably, the most widely cited and influential framework to build on Shulman's original content categories was proposed by Ball et al. (2008). In this framework (Figure 1), the content knowledge used in teaching falls into two larger domains: content knowledge and pedagogical content knowledge. Building on the work of Schulman and his colleagues (see, e.g., Grossman, 1990; Marks, 1990; Shulman, 1986, 1987; Wilson et al., 1987; Wineburg, 1990), PCK is defined as a mix or amalgam of content knowledge and knowledge of students, teaching, or curriculum. *Knowledge of content and students* includes, for example, knowledge of the types of content problems students are most likely to find challenging or of common student misconceptions. *Knowledge of content and teaching* includes knowledge of how to represent a concept in ways that meet a particular instructional goal. *Knowledge of content and curriculum* includes knowing how to use curriculum materials to teach specific content concepts.

The CKT framework developed by Ball et al. (2008) also includes three categories of content knowledge. *Common content knowledge* is defined by its shared use in common across adult pursuits other than just teaching. For example, both engineers and mathematics teachers use knowledge of how to solve algebraic problems in conducting their day-to-day work. *Horizon content knowledge* is defined by knowledge of how different content ideas are connected across the content domain from the beginning concepts to ones that are more advanced. The category, however, that has received the most attention is *specialized content knowledge* (SCK).

SCK includes types of content knowledge that are only used in teaching. For example, while someone skilled in mathematics may use a particular mathematical strategy or method to solve a set of problems, teachers must understand more than how simply to do the mathematics themselves. They need to know, for example, a range of strategies that could be used to represent the mathematics in different ways to address different types of student learning needs, differences in the mathematics involved in each of these strategies, and whether such strategies are mathematically valid and generalizable. SCK, in this context, is a type of pure mathematics that is only needed for the work of teaching mathematics. It is different from PCK, because SCK does not involve knowledge of students or teaching or curriculum. However, like PCK, it is a form of professional content knowledge that is not likely to be understood by adults who have not had relevant professional learning opportunities.

Assessing Content Knowledge for Teaching

Interest in CKT has led to many projects that have focused on developing assessments of CKT. While some of these assessments have still focused on the content of the student curriculum, emphasizing the content found in new standards for student learning (e.g., Saderholm et al., 2010), the vast majority have focused on providing evidence of content knowledge that is unique to the work of teaching—a form of professional content knowledge distinct from the more general content competencies that have traditionally been assessed. For example, assessments of this type have been developed for elementary or middle school reading language arts (RLA; Carlisle et al., 2009; Phelps et al., 2014; Phelps & Schilling, 2004), mathematics (Hill et al., 2004; Kersting, 2008; Phelps et al., 2014; Tatto et al., 2008), and science (Mikeska et al., 2017; Mikeska et al., 2018; Sadler et al., 2013; Smith & Banilower, 2015). Other projects have focused on assessing secondary-level subjects, including English (Phelps et al., 2014), algebra and geometry (Herbst & Kosko, 2014; Krauss et al., 2008; McCrory et al., 2012; Mohr-Schroeder et al., 2017; Phelps et al., 2014), and physics (Iaconangelo et al., 2017; Phelps et al., 2020). While most of these projects have focused on a single subject—with assessment frameworks that only refer to the topical organization of that particular subject—a number of projects have set out to develop a common or shared framework that can be used across subjects (Gitomer et al., 2014; Phelps et al., 2020). Given their potential to support the development and use of coordinated programs of assessment, projects and the associated frameworks that are deliberately cross-subject are of particular interest in the licensure context, where there is a need for tests that can be interpreted in similar ways across all grades and subjects.

Many of these CKT assessments have been used in studies to evaluate validity arguments that focus on the nature, development, and role of teacher content knowledge. This new generation of CKT assessments has successfully been used for multiple research purposes, including comparing the knowledge of contrasting groups (including prospective teachers, practicing teachers, and nonteachers) with the goal of supporting the claim that CKT is a form of professional knowledge (Hill et al., 2007; Iaconangelo et al., 2017; Kleickmann et al., 2013; Krauss et al., 2008; Phelps, 2005, 2009; Phelps et al., 2019); studying and evaluating professional development and teacher learning to show that CKT assessments are sensitive to professional learning opportunities (Goldschmidt & Phelps, 2010; Hill & Ball, 2004; Phelps et al., 2016; Tröbst et al., 2018; van Driel et al., 1998); examining differences among types of content knowledge to show that CKT includes a complex of knowledge types (Copur-Gencturk et al., 2018; Hill et al., 2004; Mikeska et al., 2018; Phelps, 2009; Phelps & Schilling, 2004); and investigating how content knowledge contributes to both teaching quality and student learning outcomes to support the argument that CKT provides evidence directly associated with teacher quality (Baumert et al., 2010; Carlisle et al., 2009; Correnti & Phelps, 2010; Hill et al., 2005; Hill et al., 2008; Kersting et al., 2012; Phelps et al., 2012).

Using Evidence-Centered Design to Develop Assessments of Content Knowledge for Teaching

To provide stronger backing for assessments of CKT, projects are beginning to apply ECD principles to assessment development. For example, Phelps et al. (2020) described the use of an ECD approach (e.g., Mislevy et al., 2003) to develop a CKT assessment for secondary physics. ECD is a set of design principles that enable the specification of the knowledge, skills, and abilities that become the focus of the test development process. Assessment design, item development, scoring, and reporting are all guided by the specified knowledge, skills, and abilities that will be assessed to support explicit interpretations about score meaning.

In developing an ECD approach for assessing CKT, Phelps et al. (2020) began by distinguishing between *componential* and *integrated* approaches to assessing teacher knowledge. In a componential model,

> relevant aspects of knowledge are identified, and specific assessment tasks are designed to assess discrete knowledge components .... Adopting such an approach to assessing teaching breaks complex performances into discrete skills, and the associated knowledge can fall short in capturing the coordination among skills and the range of knowledge that is necessary to carry out an integrated task. (Phelps et al., 2020, p. 3)

> Phelps et al. (2020) argued instead for an integrated approach that starts by modeling how teachers use their knowledge (CKT) to engage in multiple and coordinated tasks such as asking questions, interpreting student responses, providing explanations and helping students develop their own explanations, asking follow-up questions, promoting discussions, etc .... Rather than starting with a taxonomy of discrete knowledge types, this approach starts with a taxonomy of knowledge as it is combined to execute the more complex tasks encountered in teaching. (p. 3)

Taking an integrated approach to assessing CKT is consistent with the work of Ball et al. (2008), who described the underlying theory and the task design logic as "practice-based" or "concerned with the tasks involved in teaching and the mathematical demands of these tasks" (p. 395). Content knowledge is defined by its use in teaching and, accordingly, frameworks for assessing CKT in mathematics are organized by the mathematical tasks that make up the moment-to-moment and day-to-day work of mathematics teaching. Tasks such as "presenting mathematical ideas, responding to students' 'why' questions, finding an example to make a specific mathematical point," and many more, require "a kind of unpacking of mathematics that is not needed—or even desirable—in settings other than teaching" (Ball et al., 2008, p. 400). Ball et al. were also careful to point out that "it is not always easy to discern where one of our categories divides from the next, and this affects the precision (or lack thereof) of our definitions" (p. 403). Rather than drawing hard distinctions among the types of knowledge that make up CKT in mathematics, the emphasis is instead placed on identifying the tasks that make up the mathematical work of teaching and then identifying the complex of mathematical knowledge that is characteristic of this work.

Phelps et al. (2020) formalized these arguments into an ECD approach to assessing CKT. They argued for a framework that has two major components—tasks of teaching and student learning targets—and argued that it is necessary to represent both in frameworks for assessing CKT:

> Tasks of teaching describe how teachers work with content in their moment-to-moment interactions with students. Tasks of teaching are the core features of content teaching that are frequent and recurrent across lessons, content areas and grade levels. They provide a common structure to describe CKT generally and are then elaborated in ways specific to subject matter teaching that supports students in developing the core concepts . . . . [Learning] targets then provide a starting point for modeling CKT by identifying the core concepts that students should come to understand through instruction in a domain. (Phelps et al., 2020, p. 6)

These ideas lead to an ECD approach that situates the assessment in the work of teaching a subject. Assessments tasks are designed to present test takers with tasks that are actually encountered in the work of teaching and to assess their ability to integrate and apply the relevant content knowledge that is called for. Instead of setting out to define components of CKT and develop items that are uniquely associated with these components, this theory starts with the work of teaching and sets out to develop items that assess whatever complex of knowledge is integrated when carrying out the particular work of teaching. This practice-based approach provides a means to directly link student learning, tasks of teaching, and teacher content knowledge. As illustrated by Phelps et al. (2020), CKT and the associated assessment tasks are defined at the intersection of learning targets and tasks of teaching.

The assessment items for the *PRAXIS*® CKT test discussed in this report build from and capitalize on the preceding research, development, and recent arguments for how to apply ECD principles. While this report focuses on results from the administration of the newly developed CKT items, a centrally important contribution of this work is to explore the potential of both designing licensure assessment around a practice-based theory of teacher competence and providing evidence of the knowledge that is needed to teach effectively. Following the ECD approach, CKT is modeled on both tasks of teaching and student learning targets (see Appendix A for tasks of teaching and content frameworks). To provide a concrete illustration of the tasks that were developed for this pilot and then incorporated into a new PRAXIS test of CKT, a series of items for both RLA and mathematics are presented in Appendix B. These items were all designed to focus on tasks of teaching and then to assess the combination of knowledge types (i.e., both common and specialized forms of content knowledge) that is activated and applied in practice as teachers carry out the work of teaching a subject (Phelps et al., 2020).

Navigating This Report

The primary analyses presented in this report investigate the measurement properties for the CKT subject matter sections and associated assessment items in RLA and mathematics. In addition to the primary analyses, three additional analyses were conducted to explore whether the assessment scores were sensitive to differences in participants' educational and professional backgrounds, to their race/ethnicity, and to scores on comparable PRAXIS tests. Each of these additional analyses drew from the pilot data used for the primary analyses. These additional preliminary analyses made use of these data in different ways, either combining scores across subject matter test sections, sampling participants from the larger data, or augmenting the pilot sample with additional participants or test scores from other administrations. To help orient readers to the different sections of the report, a brief summary of each section is provided. Additional details on the samples and methods are provided in each of the respective sections later in the report:

1. Examining measurement properties for the newly designed CKT items and subject matter sections. The main pilot study was designed to investigate the measurement properties for two parallel versions of the RLA assessment and two parallel versions of the mathematics assessment. At administration, each participant was randomly assigned one of the RLA sections and one of the mathematics sections. During Phase 1 of the administration, the RLA section was administered before the mathematics section, and during Phase 2 of the administration, the mathematics section was administered before the RLA section. The analyses of the assessment properties of the items and subject matter sections were conducted separately for each subject matter section.
2. Investigating performance differences for groups of test takers. This analysis explored whether differences in participants' degree-granting institutions, content preparation, professional preparation, and teaching experience were associated with score differences on the RLA and mathematics sections. To increase the sample size, the total point or raw count scores were converted to percent correct scores, and the resulting percent correct scores were combined for the two RLA sections and for the two mathematics sections.
3. Investigating score differences for Latinx and White candidates. This analysis drew from the combined percent correct scores to investigate score differences for Latinx and White candidates. Only those candidates who self-identified as Latinx or White with valid scores in both subject matter areas were included in the analysis.
4. Investigating score differences for African American and White candidates. The sample of African American candidates was initially insufficient for conducting even preliminary analysis as described for Latinx candidates. To increase the number of African American candidates, an additional data collection was conducted. RLA Assessment 2 (RLA2) and Mathematics Assessment 1 (Math1) were administered to 78 newly recruited African American candidates. These newly collected data were combined with data from African American candidates who had completed the same subject matter sections during the main pilot.
5. Comparing CKT scores and comparable PRAXIS content tests. To compare item difficulty for CKT and extant PRAXIS tests, participants were selected from the CKT pilot who had previously taken a comparable PRAXIS title in RLA (PRAXIS 5002) and/or a comparable PRAXIS title in mathematics (PRAXIS 5003). Using the procedures described, the percent correct values were calculated for the two versions of the RLA and mathematics CKT assessments. Then percent correct values were also calculated for the RLA PRAXIS test title and the mathematics PRAXIS test title. This calculation provided a basis for comparing the relative difficulties of CKT and PRAXIS test items and to examine to what extent the scores were correlated.

Pilot Design and Methods

The pilot study sample was recruited from educator preparation programs (EPPs) and was made up primarily of prospective teachers near the end of their teacher preparation programs. Each participant completed an RLA and mathematics assessment and a background information questionnaire (BIQ). In the following sections, the assessment forms, administration procedures, and study sample are described in more detail.

Instruments

Two parallel assessments were created for RLA, and two parallel assessments were created for mathematics. Each set of parallel assessments followed the same format and included selected-response items and constructed-response items, as shown in Table 1. Study participants completed one RLA assessment and one mathematics assessment.

The BIQ collected information on demographics (gender, native language, state, disability, etc.), education (undergraduate major/minors, highest educational attainment, technical training, etc.), professional preparation (EPP, certifications, endorsements, grade level, etc.), and teaching experience (student teaching, placements, employment status, location, etc.).

Participant Recruiting and Enrollment

Recruitment for the pilot involved multiple steps. First, EPPs were identified in target states. Next, students in their final year of their programs or within 1 year of graduation were invited to participate in the study by onsite coordinators at each

Table 1 Section Design by Item Type for Reading Language Arts and Mathematics

	Reading language arts ($n = 54$)		Mathematics ($n = 48$)		Total ($n = 102$)	
Task type	n	%	n	%	n	%
SR (1 pt)	48	88.9	40	83.3	88	86.3
SR (2 pt)	3	5.6	5	10.4	8	7.8
CR (2 pt)	2	3.7	3	6.3	5	4.9
CR (4 pt)	1	1.9	0	0.0	1	1.0

Note. CR = constructed response; SR = selected response.

EPP. Interested students contacted Educational Testing Service (ETS). ETS then sent eligible students a recruitment letter via e-mail with a link to a consent form and an eligibility survey to be completed online. Of the 711 recruitment e-mails sent to potential participants, 607 responded expressing interest and were sent an e-mail to register. Of these, 484 sent back W-9 forms formally enrolling to participate. After each test taker received a date and time to take the assessment, a unique link was sent to each test-taker to complete a self-administered, online BIQ. Participants were paid $250 for completing the assessment and BIQ survey.

Administration Procedures

The assessment was administered at designated ETS testing centers from October 2015 through December 2015. Each participant was randomly assigned to one of the RLA sections and to one of the mathematics sections. During Phase 1 of the administration, the RLA section was administered before the mathematics section, and during Phase 2 of the study, the mathematics section was administered before the RLA section. There were 160 administrations in the first phase and 220 administrations in the second phase. Of the 484 participants who were assigned test dates, 380 participants were tested. Table 2 provides basic demographic information on the administration sample organized by the institutions from which participants were recruited.

Constructed-Response Item Scoring

The assessment forms included six constructed-response items, three in RLA and three in mathematics. Study participants could receive scores of either 0–2 points or 0–4 points on the constructed-response items. All constructed-response items had associated scoring rubrics, specification of the evidence that would count toward a specific score, and example responses illustrating each level of the rubric. Scoring took place over a 2-day weekend and was conducted by raters with previous experience scoring either elementary teaching or content tests. Before attending the scoring session, each rater was assigned to either the RLA or mathematics scoring team and provided prework to review all items in the subject area, scoring rubrics, evidence inventories, and example responses. During the 2-day scoring session, scoring leaders worked in separate subject teams. The scoring teams worked on one item at a time, beginning by reviewing the item and scoring rules and scoring a number of practice items in small teams. Next, each item was scored independently by two raters. After initial scoring, all responses for which there was not agreement were identified, and the rating pair was asked to review and try to come to agreement. Where agreement could not be reached, scoring leaders provided the final ratings.

Participant Scores Not Reported

Responses for constructed-response and selected-response items were combined for initial scoring review. A small number of participants were not scored on either the RLA or mathematics sections for the following reasons. For RLA1, two participants were removed, one because the participant's responses to the constructed-response items were not captured and the other because the participant answered fewer than half of the selected-response items. For RLA2, one participant was removed for answering fewer than half of the selected-response items. For MATH1, one participant was removed because the constructed-response item responses were not captured. The final sample of valid scores for each subject matter section was as follows: RLA1 = 200, RLA2 = 177, MATH1 = 203, and MATH2 = 176. Overall, 377 participants had valid RLA scores, and 379 participants had valid mathematics scores.

Table 2 Institutional Characteristics for Assessed Pilot Participants

	Institutions ($n = 58$)		Participants ($n = 380$)	
Characteristic	*n*	%	*n*	%
Geographic region[a]				
Northeast	22	37.9	203	53.4
Midwest	9	15.5	36	9.5
South	21	36.2	132	34.7
West	3	5.2	5	1.3
Not available	3	5.2	4	1.1
Location				
City	26	44.8	126	33.2
Suburb/town	28	48.3	239	62.9
Rural	1	1.7	1	0.2
Not available	3	5.2	14	3.7
Carnegie classification				
Doctoral/research university	19	32.8	147	38.7
Master's college and university	30	51.7	211	55.5
Baccalaureate college	6	10.3	8	2.1
Not available	3	5.2	14	3.7
Institution size				
<5,000	12	20.7	18	4.7
5,000–9,999	13	22.4	56	14.7
10,000–19,999	11	19.0	109	28.7
≥20,000	19	32.8	183	48.2
Not available	3	5.2	14	3.7
Minority-serving institution[b]				
Yes	12	20.7	91	23.9
No	46	79.3	289	76.1
Institutional selectivity				
Most competitive	2	3.4	3	0.8
Highly competitive	5	8.6	82	21.6
Very competitive	14	24.1	69	18.2
Competitive	25	43.1	162	42.6
Less competitive	4	6.9	11	2.9
Noncompetitive	3	5.2	21	5.5
Special/not listed	5	8.6	32	8.4

Note. The table only includes categories with at least one participant.
[a]Geographic region was based on the institution's location in the United States. For institutions located in multiple states, the state of residence of the participant at the time of enrollment in the study was used. [b]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions.

Analysis Sample

The analyses in this section of the report examined each of the subject matter sections independently. The participant demographics for each section are presented in Table 3.

The sample was over 90% female and nearly two thirds White. These sample characteristics are relatively consistent with demographic characteristics for elementary licensure, at least in the context of PRAXIS (Steinberg et al., 2016).

Item and Test Analysis Results

Preliminary Item Analysis

RLA1 initially consisted of 54 CKT RLA items. Two items were designated not to be scored (DNS) due to problematic answer keys. Five others were removed based on results from the preliminary item analysis (PIA) for reasons such as proportionally more participants with higher ability (top 10%) selecting incorrect options (distractors) instead of the correct option. This left 47 RLA items for RLA1. RLA2 also initially consisted of 54 RLA items, with one item designated

Table 3 Participants' Gender and Race/Ethnicity

Race/ethnicity	RLA1 ($n = 200$) *n*	RLA1 %	RLA2 ($n = 177$) *n*	RLA2 %	MATH1 ($n = 203$) *n*	MATH1 %	MATH2 ($n = 176$) *n*	MATH2 %
Gender								
Male	13	6.5	13	7.3	16	7.9	10	5.7
Female	187	93.5	164	92.7	187	92.1	166	94.3
Race/ethnicity								
White	128	64.0	130	73.4	135	66.5	123	69.9
Asian	12	6.0	2	1.1	12	5.9	3	1.7
Latinx	41	20.5	29	16.4	39	19.2	32	18.3
African American	11	5.5	9	5.1	10	4.9	10	5.7
Two or more races	4	2.0	6	3.4	6	3.0	4	2.3
Other	2	1.0	1	.6	1	0.5	2	1.1

Note. RLA = reading language arts. No participants in the sample identified as Native American. The respective counts and proportions of those with missing race/ethnicity were RLA1 ($n = 2$) and MATH2 ($n = 2$).

Table 4 Number of Items Representing Each Core Idea

Core idea	RLA1 ($n = 47$) *n*	RLA1 %	RLA2 ($n = 48$) *n*	RLA2 %	MATH1 ($n = 43$) *n*	MATH1 %	MATH2 ($n = 42$) *n*	MATH2 %
Foundational literacy skills	17	36.2	17	35.4				
Language	4	8.5	4	8.3				
Constructing meaning	26	55.3	27	56.3				
Counting and operations with whole numbers					13	30.2	14	33.3
Place value and decimals					6	14.0	5	11.9
Fractions, operations with fractions, and ratios					13	30.2	12	28.6
Early equations and expressions, measurement, and geometry					11	25.6	11	26.2

Note. RLA = reading language arts.

as DNS for a problematic answer key and five others removed during PIA, leaving 48 CKT RLA items for RLA2. MATH1 initially consisted of 48 mathematics items with no items designated as DNS and five others removed during PIA, leaving 43 CKT mathematics items for MATH1. MATH2 initially consisted of 48 mathematics items, with one item designated as DNS due to incomplete item responses with a constructed-response item and five others removed during PIA, leaving 42 CKT mathematics items for MATH2. Each content area consisted of a number of content subareas known as *core ideas*. A breakdown of items by content area and core idea can be found in Table 4.

Order Effects and Latency

Each participant's response time (latency) for each item was measured in seconds. For each section taken by a participant, total test time was calculated as the sum of the participant's individual item response times on that section. Response times were then averaged across participants for that section within each testing phase and with the phases also combined (see Table 5). In Phase 1, the RLA items were presented first and the mathematics items second. This ordering was reversed for the two forms in Phase 2. For both RLA and mathematics, participants took longer on average to complete a subject matter section when it appeared first. The combined time for the two sections of the RLA form differed by 6 min, and for mathematics, it differed by 4 min. This suggests that, at least in respect to time taken, the sections may not be parallel.

Item Difficulty

Table 6 displays the frequency distribution of item difficulties across forms and subject areas. These results suggest that for both mathematics and RLA, the items are distributed across the spectrum of difficulty, with the majority of items clustered in the mid-difficulty range. The two sections of RLA and the two sections of mathematics have similar distributions.

Table 5 Testing Time for Each Phase and Combined

Section	Section position	Average total time (s)	*SD* total test time (s)	Average time (min)	*SD* total time (min)
RLA1					
Phase 1 ($n = 83$)	First	4,251	1,429	71	24
Phase 2 ($n = 117$)	Second	3,734	1,365	62	23
Combined ($n = 200$)		3,949	1,412	66	24
RLA2					
Phase 1 ($n = 76$)	First	3,945	1,318	66	22
Phase 2 ($n = 101$)	Second	3,326	1,221	55	20
Combined ($n = 177$)		3,592	1,296	60	22
MATH1					
Phase 1 ($n = 77$)	Second	3,840	1,233	64	21
Phase 2 ($n = 126$)	First	4,052	1,349	68	22
Combined ($n = 203$)		3,972	1,307	66	22
MATH2					
Phase 1 ($n = 83$)	Second	4,095	1,297	68	22
Phase 2 ($n = 93$)	First	4,284	1,299	71	22
Combined ($n = 176$)		4,195	1,298	70	22

Note. RLA = reading language arts.

Table 6 Distribution of Item Difficulty

	RLA1 ($n = 47$)		RLA2 ($n = 48$)		MATH1 ($n = 43$)		MATH2 ($n = 42$)	
Answering correctly (%)	*n*	%	*n*	%	*n*	%	*n*	%
<20	7	14.9	4	8.3	4	9.3	3	7.1
20–39.9	9	19.1	7	14.6	13	30.2	8	19.0
40–59.9	8	17.0	17	35.4	12	27.9	11	26.2
60–79.9	13	27.7	9	18.8	12	27.9	14	33.3
80–94.9	8	17.0	11	22.9	2	4.7	5	11.9
≥95	2	4.3	0	0.0	0	0.0	1	2.4

Note. RLA = reading language arts. For constructed-response items, the observed average item scores were divided by the total number of points available to obtain the difficulty.

Item Discrimination

Table 7 displays the point-biserial correlations for piloted items across forms and subject areas. The point-biserial correlations provide useful information on how strongly related the items are to the overall test score. In RLA, there are more items that have lower point-biserial correlations compared to mathematics. However, it is important to recognize that this is not an indicator of content validity. Perfectly valid items might have low point-biserial correlations because they are measuring important content that differs from the majority of items on the test.

Measurement Properties for Subject Matter Sections

Table 8 shows descriptive score statistics and the coefficient alpha reliability for each subject matter area. The mean proportion correct for RLA was comparable between RLA1 (54.0%) and RLA2 (55.3%). For mathematics, MATH2 appeared to be slightly easier (56.2%) compared to MATH1 (48.6%). For RLA, the reliability coefficients were above .70 for both forms. For mathematics, the reliability coefficients were above .80 for both forms.

Table 9 further breaks down performance by the three core ideas for RLA, and Table 10 does so for the four core ideas for mathematics. Because the number of items in each core idea is generally small, it is important not to make substantive claims about the reliability coefficients. This information is included here for reference.

Classification Accuracy and Consistency

Livingston and Lewis (1995) described methods for evaluating the accuracy and consistency of decisions regarding whether a participant's test score can be classified as passing or not relative to a given passing score. The rate of

Table 7 Distribution of Item Point-Biserial Correlations

Range	RLA1 ($n = 47$)		RLA2 ($n = 48$)		MATH1 ($n = 43$)		MATH2 ($n = 42$)	
	n	%	*n*	%	*n*	%	*n*	%
0.80–0.89	0	0.0	0	0.0	0	0.0	0	0.0
0.70–0.79	2	4.3	0	0.0	0	0.0	0	0.0
0.60–0.69	1	2.1	3	6.3	5	11.6	7	16.7
0.50–0.59	4	8.5	7	14.6	10	23.3	11	26.2
0.40–0.49	12	25.5	9	18.8	12	27.9	11	26.2
0.30–0.39	18	38.3	19	39.6	8	18.6	7	16.7
0.20–0.29	4	8.5	6	12.5	5	11.6	5	11.9
<0.20	6	12.8	4	8.3	3	7.0	1	2.4

Note. RLA = reading language arts.

Table 8 Score Characteristics for Subject Matter Test

Characteristic	RLA1 ($n = 200$)	RLA2 ($n = 177$)	MATH1 ($n = 203$)	MATH2 ($n = 176$)
Total available score points	50	52	50	48
Minimum score	12	12	7	11
Maximum score	40	44	44	43
Mean score	27.0	28.8	24.3	27.0
SD	5.3	6.6	7.2	7.3
Mean percent correct	54.0	55.3	48.6	56.2
Coefficient alpha	0.70	0.77	0.80	0.82

Note. RLA = reading language arts.

classification accuracy ranges from 0% to 100% and represents the extent to which decisions made on the basis of a section of a test would agree with those made from all possible sections of the test without participants having an opportunity to practice (i.e., an estimate of the participants' true scores). The rate of classification consistency also ranges from 0% to 100% and represents the extent to which decisions made based on one section of a test would agree with those made using alternative sections of the same test.

For tests with only a single cut score, the method for determining accuracy would be based on a 2×2 table in which observed scores and "true" scores or "all forms average" (Livingston & Lewis, 1995, p. 180) are classified into "passing" or "failing" status. Classification accuracy is defined as the sum of the percentage of cases in which both the observed and true classifications agree. Classification consistency is defined similarly with the observed classifications and estimates of the alternate section classifications compared. Livingston and Lewis (1995) described the method used to estimate the true and alternate section score distributions given only the observed score distribution, the cut score, and the effective test length.

As the results in the present study were based on a pilot administration of the four subject matter sections prior to standard setting, the analyses in this section were conducted using single cut scores of 50%, 60%, and 70% of the available raw score points. These scores were selected to encompass the likely possible cut scores that might be identified in standard setting or selected by states for use in making licensure decisions. Table 11 displays the results of these analyses shown for all participants completing RLA2 and MATH1.

Associations of Content Knowledge for Teaching Scores With Test Taker Background Characteristics

The analyses presented in this section provide initial validity evidence describing differences in test performance according to participant characteristics that could be associated with different levels of CKT. For example, participants with more preparation in RLA or mathematics could demonstrate higher levels of CKT than participants with less preparation. Participants with higher grade point averages (GPAs) or who have earned entry into more selective institutions of higher

Table 9 Characteristics for Reading Language Arts Core Idea Subscores

Core idea	RLA1 ($n = 200$)	RLA2 ($n = 177$)
Foundational literacy skills		
Total available score points	17	17
Minimum score	2	3
Maximum score	12	15
Mean score	7.8	8.0
SD	1.9	2.3
Mean percent correct	45.8	46.8
Coefficient alpha	0.33	0.43
Language		
Total available score points	4	4
Minimum score	0	0
Maximum score	4	4
Mean score	2.7	2.4
SD	.9	.9
Mean percent correct	67.5	60.9
Coefficient alpha	0.36	0.07
Constructing meaning		
Total available score points	29	31
Minimum score	7	4
Maximum score	26	29
Mean score	17.0	18.4
SD	3.9	4.9
Mean percent correct	57.0	59.2
Coefficient alpha	0.62	0.73

Note. RLA = reading language arts.

education could show comparatively higher performance on CKT than participants with lower GPAs or attending less-selective institutions.

To increase the sample size, scores for participants completing RLA1 and RLA2 were combined, and likewise, the scores for participants completing MATH1 and MATH2 were combined. To account for the slightly different number of items on the two versions of each subject matter form, all analyses in this section were conducted using percent correct values instead of raw scores. Differences between groups were then analyzed using either t-tests or ANOVAs. Where statistically significant differences were found in the omnibus ANOVA tests, post hoc comparisons employing the Bonferroni procedure are discussed. Participant counts across the different comparisons vary depending on the total number of participants responding to the relevant survey questions with valid RLA and mathematics scores.

Institutional Characteristics

Table 12 displays performance results with respect to characteristics of the institutions participants attended. Characteristics of institutions, unless otherwise noted, were drawn from the Integrated Postsecondary Education Data System (U.S. Department of Education, 2017). The variables examined were geographic region, degree of urbanicity, Carnegie classification (Indiana University Center for Postsecondary Research, n.d.), institutional size, minority-serving institution status, and the institutional selectivity ranking (Barron's, 2015).

There were significant mean differences in percent correct ($p < .05$) for both RLA and mathematics for Carnegie classification, institutional size, minority-serving institution status, and institutional selectivity. For institution location, there was a significant difference only for mathematics, and for geographic region, there was a significant difference only for RLA.

Post hoc analyses for geographic region for RLA showed that the only significant difference was for participants attending northeastern schools compared to participants attending southern schools ($p < .01$). Post hoc analyses for institutional locale for mathematics showed that participants from institutions in the suburbs/towns on average performed better on mathematics compared to those attending institutions in urban areas ($p < .02$).

For Carnegie classification, average performance was significantly higher for both RLA and mathematics among those attending doctoral/research universities ($p < .01$) compared to those attending master's universities. Post hoc analyses for

Table 10 Characteristics for Mathematics Core Idea Subscores

Core idea	MATH1 ($n = 203$)	MATH2 ($n = 176$)
Counting and operations with whole numbers		
Total available score points	15	18
Minimum score	1	4
Maximum score	13	18
Mean score	7.4	11.1
SD	2.6	3.0
Mean percent correct	49.3	61.6
Coefficient alpha	0.48	0.62
Place value and decimals		
Total available score points	7	6
Minimum score	0	1
Maximum score	7	6
Mean score	3.9	3.9
SD	1.4	1.2
Mean percent correct	56.1	65.2
Coefficient alpha	0.35	0.18
Fractions, operations with fractions, and ratios		
Total available score points	16	13
Minimum score	2	0
Maximum score	15	12
Mean score	7.4	6.6
SD	3.0	2.7
Mean percent correct	46.4	50.6
Coefficient alpha	0.57	0.60
Early equations and expressions, measurement, and geometry		
Total available score points	12	11
Minimum score	0	0
Maximum score	11	10
Mean score	5.5	5.4
SD	2.0	2.0
Mean percent correct	46.1	49.3
Coefficient alpha	0.50	0.47

Table 11 Reliability of Classification

Qualifying score at points in the score distribution[a] (%)	Reading language arts (RLA2)		Mathematics (MATH1)	
	Classified accurately (%)	Classified consistently (%)	Classified accurately (%)	Classified consistently (%)
50	86.5	81.9	85.8	80.2
60	84.6	80.0	88.5	84.0
70	91.7	88.2	94.6	92.2

[a]The qualifying score is shown as the percentage of the raw points possible for the form. For RLA2, the qualifying score for each point in the score distribution is 50% = 26, 60% = 32, and 70% = 37. For MATH1, the qualifying score for each point in the score distribution is 50% = 25, 60% = 30, and 70% = 35.

institutional size for both RLA and mathematics showed that only participants attending large institutions (20,000 undergraduates or more) scored significantly higher compared to those attending institutions with 10,000–19,999 undergraduates (RLA, $p = .01$; mathematics, $p < .01$). There were no other significant post hoc differences in average performance. For minority-serving institutions, participants performed about 5 percentage points lower in both subject areas than those not attending those types of institutions (RLA, $p < .01$; mathematics, $p < .01$). Post hoc analyses for institutional selectivity for both RLA and mathematics showed that, on average, participants attending highly/most competitive schools outperformed those attending less than competitive institutions (RLA, $p = .03$; mathematics, $p = .03$). Additionally, participants attending very competitive institutions, on average, outperformed those attending less than competitive institutions (RLA, $p < .01$; mathematics, $p < .01$). There were no other significant differences for institutional selectivity.

Table 12 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct Performance by Institutional Characteristics

	Reading language arts (RLA1 and RLA2)				Mathematics (MATH1 and MATH2)			
Characteristic	*n*	Mean	*SD*	*p*-Value	*n*	Mean	*SD*	*p*-Value
Geographic region								
Northeast	200	54.69	13.87	0.01	200	55.19	13.98	0.05
Midwest	36	54.52	11.71		36	52.53	11.42	
South	122	50.12	13.88		122	51.57	12.78	
Location								
City	122	51.60	14.20	0.14	122	51.29	12.37	0.02
Suburb/town	239	53.87	13.54		239	54.84	13.77	
Carnegie classification								
Doctoral/research	145	57.38	13.34	<0.01	145	57.83	13.04	<0.01
Master's	209	50.12	13.30		209	50.76	12.76	
Institutional size								
<5,000	18	48.02	16.99	<0.01	18	50.75	14.09	<0.01
5,000–9,999	56	52.40	13.82		56	53.06	14.19	
10,000–19,999	107	50.08	12.69		107	50.29	12.24	
≥20,000	181	55.60	13.62		181	56.07	13.31	
Minority-serving institution[a]								
Yes	88	49.07	15.23	<0.01	88	50.02	13.65	<0.01
No	274	54.39	13.04		274	54.79	13.11	
Institutional selectivity								
Highly/most competitive[b]	85	58.45	12.79	<0.01	85	60.26	12.41	<0.01
Very competitive	68	54.08	14.82		68	55.11	12.45	
Competitive	162	51.24	12.94		162	50.39	13.08	
Less than competitive	31	50.65	13.59		31	52.77	13.01	

[a]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions. [b]Owing to the extremely small sample size of participants from the most competitive institutions ($n = 3$), this group was combined with that representing highly competitive institutions ($n = 82$).

Table 13 reports CKT performance based on the type of teacher preparation program. There were two primary categories analyzed: undergraduate and master's degree. Fifth-year undergraduate and alternative route programs were not included because the sample sizes were too small. The t-tests showed significantly higher average performance in both content areas for those in master's programs compared to those in undergraduate programs ($p < .01$).

Educational Attainment

Table 14 displays CKT performance based on degree field. Differences in educational attainment were only significant for mathematics ($p < .01$) with participants with a major or minor in mathematics scoring about 10 percentage points higher (63.7%) compared to those not majoring or minoring in mathematics (53.0%). Note, however, that the sample sizes were very low for participants with a major or minor, especially for mathematics.

Table 15 presents results for CKT RLA and mathematics performance for participants with different levels of educational attainment. To account for low sample sizes in some of the reporting categories, sophomores and juniors were

Table 13 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct by Teacher Preparation Program Type

	Reading language arts (RLA1 and RLA2)				Mathematics (MATH1 and MATH2)			
Program type	*n*	Mean	*SD*	*p*-Value	*n*	Mean	*SD*	*p*-Value
Undergraduate	248	50.88	13.51	<0.01	248	51.07	12.39	<0.01
Master's	87	58.52	13.12		87	59.66	12.52	

Table 14 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct by English and by Mathematics Major or Minor

	Reading language arts (RLA1 and RLA2)					Mathematics (MATH1 and MATH2)			
English	*n*	Mean	*SD*	*p*-Value	Mathematics	*n*	Mean	*SD*	*p*-Value
No major or minor	311	53.33	13.89	0.44	No major or minor	349	53.04	13.46	<0.01
Major or minor	53	51.74	14.04		Major or minor	15	63.65	6.54	

Table 15 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct by Educational Attainment

	Reading language arts (RLA1 and RLA2)				Mathematics (MATH1 and MATH2)			
Educational attainment	*n*	Mean	*SD*	*p*-Value	*n*	Mean	*SD*	*p*-Value
Sophomore or junior	25	50.43	14.40	0.01	25	56.04	13.20	<0.01
Senior	175	51.31	13.91		175	51.19	12.52	
Bachelor's degree	62	53.16	11.38		62	52.00	12.97	
Bachelor's degree plus additional credits and above	102	56.79	14.58		102	57.68	14.25	

combined, and all options above those earning a bachelor's degree were also combined.[1] In both RLA and mathematics, the omnibus test showed significant differences in performance by level of educational attainment (RLA, $p = .01$; mathematics, $p < .01$). Post hoc analyses for RLA showed that the only significant difference was for participants in the highest educational attainment category (earning a bachelor's degree plus additional credits and above) who outperformed seniors ($p = .01$). Post hoc tests for mathematics showed that participants earning at least a bachelor's degree and additional credits outperformed both seniors ($p = .01$ and those earning a bachelor's degree ($p = .046$). There was no significant difference relative to performance for sophomores or juniors.

Table 16 shows average CKT performance based on undergraduate GPA. The omnibus ANOVA showed significant differences among GPA ranges and scores for both content areas ($p < .01$) in each case. Post hoc tests revealed that participants in the highest range (3.5–4.0) significantly outperformed participants with a 3.0–3.49 GPA (RLA, $p < .01$; mathematics, $p < .01$ and a 2.5–2.99 GPA (RLA, $p < .01$; mathematics, $p = .01$).

Teaching Experience

Table 17 displays performance differences on CKT RLA and mathematics by professional experience. For this analysis, each participant was categorized as a student teacher, a classroom teacher,[2] or neither. The omnibus ANOVA results indicated significant differences in average performance across groups for both content areas (RLA, $p = .01$; mathematics, $p = .02$). Post hoc comparisons for both subject areas showed that participants currently student teaching or currently

Table 16 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct by Undergraduate Grade Point Average

	Reading language arts (RLA1 and RLA2)				Mathematics (MATH1 and MATH2)			
Undergraduate GPA	*n*	Mean	*SD*	*p*-Value	*n*	Mean	*SD*	*p*-Value
3.5–4.0	213	56.29	12.66	<0.01	213	56.07	12.87	<0.01
3.0–3.49	123	49.43	14.45		123	50.23	13.02	
2.5–2.99	28	44.93	13.62		28	48.06	14.81	

Note. GPA = grade point average.

Table 17 Content Knowledge for Teaching Combined Reading Language Arts and Combined Mathematics Percent Correct by Current Teaching Status

	Reading language arts (RLA1 and RLA2)				Mathematics (MATH1 and MATH2)			
Professional experience	*n*	Mean	*SD*	*p*-Value	*n*	Mean	*SD*	*p*-Value
Neither student nor classroom teacher	61	48.65	14.28	0.01	61	49.41	11.96	0.02
Currently student teacher	200	54.59	13.48		200	54.98	13.31	
Currently classroom teacher[a]	103	52.85	14.04		103	52.98	13.97	

[a]Combines full-time, part-time, and substitute teaching.

acting as classroom teachers scored significantly higher than participants who were not student or classroom teachers (RLA, $p < .01$; mathematics, $p < .01$).

Score Differences for White, Latinx, and African American Candidates

Score differences for minority and White candidates are typically observed on licensure exams (see, e.g., Nettles et al., 2011; Steinberg et al., 2014). For example, on the PRAXIS Elementary Education: Content Knowledge (EE:CK) exam for the period November 2005 to November 2009, the standardized score difference was 0.85 for Latinx and White candidates and 1.27 for African American and White candidates (Nettles et al., 2011). Scores and pass rates significantly lower for minority candidates can adversely affect diversity in the teaching workforce. Therefore, with any new licensure assessment, it is important to evaluate score differences to determine whether the exam may have an adverse impact on pass rates for minority candidates and specifically whether these differences are due to construct-irrelevant factors that would raise concerns about test fairness.[3] However, score gaps or significant differences in pass rates that are the result of construct-relevant differences in performance can provide important guidance on whether there is a need for interventions that support minority candidates in passing licensure tests. Such interventions might focus on incentives to recruit promising individuals into education or access to additional opportunities to develop the competencies assessed on the new licensure tests.

Ideally, information about group differences by candidate race/ethnicity should be collected as part of an operational administration. Actual administration data allow for comparing scores on the actual populations that are being assessed and under the same stakes and conditions encountered when candidates are pursuing their teaching licenses. The results presented from this study should be interpreted with caution because the sample sizes are very small, the degree to which these samples are representative is unknown, and there are no stakes attached to the test results.

Content Knowledge for Teaching Comparison: Latinx and White Participants

Sample for Latinx and White Participants

The sample for the score comparisons included Latinx and White candidates with valid scores for both RLA and mathematics. The combined sample included 64 participants who self-identified as Latinx (Mexican, Mexican American, or Chicano; Puerto Rican; other Hispanic, Latino, or Latin American) and 254 participants who self-identified as White. Table 18 shows the institutional characteristics of the two pilot subgroups. Latinx participants attended 16 different institutions, whereas White participants attended 48 institutions with broader geographic representation. Whereas the majority of participating institutions in both race/ethnicity groups were located in suburbs or towns and with very large student populations (20,000 or more), Latinx participants tended to come more from urban-based and less selective institutions compared to White participants, who tended to come more from suburban or town-based and more highly selective institutions.

Other primary background variables of interest were professional/academic preparation (undergraduate GPA, degree attainment, type of EPP, degree field of study, and teaching status) with summaries provided in Table 19 for those with relevant data (Latinx = 61; White = 248). With respect to professional/academic preparation, the prevailing distribution of overall undergraduate GPA showed that a larger proportion of White participants (62.9%) reported values in the highest range (3.5–4.0) compared to Latinx participants (50.8%), and the proportion of White participants with a bachelor's

Table 18 Combined Institutional Characteristics of Reading Language Arts and Mathematics Pilot Participants Comparing Latinx and White Candidates

	Latinx institutions ($n = 16$)		Latinx participants ($n = 64$)		White institutions ($n = 48$)		White participants ($n = 254$)	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Geographic region								
Northeast	6	37.5	11	17.2	20	41.7	161	63.4
Midwest	2	12.5	2	3.1	8	16.7	30	11.8
South	8	50.0	51	79.7	16	33.3	57	22.4
West	0	0.0	0	0.0	2	4.2	3	1.2
Other	0	0.0	0	0.0	2	4.2	3	1.2
Location								
City	6	37.5	44	68.8	19	39.6	63	25.1
Suburb/town	10	62.5	20	31.3	26	54.2	187	74.5
Rural	0	0.0	0	0.0	1	2.1	1	0.4
Not available	0	0.0	0	0.0	2	4.2	3	1.2
Carnegie classification								
Doctoral/research	7	43.8	24	37.5	17	36.0	103	41.0
Master’s	9	56.3	40	62.5	23	47.9	140	55.8
Baccalaureate	0	0.0	0	0.0	6	12.5	8	3.2
Not available	0	0.0	0	0.0	2	4.2	3	1.2
Institution size								
<5,000	0	0.0	0	0.0	9	18.8	15	6.0
5,000–9,999	4	25.0	6	9.4	11	22.9	43	17.1
10,000–19,999	3	18.8	20	31.3	11	22.9	74	29.5
≥20,000	9	56.3	38	59.4	15	31.3	119	47.4
Not available	0	0.0	0	0.0	2	4.2	3	1.2
Minority-serving institution[a]								
Yes	7	43.8	46	71.9	7	14.6	30	11.8
No	9	56.3	18	28.1	41	85.4	224	88.2
Institutional selectivity								
Highly competitive	3	18.8	5	7.8	7	14.6	66	26.0
Very competitive	2	12.5	3	4.7	13	27.1	60	23.6
Competitive	7	43.8	21	32.8	20	41.7	115	45.3
Less competitive	3	18.8	21	32.8	2	4.2	8	3.1
Noncompetitive	1	6.3	14	21.9	2	4.2	5	2.0
Not available	0	0.0	0	0.0	4	8.3	8	3.2

Note. The table only includes categories with at least one participant.
[a]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions.

degree or higher (51.2%) was more than twice as large as the proportion of Latinx participants with the same education(19.7%). All but five Latinx participants reported being in an undergraduate teaching program, whereas more than one fourth of White participants were in master’s degree programs. Finally, with respect to current work status, the proportion of student teachers was slightly higher among Latinx participants (62.3%) compared to White participants (53.6%), yet there was a substantially higher proportion of White participants identified as classroom teachers (31.9%) compared to Latinx participants (11.5%).

Results for Latinx and White Participants

For RLA, Latinx participants had an average performance of 47.9% correct ($SD = 12.6\%$) compared to 54.5% ($SD = 13.9\%$) for White participants. For mathematics, Latinx participants had an average performance of 50.0% correct ($SD = 11.4\%$) compared to 54.8% ($SD = 13.7\%$) for White participants. To provide a more detailed illustration, the score differences are shown at different points in the respective score distributions (Figures 2 and 3). For RLA, the 25th percentile (P25) for Latinx candidate performance corresponds to about the 15th percentile for White performance. The 75th percentile (P75) for Latinx candidate performance corresponds to about the 58th percentile for White candidate performance. A similar

Table 19 Combined Demographic Characteristics of Reading Language Arts and Mathematics Pilot Participants Comparing Latinx and White Candidates

	Latinx ($n = 61$)		White ($n = 248$)	
Characteristic	*n*	%	*n*	%
Professional experience status				
None	16	26.2	36	14.5
Student teacher	38	62.3	133	53.6
Classroom teacher	7	11.5	79	31.9
Highest education level				
Sophomore	0	0.0	4	1.6
Junior	1	1.6	15	6.0
Senior	48	78.7	102	41.1
Bachelor's/bachelor's+	11	18.0	112	45.2
Master's/master's+	1	1.6	15	6.0
Undergraduate GPA				
3.5–4.00	31	50.8	156	62.9
3.0–3.49	23	37.7	78	31.5
2.5–2.99	7	11.5	14	5.6
Teacher preparation program				
Undergraduate teacher education program	56	91.8	157	63.3
Fifth-year postbaccalaureate program	0	0.0	7	2.8
Master's degree education program	4	6.6	70	28.2
Alternate route	0	0.0	7	2.8
Other	1	1.6	7	2.8

Note. Participant counts for background information vary depending on survey responses on the background questionnaire.

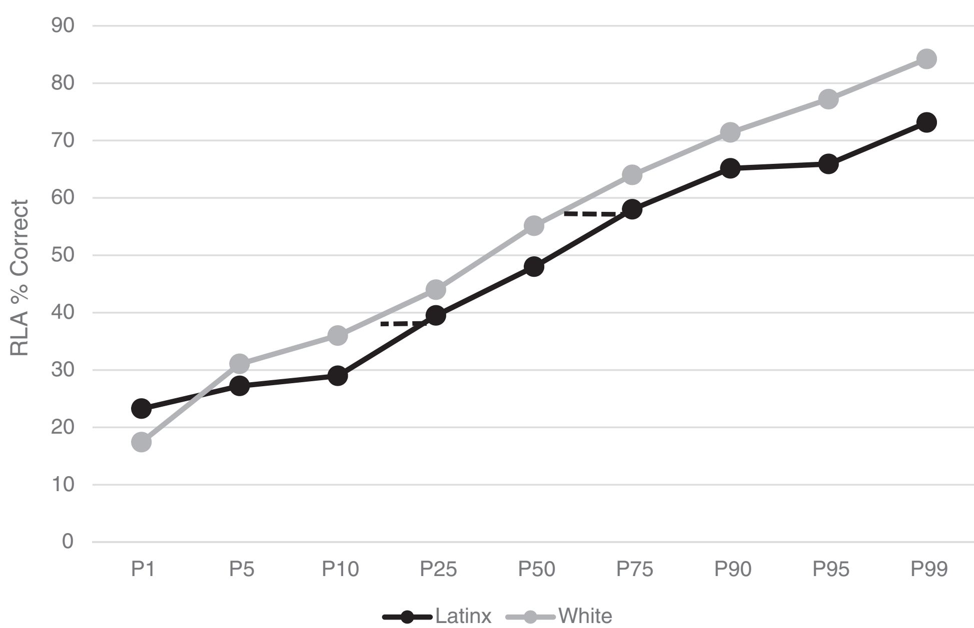

Figure 2 Content knowledge for reading language arts performance comparison at select percentiles for Latinx and White participants.

pattern was found for mathematics. The 25th percentile for Latinx candidate performance corresponds to about the 16th percentile for White candidate performance, and the 75th percentile for Latinx candidate performance corresponds to about the 61st percentile for White candidate performance.

It is also useful to examine differences in pass rates. Table 20 presents pass rates for Latinx and White candidates where participants receive 50%, 60%, and 70% of the total possible score points. These passing scores were selected to encompass

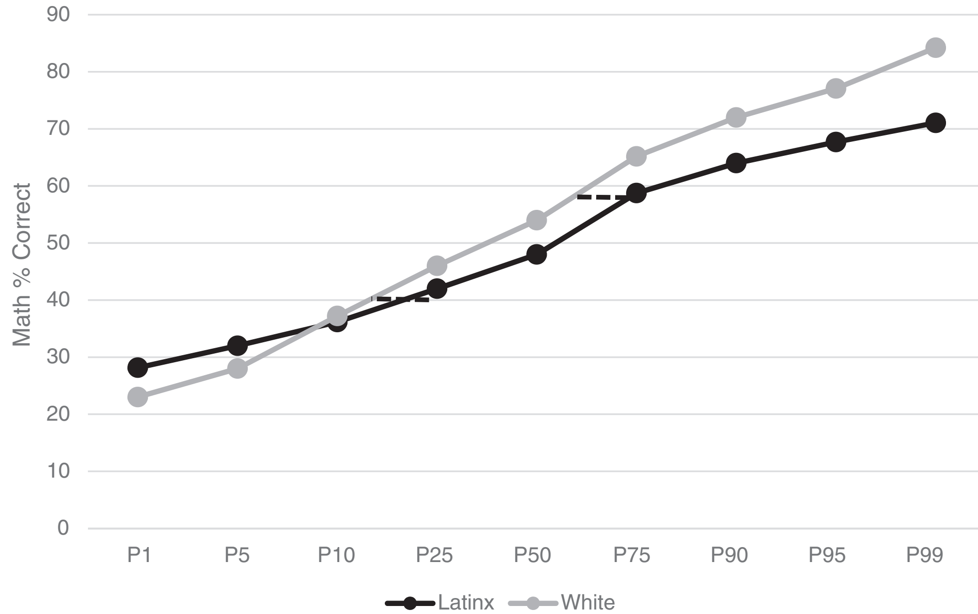

Figure 3 Content knowledge for mathematics performance comparison at select percentiles for Latinx and White participants.

Table 20 Pass Rates for Latinx and White Participants at Different Score Points

	Reading language arts (RLA2)		Mathematics (MATH1)	
Qualifying score at points in the score distribution[a] (%)	Latinx test takers passing (%)	White test takers passing (%)	Latinx test takers passing (%)	White test takers passing (%)
70	1.6	8.7	1.6	14.6
60	10.9	28.7	28.1	36.2
50	34.4	61.8	48.4	67.3

[a]The qualifying score is shown as the percentage of the raw points possible for the form. For RLA2, the qualifying score for each point in the score distribution is 50% = 26, 60% = 32, and 70% = 37. For MATH1, the qualifying score for each point in the score distribution is 50% = 25, 60% = 30, and 70% = 35.

the range of passing scores that might be determined in a standard setting or by states for making licensure decisions. Both CKT RLA and mathematics are difficult tests with only a small proportion of Latinx or White participants receiving 70% or more of the score points. The pass rate is higher for White participants at every passing score compared to Latinx candidates. The difference in pass rates is the largest for RLA at the 50% passing score, where only roughly one in three Latinx test takers pass compared to well over half of White test takers. The gap is the smallest at the 70% passing score, where a very small number of Latinx and White participants are passing.

Content Knowledge for Teaching Comparison: African American Versus White Participants

Sample for African American and White Participants

To increase the sample of African American candidates who participated in the initial pilot, an extension study was conducted from April 7 to May 31, 2016. The recruiting procedures, criteria for participation, and test administration for the extension study followed the same procedures as the pilot. Extension study participants were all administered the same version of RLA2 and MATH1 used in the original pilot. Data for African American candidates in the extension study (RLA2 = 78; MATH1 = 78) were combined with data from African American candidates from the original pilot (RLA2 = 9; MATH1 = 10) for a combined sample of African American candidates (RLA2 = 87; MATH1 = 88). Scores for this combined sample of African participants were then compared to those of White participants from the original pilot

(RLA2 = 130; MATH1 = 135). Because the participants for RLA and mathematics differ,[4] the samples for the two subject tests are described separately. Table 21 describes the institutional characteristics for the RLA sample. Table 22 describes the demographic characteristics for those in the RLA with valid data (African American participants = 87; White participants = 125). Table 23 describes the institutional characteristics for the mathematics sample. Table 24 describes the demographic characteristics for those in the mathematics sample with valid data (African American participants = 88; White participants = 133).

For both subjects, African American participants were more concentrated in the South and in urban settings compared to White participants who were concentrated in the Northeast and suburban settings. African American candidates tended to come from smaller institutions, minority-serving institutions, and institutions that were generally less competitive. African American candidates and White candidates reported different profiles for professional experience. While similar proportions of both groups reported no teaching experience, the participants with teaching experience differed in the type of teaching experience they reported. Well over half of African American candidates reported experience as a classroom teacher, whereas only about 10% reported experience student teaching. Conversely, White candidates reported

Table 21 Combined Institutional Characteristics of Reading Language Arts Pilot Participants Comparing African American and White Candidates

	African American institutions ($n = 41$)		African American participants ($n = 87$)		White institutions ($n = 38$)		White participants ($n = 130$)	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Geographic region								
Northeast	8	19.5	12	13.8	18	47.4	81	62.3
Midwest	1	2.4	1	1.1	6	15.8	13	10.0
South	30	73.2	48	55.2	12	31.6	32	24.6
Online	2	4.9	3	3.4	2	5.3	2	1.5
Not available	0	0.0	23	26.4	0	0.0	2	1.5
Location								
City	20	48.8	34	39.1	16	42.1	35	26.9
Suburb/town	17	41.5	25	28.7	20	52.6	91	70.0
Rural	2	4.9	2	2.3	0	0.0	0	0.0
Not available	2	4.9	26	29.9	2	5.3	4	3.1
Carnegie classification								
Doctoral/research	10	24.4	19	21.8	14	36.8	48	36.9
Master’s	24	49.0	30	34.5	19	50.0	75	57.7
Baccalaureate	5	10.2	12	13.8	3	7.9	3	2.3
Not available	2	4.9	26	29.9	2	5.3	4	3.1
Institution size								
<5,000	12	29.3	23	26.4	5	13.2	9	6.9
5,000–9,999	13	31.7	19	21.8	10	26.3	24	18.5
10,000–19,999	8	19.5	10	11.5	9	23.7	37	28.5
≥20,000	6	14.6	9	10.3	12	31.6	56	43.1
Not available	2	4.9	26	29.9	2	5.3	4	3.1
Minority-serving institution[a]								
Yes	21	51.2	38	43.7	6	15.8	17	13.1
No	20	48.8	49	56.3	32	84.2	113	86.9
Institutional selectivity								
Highly competitive	3	7.3	6	6.9	6	15.8	34	26.2
Very competitive	3	7.3	3	3.4	9	23.7	30	23.1
Competitive	16	39.0	29	33.3	18	47.4	59	45.4
Less competitive	16	39.0	22	25.3	3	7.9	3	2.3
Noncompetitive	1	2.4	4	4.6	0	0.0	2	1.5
Not available	2	4.9	23	26.4	2	5.3	2	1.5

Note. The table only includes categories with at least one participant.

[a]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions.

Table 22 Combined Demographic Characteristics of Reading Language Arts Pilot Participants Comparing African American and White Candidates

Characteristic	African American ($n = 87$)		White ($n = 125$)	
	n	%	*n*	%
Professional experience status				
None	21	24.1	24	19.2
Student teacher	11	12.6	55	44.0
Classroom teacher	55	63.2	46	36.8
Highest education level				
Sophomore	2	2.3	3	2.4
Junior	3	3.4	5	4.0
Senior	16	18.4	51	40.8
Bachelor's/bachelor's+	48	55.2	56	44.8
Master's/master's+	17	19.5	10	8.0
Doctorate	1	1.1	0	0.0
Undergraduate GPA				
3.5–4.00	38	43.7	77	61.6
3.0–3.49	32	36.8	41	32.8
2.5–2.99	16	18.4	7	5.6
2.0–2.49	1	1.1	0	0.0
Teacher preparation program				
Undergraduate teacher education program (BA or BS)	48	55.2	78	62.4
Fifth-year postbaccalaureate program (not leading to a master's degree)	3	3.4	5	4.0
Master's degree education program	21	24.1	33	26.4
Alternate route	9	10.3	7	5.6
Other	6	6.9	2	1.6

Note. Participant counts for background information vary depending on survey responses on the background questionnaire. GPA = grade point average.

Table 23 Combined Institutional Characteristics of Mathematics Pilot Participants Comparing African American and White Candidates

Characteristic	African American institutions ($n = 42$)		African American participants ($n = 88$)		White institutions ($n = 33$)		White participants ($n = 135$)	
	n	%	*n*	%	*n*	%	*n*	%
Geographic region								
Northeast	6	14.3	8	9.1	15	45.5	88	65.2
Midwest	2	4.8	3	3.4	7	21.2	18	13.3
South	31	73.8	49	55.7	9	27.3	26	19.3
West	0	0.0	0	0.0	1	3.0	2	1.5
Online	3	7.1	4	4.5	1	3.0	1	0.7
Not available	0	0.0	24	27.3	0	0.0	0	0.0
Location								
City	22	52.4	37	42.0	14	42.4	32	23.7
Suburb/town	15	35.7	21	23.9	17	51.5	101	74.8
Rural	2	4.8	2	2.3	1	3.0	1	0.7
Not available	3	7.1	28	31.8	1	3.0	1	0.7
Location								
City	22	52.4	37	42.0	14	42.4	32	23.7
Suburb/town	15	35.7	21	23.9	17	51.5	101	74.8
Rural	2	4.8	2	2.3	1	3.0	1	0.7
Not available	3	7.1	28	31.8	1	3.0	1	0.7

Table 23 Continued

Characteristic	African American institutions ($n = 42$)		African American participants ($n = 88$)		White institutions ($n = 33$)		White participants ($n = 135$)	
	n	%	*n*	%	*n*	%	*n*	%
Carnegie classification								
Doctoral/research	11	26.2	19	21.6	13	39.4	55	40.7
Master's	23	54.8	29	33.0	16	48.5	75	55.6
Baccalaureate	5	11.9	12	13.6	3	9.1	4	3.0
Not available	3	7.1	28	31.8	1	3.0	1	0.7
Institution size								
<5,000	13	31.0	24	27.3	4	12.1	6	4.4
5,000–9,999	12	28.6	18	20.5	7	21.2	21	15.6
10,000–19,999	8	19.0	11	12.5	9	27.3	46	34.1
≥20,000	6	14.3	7	8.0	12	36.4	61	45.2
Not available	3	7.1	28	31.8	1	3.0	1	0.7
Minority-serving institution[a]								
Yes	22	52.4	39	44.3	5	15.2	14	10.4
No	20	47.6	49	55.7	28	84.8	121	89.6
Institutional selectivity								
Highly competitive	3	7.1	4	4.5	5	15.2	34	25.2
Very competitive	3	7.1	3	3.4	9	27.3	32	23.7
Competitive	16	38.1	30	34.1	15	45.5	63	46.7
Less competitive	16	38.1	22	25.0	2	6.1	4	3.0
Noncompetitive	1	2.4	4	4.5	2	6.1	2	1.5
Not available	3	7.1	25	28.4	0	0.0	0	0.0

Note. The table only includes categories with at least one participant.
[a]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions.

Table 24 Combined Demographic Characteristics of Mathematics Pilot Participants Comparing African American and White Candidates

Characteristic	African American ($n = 88$)		White ($n = 133$)	
	n	%	*n*	%
Professional experience status				
None	21	23.9	19	14.3
Student teacher	12	13.6	69	51.9
Classroom teacher	55	62.5	45	33.8
Highest education level				
Sophomore	2	2.3	1	0.8
Junior	3	3.4	10	7.5
Senior	18	20.5	57	42.9
Bachelor's/bachelor's+	47	53.4	57	42.9
Master's/master's+	17	19.3	8	6.0
Doctorate	1	1.1	0	0.0
Undergraduate GPA				
3.5–4.00	36	40.9	80	60.2
3.0–3.49	34	38.6	48	36.1
2.5–2.99	17	19.3	5	3.8
2.0–2.49	1	1.1	0	0.0
Teacher preparation program				
Undergraduate teacher education program (BA or BS)	50	56.8	91	68.4
Fifth-year postbaccalaureate program (not leading to a master's degree)	3	3.4	2	1.5
Master's degree education program	20	22.7	33	24.8
Alternate route	9	10.2	4	3.0
Other	6	6.8	3	2.3

Note. Participant counts for background information vary depending on survey responses on the background questionnaire. GPA = grade point average.

Table 25 Pass Rates for African American and White Participants at Different Score Points

Qualifying score at points in the score distribution[a] (%)	Reading Language Arts (RLA2): African American test takers passing (%)	Reading Language Arts (RLA2): White test takers passing (%)	Mathematics (MATH1): African American test takers passing (%)	Mathematics (MATH1): White test takers passing (%)
70	3.4	11.5	3.4	12.6
60	16.1	30.0	10.2	34.1
50	49.4	63.1	28.4	63.7

[a]The qualifying score shown as the percentage of the raw points possible for the form is RLA2 50% = 26, 60% = 32, and 70% = 37; MATH1 50% = 25, 60% = 30, and 70% = 35.

Table 26 Institutional Characteristics of Pilot Participants With Both Content Knowledge for Teaching and PRAXIS Scores

	RLA institutions (n = 14)		RLA participants (n = 53)		Math institutions (n = 16)		Math participants (n = 63)	
Characteristics	n	%	n	%	n	%	n	%
Geographic region								
Northeast	10	71.4	50	94.3	12	75.0	60	95.2
South	2	14.3	2	3.8	2	12.5	2	3.2
West	1	7.1	1	1.9	1	6.3	1	1.6
Not available	1	7.1	0	0.0	1	6.3	0	0.0
Location								
City	3	21.4	14	26.4	4	25.0	16	25.4
Suburb/town	10	71.4	38	71.7	11	68.8	46	73.0
Not available	1	7.1	1	1.9	1	6.3	1	1.6
Carnegie classification								
Doctoral/research university	6	42.9	21	39.6	6	37.5	28	44.4
Master's college and university	7	50.0	31	58.5	9	56.3	34	54.0
Not available	1	7.1	1	1.9	1	6.3	1	1.6
Institution size								
<5,000	1	7.1	1	1.9	1	6.3	1	1.6
5,000–9,999	3	21.4	11	20.8	4	25.0	12	19.0
10,000–19,999	4	28.6	15	28.3	5	31.3	17	27.0
≥20,000	5	35.7	25	47.2	5	31.3	32	50.8
Not available	1	7.1	1	1.9	1	6.3	1	1.6
Minority-serving institution[a]								
Yes	2	14.3	2	3.8	2	12.5	2	3.2
No	12	85.7	51	96.2	14	87.5	61	96.8
Institutional selectivity								
Highly competitive	3	21.4	15	28.3	3	18.8	21	33.3
Very competitive	4	28.6	18	34.0	4	25.0	19	30.2
Competitive	6	42.9	19	35.8	8	50.0	22	34.9
Not available	1	7.1	1	19.9	1	6.3	1	1.6

Note. The table only includes categories with at least one participant. RLA = reading language arts.
[a]Minority-serving institutions include historically Black colleges and universities, Hispanic-serving institutions, tribal colleges and universities, and Asian American and Pacific Islander–serving institutions.

a much higher incidence of student teaching experience and a lower incidence of classroom experience. African American candidates reported higher levels of educational attainment but lower GPAs than their White counterparts.

Results for African American and White Participants

For RLA, African American participants had an average performance of 47.9% correct (*SD* = 13.5%) compared to 54.6% (*SD* = 15.0%) for White participants. For mathematics, African American participants had an average performance of 40.6% (*SD* = 14.4%) compared to 53.5% (*SD* = 12.6%) for White participants. For both tests, it is worth exploring key

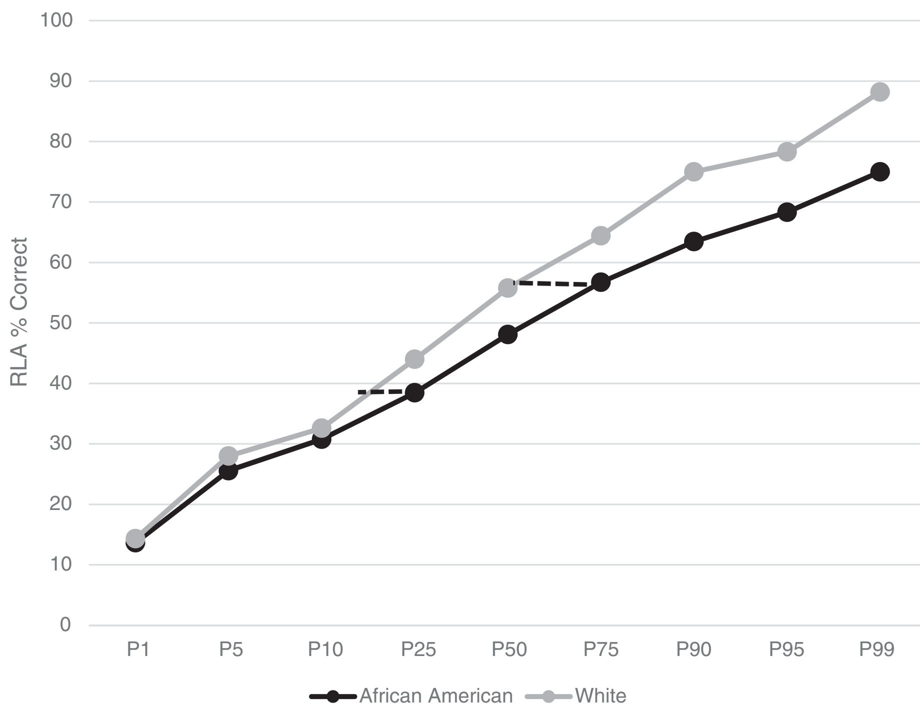

Figure 4 Reading Language Arts percent correct performance at select percentiles for African American and White participants.

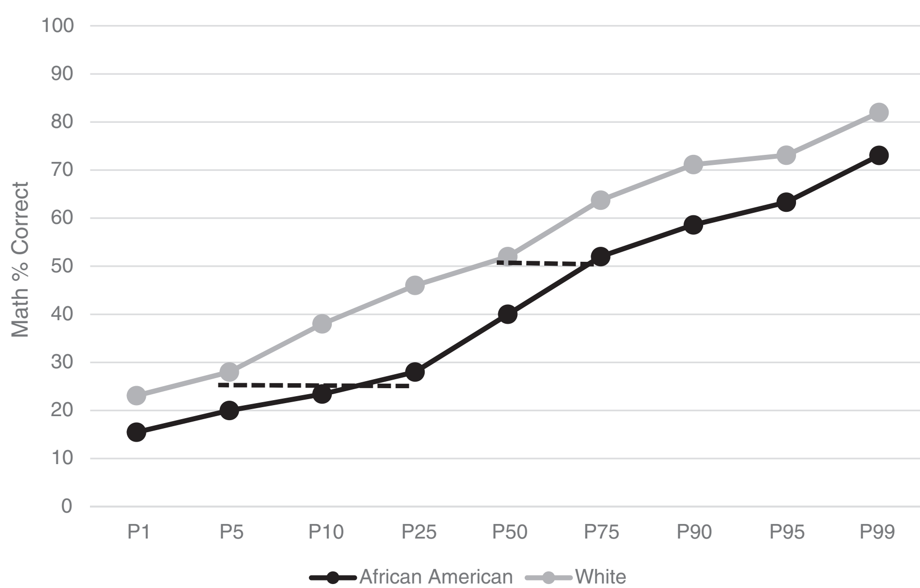

Figure 5 Mathematics percent correct performance at select percentiles for African American and White participants.

Table 27 Percent Correct for Pilot Participants With Both Content Knowledge for Teaching and PRAXIS Scores

Test	Mean	*SD*	Minimum	Maximum
Reading language arts ($n = 53$)				
PRAXIS 5002 RLA	77.9	8.6	47.7	95.3
CKT RLA	54.1	15.3	14.0	88.0
Mathematics ($n = 63$)				
PRAXIS 5003 Mathematics	80.7	12.0	35.0	100.0
CKT Mathematics	55.3	14.1	24.0	84.6

Note. CKT = content knowledge for teaching. RLA = reading language arts.

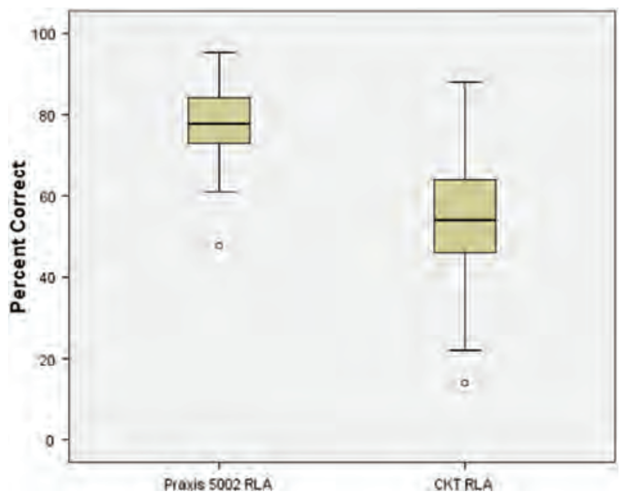

Figure 6 Percent correct for PRAXIS 5002 Reading Language Arts and Content Knowledge for Teaching Reading Language Arts.

percentiles in describing performance. Figure 4 shows more detail at key percentiles in the distribution. For RLA, the 25th percentile for African American candidate performance corresponds to about the 15th percentile for White candidate performance. The 75th percentile for African American candidate performance corresponds to just above the 50th percentile for White candidate performance. A similar pattern was found for mathematics, as shown in Figure 5.

Table 25 presents pass rates for African American and White candidates according to three hypothetical passing scores set at a score point where participants receive 50%, 60%, and 70% of the total possible score points. The pass rates for White candidates are substantially higher at all passing scores. The difference is the largest in mathematics at the 50% passing score where just below one in three African American candidates would pass compared to almost two in three White candidates.

Comparing PRAXIS and Content Knowledge for Teaching Assessments

In this section, we compare performance for pilot participants who also completed the PRAXIS Elementary Education Multiple Subjects (EE:MS) exams in RLA (5002) and mathematics (5003). Background characteristics for this sample are presented in Table 26. The results indicate strong similarity between the RLA and mathematics samples. Both samples are predominantly from the Northeast region.

Table 27 shows the performance comparison within content area between tests. The results indicate that CKT is a somewhat more difficult test. The mean percent correct values are approximately 25 percentage points lower for both CKT RLA and mathematics compared to PRAXIS. The observed correlations were .72 for RLA and .62 for mathematics. After adjusting for attenuation,[5] these values were .93 for RLA and .74 for mathematics.

As shown in Figures 6 and 7, the respective relative performance distributions for RLA and mathematics differ substantively between PRAXIS and CKT, reflecting the overall difficulty of each test.

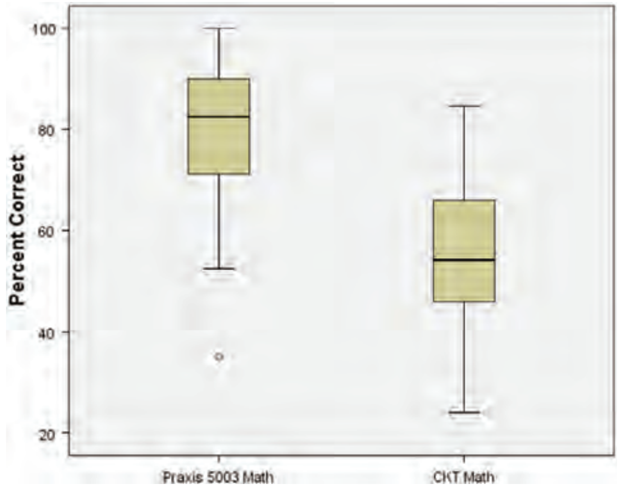

Figure 7 Percent correct for PRAXIS 5003 Mathematics and Content Knowledge for Teaching Mathematics.

Discussion

The purpose of the of CKT Elementary RLA and Mathematics pilot was to collect initial evidence on test quality (ETS, 2014). One focus was on "pretesting" the items and subject matter forms to learn about the statistical characteristics of the new items (ETS, 2014, p. 32). The main analyses examined how the newly developed CKT assessment items function, including item difficulty and discrimination, testing time, reliability of the pool of administered items for each content area, and classification accuracy and consistency. The results from these analyses were favorable and suggest that CKT items are likely to perform adequately in an operational context for use in making licensure decisions.

A second set of analyses provided preliminary validity evidence for the CKT items and for their intended use in the licensure context. First, we examined whether the scores are sensitive to professional preparation (i.e., methods preparation and professional experience). Even though the variation in the sample on these covariates was limited (because the participants were enrolled in teacher preparation programs and selected to be at similar points in their programs), there was evidence that scores are associated with indicators of general academic success and rigor, academic attainment, and teaching experience.

We also examined the extent to which the newly developed CKT items assess competencies that differ from assessments that focus on the content that students are learning in K–12 schooling (i.e., the content of the student curriculum). We found that, on average, the newly developed CKT items are more difficult for candidates than more traditional items used on other versions of PRAXIS subject matter tests. However, when the correlations of the CKT and PRAXIS tests were adjusted for measurement error, both RLA and mathematics tests provided similar rank ordering of the participants.

A third set of analyses provided evidence on score differences for teachers from different demographic groups. We focused on differences for Latinx and African American prospective teachers because teachers from both of these groups have historically been underrepresented in the teaching workforce. Any new licensure test has the potential of having an adverse impact on the diversity of the teaching workforce. It is important to start examining score differences as soon as possible to consider the impact on pass rates. Like other assessments that focus on teacher content knowledge, we did find score differences with the sample of minority teachers generally scoring lower than teachers who identified as White. These score differences also led to disproportionate pass rates at different hypothetical passing scores. However, when compared to other similar licensure assessments, these differences are comparable or smaller in magnitude.

Limitations

Several important limitations need to be considered when interpreting the results from this study. First, the study was conducted with a convenience sample of prospective teachers. While the sample roughly matches the general demographics of candidates who might be administered a CKT exam, the actual sample of participants could differ in important ways that might lead to different test results.

One important difference is the condition of the testing. Participants in this pilot were volunteers paid to participate in the study. They were not offered the opportunity to participate in test preparation and their professors had not prepared them explicitly for the content assessed by the CKT test items. In an operational administration the stakes for test takers are not just different but much higher. Tested candidates are more likely to have attended preparation programs where there is both awareness and a focus on test content, where they have access to test preparation materials, and where they have no choice about taking the licensure exam. These differences could lead to different test scores. This could also lead to different associations with other tests that the candidates previously completed. For example, the difference in test scores between the CKT items and other test administrations for PRAXIS could be due to effort or preparation and not the relative difficulty of the items and tests.

The differences in the comparisons across groups of test takers could also be influenced by the particular samples examined in the study. It is possible that candidates who decided to participate in this study may not represent the relevant populations. Given that little is known about selection mechanisms and how the testing conditions might differentially impact the composition of groups by characteristics such as, for example, race/ethnicity or educational and professional attainment, the score comparisons should be interpreted with caution and be considered provisional.

Next Steps

The results presented in this report represent an initial pilot of newly developed items designed to assess CKT. The number of items piloted for RLA and mathematics approximate the plan for an operational test, and the participants who completed the pilot are similar to teacher candidates who will complete an operational test. Overall, the performance of the items and the initial validity evidence are encouraging and indicate that these items are likely to perform adequately in the licensure context. However, because the forms, testing conditions, participant samples, and use of test scores will be different for an operational administration, it will be important to replicate the item and test analyses presented in this report with operational data to ensure that the intended CKT tests meet ETS standards for quality and fairness (ETS, 2014).

The ECD approach that was used to develop the CKT items and subject matter forms represents a new way of both conceptualizing and assessing teacher content knowledge (Phelps et al., 2020). This approach grounds the design of the assessments in the work of teaching. The resulting assessment tasks and associated tests are therefore designed to assess the complex of content knowledge that is needed or called on in the central tasks of teaching these subjects. The generally encouraging results from this pilot study provide support both for continued development of CKT items and for the use of CKT tests to make licensure decisions.

Acknowledgments

The authors thank Ben Baehr and Allison Brettschneider, who led teams developing the Content Knowledge for Teaching (CKT) subject matter sections. The authors would also like to thank the many test developers, analysts, and others involved with the test development and administration, including Burt Fried, Nicole Garcia, Wyatt Gordon, Colleen Knorr, Kevin Larkin, Shuangshuang Liu, Dan McCaffrey, Kathy Miller, Donna Ott, David Pearson, Katie Pedley, Sarah Scott, Sarah Kate Selling, Adele Tan, and Sharon Temmer. The authors would also like to thank our partners at Teaching Works for their input through all stages of the work.

Notes

1 These included bachelor's degree plus additional credits, master's degree, master's degree plus additional credits, education specialist or professional diploma based on at least 1 year's work past a master's degree, doctoral degree, and professional degree (e.g., MD, LLM, JD, DDS).
2 The category of classroom teacher includes full-time, part-time, and substitute teachers as well as those currently working in a school in some other capacity.
3 For a discussion of test fairness and associated testing guidelines, see Educational Testing Service (2014).
4 These differences are largely driven by the original study given the design specified earlier.
5 Over the period of September 2014 to January 2016, the average reliability for PRAXIS 5002 (Reading Language Arts) was .78, and that for PRAXIS 5003 Mathematics was .81.

References

Ball, D. L., & McDiarmid, G. (1990). The subject matter preparation of teachers. In W. R. Houston (Ed.), *Handbook of research on teacher education: A project of the Association of Teacher Educators* (pp. 437–449). Macmillan.

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

Barron's. (2015). *Barron's profile of American colleges* (32nd ed.). Author.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Carlisle, J. F., Correnti, R., Phelps, G., & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing*, *22*(4), 457–486. https://doi.org/10.1007/s11145-009-9165-y

Copur-Gencturk, Y., Tolar, T., Jacobson, E., & Fan, W. (2018). An empirical study of the dimensionality of the mathematical knowledge for teaching construct. *Journal of Teacher Education*, *70*(5), 485–497. https://doi.org/10.1177/0022487118761860

Correnti, R., & Phelps, G. (2010, April 30–May 4). *Investigating the relationship between teachers' knowledge, literacy practice and growth in student learning* [Paper presentation]. Annual meeting of the American Educational Research Association, Denver, CO.

Educational Testing Service (ETS). (2014). *ETS standards for quality and fairness*. Author.

Gitomer, D. H., Phelps, G., Weren, B., Howell, H., & Croft, A. J. (2014). 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). Jossey-Bass. https://doi.org/10.1002/9781119210856.ch15

Gitomer, D. H., & Zisk, R. C. (2015). Knowing what teachers know. *Review of Research in Education*, *39*(1), 1–53. https://doi.org/10.3102/0091732X14557001

Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review*, *29*(3), 432–439. https://doi.org/10.1016/j.econedurev.2009.10.002

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.

Herbst, P., & Kosko K. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. J. Lo, K. Leatham, & L. 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., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, *35*(5), 330–351. https://doi.org/10.2307/30034819

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

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2–3), 81–92. https://doi.org/10.1080/15366360701486999

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028312042002371

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

Iaconangelo, C., Phelps, G., & Gitomer, D. (2017, April 27–May 1). *Cognitive diagnostic modeling and validation using known groups* [Paper presentation]. Annual meeting of the National Conference on Measurement in Education, New York, NY.

Indiana University Center for Postsecondary Research. (2015). 2015 data file [Data set]. https://carnegieclassifications.iu.edu/downloads.php.

Kersting, N. B. (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

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568–589. https://doi.org/10.3102/0002831212437853

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, *64*, 90–106. https://doi.org/10.1177/0022487112460398

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM*, *40*(50), 873–892. https://doi.org/10.1007/s11858-008-0141-9

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197. https://doi.org/10.1111/j.1745-3984.1995.tb00462.x

Marks, R. (1990). *Pedagogical content knowledge in elementary mathematics* (Unpublished doctoral dissertation). Stanford University, Palo Alto, CA.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M., & Senk, S. (2012). Knowledge of algebra for teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education*, *43*(5), 584–615. https://doi.org/10.5951/jresematheduc.43.5.0584

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). Educational Testing Service. https://doi.org/10.1002/ets2.12207

Mikeska, J. N., Phelps, G., & Croft, A. (2017). *Practice-based measures of elementary science teachers' content knowledge for teaching: Initial item development and validity evidence* (Research Report No. RR-17-43). Educational Testing Service. https://doi.org/10.1002/ets2.12168

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Mohr-Schroeder, M., Ronau, R. N., Peters, S., Lee, C. W., & Bush, W. S. (2017). Predicting student achievement using measures of teachers' knowledge for teaching geometry. *Journal for Research in Mathematics Education*, *48*(5), 520–566. https://doi.org/10.5951/jresematheduc.48.5

Nettles, M. T., Scatton, L. H., Steinberg, J. H., & Tyler, L. L. (2011). *Performance and passing rate differences of African American and White prospective teachers on PRAXIS examinations* (Research Report RR-11-08). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2011.tb02244.x

Phelps, G. (2005). *Content knowledge for teaching reading* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor.

Phelps, G. (2009). Just knowing how to read isn't enough! What teachers know about the content of reading. *Educational Assessment, Evaluation, and Accountability*, *21*(2), 137–154. https://doi.org/10.1007/s11092-009-9070-6

Phelps, G., Corey, D., DeMonte, J., Harrison, D., & Ball, D. L. (2012). How much English language arts and mathematics instruction do students receive? Investigating variation in instructional time. *Educational Policy*, *26*(5), 631–662. https://doi.org/10.1177/0895904811417580

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., Gitomer, G., Weren, B., & Croft. A. (2014). *Developing content knowledge for teaching assessments for the measures of effective teaching study* (Research Report No. RR-14-33). Educational Testing Service. https://doi.org/10.1002/ets2.12031

Phelps, G., Howell, H., & Liu, S. (2019). Exploring differences in mathematical knowledge for teaching for prospective and practicing teachers. *ZDM*, *52*(2), 252–268. https://doi.org/10.1007/s11858-019-01097-x.

Phelps, G., Kelcey, B., Jones, N., & Liu, S. (2016). Informing estimates of program effects for studies of mathematics professional development using teacher content knowledge outcomes. *Evaluation Review*, *40*(5), 383–409. https://doi.org/10.1177/0193841X16665024

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

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, *8*, 255–281. https://doi.org/10.1007/s10857-005-0853-5

Saderholm, J., Ronau, R., Todd, B. E., & Collins, G. (2010). Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) instrument. *School Science and Mathematics*, *110*(4), 180–192. https://doi.org/10.1111/j.1949-8594.2010.00021.x

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

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

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

Steinberg, J., Brenneman, M., Castellano, K., Lin, P., & Miller, S. (2014). *A comparison of achievement gaps and test-taker characteristics on computer-delivered and paper-delivered PRAXIS I tests* (Research Report No. RR-14-13). Educational Testing Service. https://doi.org/10.1002/ets2.12033

Steinberg, J., Ling, G., & Delaney, C. (2016, April 8-22). *Balancing quality and opportunity for elementary education licensure candidates within multiple frameworks* [Roundtable presentation]. Annual meeting of the American Educational Research Association, Washington, DC.

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*. Teacher Education and Development International Study Center, College of Education, Michigan State University.

Tröbst, S., Kleickmann, T., Heinze, A., Bernholt, A., Rink, R., & Kunter, M. (2018). Teacher knowledge experiment: Testing mechanisms underlying the formation of preservice elementary school teachers' pedagogical content knowledge concerning fractions and fractional arithmetic. *Journal of Educational Psychology*, *110*(8), 1049–1065. https://doi.org/10.1037/edu0000260

U.S. Department of Education. 2017. *Institutional characteristics*. National Center for Education Statistics, Integrated Postsecondary Education Data System. https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*(6), 673–695. https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<673::AID-TEA5>3.0.CO;2-J

Wilson, S., Shulman, L., & Richert, A. (1987). "150 different ways of knowing": Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 104–123). Cassell.

Wineburg, S. (1990). *Historical problem-solving: A study of the cognitive processes used in the evaluation of documentary evidence* (Unpublished doctoral dissertation). Stanford University, Palo Alto, CA.

Appendix A

Content and Task of Teaching Frameworks for Reading Language Arts and Mathematics

Table A1 Reading Language Arts Content Framework

Core ideas	Component ideas
Foundational literacy skills	Print concepts
	Alphabetic principle
	Phonological awareness
	Phonics and word recognition
	Fluency
Language	Conventions of standard academic English
	Vocabulary
	Forms and functions of language
Constructing meaning	Key ideas and details
	Author's craft and structure
	Integration and application of knowledge
	Text types
	Production of written texts
	Research to build and present knowledge
	Discussion and collaboration
	Presentation of knowledge and ideas

Table A2 Reading Language Arts Tasks of Teaching Framework

Core task of teaching	Component task of teaching
Planning and facilitating instruction	Evaluating texts, examples, and graphic representations for their support of particular RLA instructional goals
	Creating and modifying texts, examples, and graphic representations to support particular RLA instructional goals, including differentiation for particular learners
	Analyzing language and language systems
	Explaining, defining, and demonstrating RLA processes and concepts for students
	Facilitating class discussions and conversations with individual students to elicit or develop their thinking about particular RLA content
	Evaluating instructional strategies and activities to elicit, develop, or assess students' thinking about particular RLA content or to develop or assess their facility with particular RLA processes
Analyzing student learning	Evaluating student reading, writing, speaking, and listening to identify specific strengths and/or areas for improvement or instructional focus
	Evaluating student reading, writing, speaking, or listening to classify students' level of literacy development
	Analyzing student reading, writing, speaking, or listening to identify patterns of thinking, cuing systems, misconceptions, and partial conceptions
	Responding to student reading, writing, speaking, or listening to target the particular content issue in need of attention

Table A3 Mathematics Content Framework

Core ideas	Component ideas
Counting	Know number names and count sequence
	Count to tell the number of objects
	Compare numbers
	Extend the counting sequence
	Understand place value
Operations with whole numbers	Understand the four operations
	Represent/solve problems using four operations
	Understand and apply properties of operations
	Work with equations using four operations
	Identify and explain patterns in arithmetic
	Use the four operations with whole numbers
	Gain familiarity with factors and multiples
	Perform multidigit arithmetic
	Solve problems involving measurement
	Compute fluently with multidigit numbers
Place value and decimals	Understand place value
	Use place value with operations
	Generalize understanding to multidigit numbers
	Perform multidigit arithmetic
	Understand decimal notation and compare
	Understand the place value system
	Perform operations with multidigit numbers
Fractions, operations with fractions, and ratios	Reason with shapes and their attributes
	Develop understanding of fractions as numbers
	Understand fraction equivalence and ordering
	Build fractions from unit fractions
	Understand decimal notation for fractions
	Solve problems involving measurement
	Use equivalent fractions to perform operations
	Understand ratio concepts
	Use ratio reasoning to solve problems
Early equations and expressions	Represent and solve equations with operations
	Identify and explain patterns in arithmetic
	Use operations with whole numbers
	Write and interpret numerical expressions
	Extend understanding to algebraic expressions
	Solve one-variable equations and inequalities
	Analyze dependent and independent variables
Measurement	Describe/compare measurable attributes
	Measure lengths indirectly
	Measure and estimate length in standard units
	Relate operations to length
	Reason with shapes and their attributes
	Solve problems using estimations
	Understand concept of area
	Recognize perimeter of plane figures
	Distinguish between linear and area measures
	Conversion of measurements
	Understand concept of volume
Geometry	Identify and describe key shapes
	Analyze, compare, create, and compose shapes
	Reason with shapes and their attributes
	Understand concept of angle and measure angles
	Draw and identify lines and angles and measure
	Classify shapes by properties of lines and angles
	Classify two-dimensional figures into categories
	Solve problems of area, surface area, and volume

Table A4 Mathematics Tasks of Teaching Framework

Core task of teaching	Component task of teaching
Explanations, conjectures, and definitions	Giving mathematically valid explanations for a process, conjecture, or relationship
	Evaluating mathematical explanations for their validity, generalizability, explanatory power, and/or completeness
	Determining the changes that would improve the validity, generalizability, completeness, and/or precision of a mathematical explanation
	Evaluating a student conjecture for its validity and/or generalizability on a given domain
	Evaluating mathematical definitions or other mathematical language for precision, validity, generalizability, and/or usefulness in a particular context and/or support for an instructional goal
Problems, examples, and structure	Evaluating mathematical problems for how well they elicit a particular idea, support the use of a particular solution strategy or practice, fit a particular mathematical structure, address the same concept as another problem, or assess a particular student conception or error
	Writing mathematical problems that fit a particular solution strategy or mathematical structure
	Evaluating examples for how well they introduce a concept; illustrate an idea or relationship; illustrate the appropriateness of a strategy, procedure, or practice; or address particular student questions, misconceptions, or partial conceptions
	Generating or identifying nonexamples or counterexamples to highlight a mathematical distinction or to demonstrate why a student conjecture is incorrect or partially incorrect
	Choosing which mathematical topics are most closely related to a particular instructional goal
Representations and manipulatives	Selecting, creating, or evaluating representations or manipulatives for a mathematical purpose or to show a particular mathematical idea
	Evaluating how representations or manipulatives have been used to show particular mathematical ideas, relationships between ideas, mathematical processes, or strategies in a text, talk, or written work
Student strategies and errors	Determining whether student work demonstrates the use of a particular mathematical idea or strategy
	Determining whether a strategy is mathematically valid or generalizable
	Interpreting a student's mathematical error, including anticipating how it would replicate across similar problems, and choosing other work samples that demonstrate the same error
	Identifying tasks or situations in which student work or talk that seems mathematically valid might mask incorrect thinking

Appendix B

Sample Content Knowledge for Teaching Items

The sample items presented below are from ETS's PRAXIS Study Companion, Elementary Education: Content Knowledge for Teaching (7801) and represent the task design of the items discussed in this report. The study companion can be retrieved from https://www.ets.org/praxis/prepare/materials/

Reading Language Arts Items

1. A teacher is administering an informal reading assessment that includes the following sentences.

   Monkeys like to play together. They wrestle and roll.

   When reading the sentences, one student says the word "wiggle" instead of "wrestle." The student is likely using which of the following cuing systems?

   Select all that apply.

   (A) Semantic
   (B) Syntactic
   (C) Graphophonemic

How to Answer the Question Above
This is a "select all that apply" question. You should select one, two, or all three of the answer choices—however many are correct. You earn credit for the question only if you select all of the choices that are correct answers. In the actual test, the answer choices appear next to empty check boxes. Click on a box to select the choice next to it; this causes an "x" to appear in the box. If you change your mind, click the box again to remove the "x."

3. A student uses a "W" to represent the initial sound in the word "doughnut." In this scenario, which of the following patterns of thinking is the student demonstrating?

   Select all that apply.

   (A) Believing that each sound can be represented by only one letter
   (B) Thinking that the position of a letter within a word does not affect the sound it makes
   (C) Confusing a sound in a letter name with a sound represented by the letter

How to Answer the Question Above
This is a "select all that apply" question. You should select one, two, or all three of the answer choices—however many are correct. You earn credit for the question only if you select all of the choices that are correct answers. In the actual test, the answer choices appear next to empty check boxes. Click on a box to select the choice next to it; this causes an "x" to appear in the box. If you change your mind, click the box again to remove the "x."

The remaining sample test questions for reading and language arts are organized into sets in which two or more questions refer to the same instructional scenario and text; the text is either a reading passage or an example of student work. For some sets, the scenario and text are shown only at the beginning of the set. However, for other sets that contain questions referring to highlighted portions of the text, the scenario and text appear a second or third time, to show the highlighting described in the question or questions. A scenario and text may also be repeated so that they can be viewed on the same page as each test question that refers to them. On the actual test, the scenario and text will always appear on the left-hand side of the screen, with whatever highlighting is needed for a given question, and the questions will appear on the right-hand side.

Questions 5–7 refer to the following scenario.

Ms. Weise is reading and discussing "The Tortoise and the Hare" with her class.

One day, a hare was boasting to all the other animals in the forest. "I'm faster than the wind," he said. "No one has ever beaten me. No one ever will!"

"Does anyone here want to challenge me to a race?" the hare dared the animals of the forest. The foxes and frogs and snakes all stared back silently.

Then the hare heard a modest voice say, "I will race you." It was the tortoise.

The hare guffawed. "You? What a funny joke! I can run circles around you!"

"Perhaps," the tortoise said, smiling to himself. "Shall we race?"

"On your mark, get set, go!" chanted the nearby animals. In a flash, the hare was off and out of sight. The tortoise set off slowly, just plodding along steadily, one heavy foot in front of the other.

After a while the hare looked back and could not see the tortoise anywhere. It was a hot day. The hare said to himself, "I'm so far ahead that I can win this race easily. It won't hurt if I just take a little rest." So he found a shady spot and lay down on a patch of grass.

When he awoke, the sun hung lower in the sky. The hare looked for the tortoise but didn't see him. "I might as well go finish the race," he said, heading towards the finish line. But when he got close, he saw the tortoise crawling over the finish line before him. "Oh no!" the hare said.

All the animals cheered, "Tortoise won! Tortoise won!" And the tortoise smiled to himself again.

5. Which of the following student statements about the story required the student to make a text-based inference?
   (A) The hare thinks he is going to win the race because he wins every race.
   (B) The foxes, frogs, and snakes are afraid to race against the hare.
   (C) The hare laughs at the tortoise when he says he wants to race.
   (D) The hare takes a nap because he thinks he's way ahead of the tortoise.

Ms. Weise is reading and discussing "The Tortoise and the Hare" with her class.

One day, a hare was boasting to all the other animals in the forest. "I'm faster than the wind," he said. "No one has ever beaten me. No one ever will!"

"Does anyone here want to challenge me to a race?" the hare dared the animals of the forest. The foxes and frogs and snakes all stared back silently.

Then the hare heard a modest voice say, "I will race you." It was the tortoise.

The hare guffawed. "You? What a funny joke! I can run circles around you!"

"Perhaps," the tortoise said, smiling to himself. "Shall we race?"

"On your mark, get set, go!" chanted the nearby animals. In a flash, the hare was off and out of sight. The tortoise set off slowly, just plodding along steadily, one heavy foot in front of the other.

After a while the hare looked back and could not see the tortoise anywhere. It was a hot day. The hare said to himself, "I'm so far ahead that I can win this race easily. It won't hurt if I just take a little rest." So he found a shady spot and lay down on a patch of grass.

When he awoke, the sun hung lower in the sky. The hare looked for the tortoise but didn't see him. "I might as well go finish the race," he said, heading towards the finish line. But when he got close, he saw the tortoise crawling over the finish line before him. "Oh no!" the hare said.

All the animals cheered, "Tortoise won! Tortoise won!" And the tortoise smiled to himself again.

6. Ms. Weise wants students to find examples from "The Tortoise and the Hare" to demonstrate three basic elements of story structure:

   (1) an orientation that sets the scene,

   (2) a complication that is introduced, triggering the main series of events, and

   (3) a resolution.

   Of the three highlighted paragraphs of dialogue, select the <u>one</u> paragraph in which the complication is introduced.

How to Answer the Question Above
This is a "select in passage" question with one correct answer. For this kind of question, you will select your answer directly from the passage, not from a separate list of answer choices. In this case, the three answer choices are the three highlighted paragraphs in the passage, and only one of them is the correct answer. In the actual test, you will answer this question by using your mouse to click on the appropriate paragraph, which will change the color of the highlighting from gray to black. If you change your mind, click on a different highlighted paragraph.

Ms. Weise is reading and discussing "The Tortoise and the Hare" with her class.

One day, a hare was boasting to all the other animals in the forest: "I'm faster than the wind," he said. No one has ever beaten me. No one ever will!"

"Does anyone here want to challenge me to a race?" the hare dared the animals of the forest. The foxes and frogs and snakes all stared back silently.

Then the hare heard a modest voice say, "I will race you." It was the tortoise.

The hare guffawed. "You? What a funny joke! I can run circles around you!"

"Perhaps," the tortoise said, smiling to himself. "Shall we race?"

"On your mark, get set, go!" chanted the nearby animals. In a flash, the hare was off and out of sight. The tortoise set off slowly, just plodding along steadily, one heavy foot in front of the other.

After a while the hare looked back and could not see the tortoise anywhere. It was a hot day. The hare said to himself, "I'm so far ahead that I can win this race easily. It won't hurt if I just take a little rest." So he found a shady spot and lay down on a patch of grass.

When he awoke, the sun hung lower in the sky. The hare looked for the tortoise but didn't see him. "I might as well go finish the race," he said, heading towards the finish line. But when he got close, he saw the tortoise crawling over the finish line before him. "Oh no!" the hare said.

All the animals cheered, "Tortoise won! Tortoise won!" And the tortoise smiled to himself again.

7. **Part A**

Which of the following pairs of words is most appropriate for Ms. Weise to introduce to students to support their discussion of the main characters' personality traits?

(A) Vengeful, unknowing

(B) Jealous, supportive

(C) Lazy, unathletic

(D) Determined, arrogant

Part B

Review the five highlighted sentences from the story. Which two sentences best support the answer to part A? One sentence should support the first word, and the other sentence should support the second word.

How to Answer the Question Above

This is a two-part question that is worth two points and eligible for partial credit. Each part of a two-part question may follow any of the question formats on the test. For all two-part questions, the answer to part B is reliant on the answer to part A. If you answer part A incorrectly, you cannot earn credit for part B. However, you can earn one point if you answer part A correctly but do not answer part B correctly.

In this question, part A follows a standard multiple-choice format with four answer choices and one correct answer. Part B follows a "select in passage" format, with five answer choices highlighted in the passage and two correct answers; both of these answers—and no other choices—must be selected to earn credit for part B. In the actual test, you answer part B by using your mouse to click on the appropriate highlighted sentences, which changes the color of the highlighting from gray to black. To change one of your answers, click on the selected sentence again to revert back to gray highlighting and then click on a different highlighted sentence.

Reading Language Arts Answers

1. The correct answers are (A), (B), and (C). "Wiggle" makes sense in the sentence, suggesting that the student may be using semantic cuing (A). Because "wiggle" is a verb, it fits the sentence syntactically (B), so the student may also be using syntactic cuing. Finally, "wiggle" looks similar to "wrestle," since it shares the first letter and the last two letters and is approximately the same length, suggesting that the student may also be using graphophonemic cuing (C).

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching ELA	9. Analyzing student reading, writing, speaking, or listening to identify patterns of thinking, cuing systems, misconceptions, and partial conceptions
Topic	I. Foundational Skills
Subtopic	D. Phonics and Word Recognition

3. The correct answer is (C) only. The student's use of "w" to represent the /d/ in "doughnut" suggests that the student has concluded that because the letter name for "w" starts with the sound /d/ (as in "double-u"), the correct letter to represent /d/ is "w." The substitution does not suggest a misunderstanding about the impact of position on sound (B) or about how many letters can represent a single sound (A).

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching ELA	7. Evaluating student reading, writing, speaking, and listening to identify specific strengths and/or areas for improvement or instructional focus.
Topic	I. Foundational Skills
Subtopic	B. Alphabetic Principle

5. The correct answer is (B) because the text does not state explicitly that the animals are afraid to race the hare. However, students can make this inference because it says that the animals "stared back silently" when the hare issued his challenge. The hare's intimidating previous statement "No one has ever beaten me. No one ever will!" offers additional support for the inference. (A) is incorrect because it simply paraphrases the hare's own statement "No one has ever beaten me. No one ever will!" It does not require students to make an inference. (C) is incorrect. Although students may need to consult a dictionary to understand that "guffawed" means "laughed," they do not need to make an inference. (D) is incorrect because, like (A), it simply paraphrases the hare's own statement, "I'm so far ahead that I can win this race easily. It won't hurt if I just take a little rest."

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching ELA	9. Analyzing student reading, writing, speaking, or listening to identify patterns of thinking, cuing systems, misconceptions, and partial conceptions
Topic	III. Constructing Meaning
Subtopic	A. Key Ideas and Details

6. The correct paragraph to highlight is "Then the hare heard a modest voice say, 'I will race you.' It was the tortoise." This paragraph introduces the complication: the tortoise's surprising challenge to the hare. This complication, in turn, triggers the main series of events in the story, including the hare's condescending reaction and the tortoise's confident reply (both depicted in the other two selectable paragraphs), the race, and the tortoise's victory (the resolution).

Content Knowledge Type	Work of the Student Curriculum
Task of Teaching ELA	not applicable
Topic	III. Constructing Meaning
Subtopic	B. Author's Craft and Text Structure

7. This is a two-part, or composite, test question. In order to answer part B correctly, you must answer part A correctly. For part A, the correct answer is (D). The characteristic "determined" best matches the personality of the tortoise, who is described in the story as setting "off slowly, just plodding along steadily, one heavy foot in front of the other." This is the first highlighted sentence that should be selected for part B. The characteristic "arrogant" best matches the personality of the hare, who makes haughty statements such as "I'm faster than the wind." This is the second highlighted sentence that should be selected for part B. In part A, (A) is incorrect. Although the hare might be described as "unknowing" because he is unable to imagine being beaten by the tortoise, this is not his dominant personality trait in the story. In addition, "vengeful" is not an appropriate description for either character. (B) is incorrect because although the hare is likely jealous of the tortoise's victory, this jealousy is only implied at the very end of the story; it is not a dominant personality trait. Further, neither character could be described as "supportive" of the other. Finally, (C) is incorrect because although the hare does behave in a lazy way by taking a nap in the middle of the race, neither character could be described as unathletic; the hare is a skilled racer, and the tortoise, by winning the race, shows that he is also athletic.

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching ELA	1. Evaluating texts, examples, and graphic representations for their support of particular ELA instructional goals
Topic	III. Constructing Meaning
Subtopic	A. Key Ideas and Details

Mathematics Items

1. Which <u>three</u> of the following expressions are equivalent to 3,956 ´ 4 ?

   (A) 3,000 ´ 4 + 900 ´ 4 + 50 ´ 4 + 6 ´ 4

   (B) (4,000 ´ 4 − 100 ´ 4) + (60 ´ 4 − 4 ´ 4)

   (C) 4 ´ 3 + 4 ´ 9 + 4 ´ 5 + 4 ´ 6

   (D) 4,000 ´ 4 − 40 ´ 4 − 4 ´ 4

   (E) 3 ´ 1,000 ´ 4 + 95 ´ 100 ´ 4 + 6 ´ 1 ´ 4

How to Answer the Question Above
This is a multiple-choice question with <u>three</u> correct answers. You must select all three correct answers—and no incorrect answers—to earn credit for the question. In the actual test, the answer choices appear next to empty check boxes. Click on a box to select the answer choice next to it; this causes an "x" to appear in the box. If you change your mind, click the box again to remove the 'x."

4.

385	453	321
+ 462	+ 427	+ 836
7147	8710	1157

Josh is a third-grade student in Ms. Carter's classroom. Josh's answers to three addition problems are shown. He incorrectly answered the first two problems but correctly answered the third problem.

If Josh uses the same strategy to answer the following problem, what will his answer be?

$$\begin{array}{r} 328 \\ +\,564 \\ \hline \boxed{} \end{array}$$

How to Answer the Question Above
This is a numeric-entry test question. It requires you to enter a number in the box rather than select a number from a list of answer choices. In the actual test, simply type in the number. Backspace to erase.

3. Dora made a pile of 5 counters. Then Mr. Levy asked her to add counters to her pile of 5 so that the pile would have 7 counters. Dora counted out 7 more counters and added them to the pile of 5 counters.

   Which of the following most likely explains the reason behind Dora's error?

   (A) Dora does not fully understand one-to-one correspondence between numbers and objects.

   (B) Dora does not yet have a concept of the quantity 7.

   (C) Dora does not yet understand that one quantity can be composed of two smaller quantities.

   (D) Dora does not yet know her number facts for sums greater than 10.

5. Mr. Keller's sixth-grade class is learning about algebraic equations. In his teachers' edition of the textbook, Mr. Keller finds a page that suggests he ask students to critique the following two solutions to determine whether they are valid.

$4x + 2 = 12$	$5 = 2x + 3$
$6x = 12$	$5 = 5x$
$x = 12 \div 6$	$5 \div 5 = x$
$x = 2$	$x = 1$

   Which of the following is most clearly highlighted by asking students to critique the invalid strategies?

   (A) Understanding the meaning of the equal sign

   (B) Understanding the importance of combining like terms

   (C) Understanding the use of properties of operations to simplify expressions

   (D) Understanding the use of inverse operations to solve equations

13. Ms. Kress asked her students to compare $\frac{1}{3}$ and $\frac{7}{8}$. Four of her students correctly answered that $\frac{7}{8}$ is greater than $\frac{1}{3}$, but they gave different explanations when asked to describe their strategies to the class.

Indicate whether each of the following student explanations provides evidence of a mathematically valid strategy for comparing $\frac{1}{3}$ and $\frac{7}{8}$.

Student Explanation	Provides Evidence	Does Not Provide Evidence
When you look at the numbers, you see that 7 is bigger than 1, so $\frac{7}{8}$ is the bigger fraction.		
In the first fraction, 1 is less than half of 3, but in the second, 7 is more than half of 8, so $\frac{7}{8}$ is larger than $\frac{1}{3}$.		
I multiplied 1 times 7 and 3 times 7, so $\frac{1}{3}$ is the same as $\frac{7}{21}$. This means that $\frac{7}{8}$ is bigger than $\frac{1}{3}$ because $\frac{1}{8}$ is bigger than $\frac{1}{21}$.		
I wanted to make a fraction equal to $\frac{1}{3}$ with the same bottom number as $\frac{7}{8}$, so I added 5 to 3 and got 8. Then I added 5 to 1 and got 6, but 7 is greater than 6, so $\frac{7}{8}$ is greater.		

How to Answer the Question Above

This is a table question worth two points and eligible for partial credit. It requires you to select one choice for each row. You will receive full credit (two points) if all rows are completed correctly. You will receive one point if all but one row is completed correctly. In the actual test, click on a box to select it; a check mark will appear. If you change your mind, click on the check mark to remove it, or simply click on another box in the same row, and your check mark will move to the new box.

Mathematics Answers

1. The correct answers are (A), (B), and (D). Since 3,956 can be written as $3{,}000 + 900 + 50 + 6$, the given expression is equivalent to $(3{,}000 + 900 + 50 + 6) \times 4$. Applying the distributive property yields $3{,}000 \times 4 + 900 \times 4 + 50 \times 4 + 6 \times 4$, which is the expression in (A). Since 3,956 can be written as $3{,}900 + 56$, the given expression is equivalent to $(3{,}900 + 56) \times 4$. Applying the distributive property yields $3{,}900 \times 4 + 56 \times 4$. One can rewrite 3,900 as $4{,}000 - 100$ and 56 as $60 - 4$, which yields the equivalent expression $(4{,}000 - 100) \times 4 + (60 - 4) \times 4$. Applying the distributive property again yields $(4{,}000 - 100) \times 4 + (60 - 4) \times 4$, which is the expression in (B). Since 3,956 can be written as $4{,}000 - 40 - 4$, the given expression is equivalent to $(4{,}000 - 40 - 4) \times 4$. Applying the distributive property yields $4{,}000 \times 4 - 40 \times 4 - 4 \times 4$, which is the expression in (D). Applying the distributive property to the expression in (C) yields $4 \times (3 + 9 + 5 + 6)$, which is equivalent to 4×23, but this expression is not equivalent to the given expression. Since $6 \times 1 = 6$, applying the distributive property to the expression in (E) yields $(3 \times 1{,}000 + 95 \times 100 + 6) \times 4$, which is equivalent to $(3{,}000 + 9{,}500 + 6) \times 4$. The sum of the numbers in the parentheses is 12,506; therefore, the expression in (E) is not equivalent to the given expression.

Content Knowledge Type	Work of the Student Curriculum
Task of Teaching Mathematics	not applicable
Topic	II. Place Value and Decimals

3. The correct answer is (C). Dora counted out 7 more counters, not realizing that 5 can be part of 7, so she does not seem to understand that one quantity can be composed of two smaller quantities. (A) is not the key because Dora actually counted out 7 more counters, so there is evidence that she does understand one-to-one correspondence. (B) is not the key because Dora counted out 7 counters, so there is evidence that she has a concept of the quantity 7. (D) is not the key because even though Dora made a pile of 12 counters, knowing number facts for sums greater than 10 was not necessary for the original task, which was to add counters to her pile of 5 counters so there would be 7 counters in the pile. Therefore, (D) does not explain the reason behind Dora's error.

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching Mathematics	15. Interpreting a student's mathematical error, including anticipating how it would replicate across similar problems, and choosing other work samples that demonstrate the same error
Topic	I. Counting and Operations with Whole Numbers
Subtopic	A. Counting

4. The correct answer is 8812. Josh's error is that he is not regrouping when necessary; instead, he is just writing the sum of the digits in each place value column. His written answer is correct in the third problem because 11 hundreds (the result of adding 3 hundreds and 8 hundreds) is equivalent to regrouping to get 1100. However, when he does not regroup in the first two problems, his written answers are incorrect. For example, in the first problem, Josh adds 8 tens and 6 tens to get 14 tens, but instead of regrouping 10 of those tens to get 100 and then writing the final answer as 847, Josh just adds the 3 hundreds and the 4 hundreds and then writes the final answer as 7147. Therefore, if Josh uses the same method in the last problem, he will add 8 and 4 to get 12 ones, but he will not regroup, and then he will add 2 and 6 to get 8 and 3 and 5 to get 8, and his final answer will be 8812.

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching Mathematics	15. Interpreting a student's mathematical error, including anticipating how it would replicate across similar problems, and choosing other work samples that demonstrate the same error
Topic	I. Counting and Operations with Whole Numbers
Subtopic	B. Operations with Whole Numbers

5. The correct answer is (B). In the first solution, $4x$ and 2 are added to get $6x$, but the $4x$ term contains a variable, whereas the 2 is a constant term; it is incorrect to add $4x$ and 2 because they are not like terms. Similarly, in the second solution, $2x$ and 3 are added to get $5x$, but $2x$ and 3 are not like terms, so this strategy is not valid. Therefore, understanding the importance of combining like terms is the answer choice that is most clearly highlighted by asking students to critique the two invalid strategies.

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching Mathematics	8. Evaluating examples for how well they introduce a concept; illustrate an idea or relationship; illustrate the appropriateness of a strategy, procedure or practice; or address particular student questions, misconceptions, or partial conceptions
Topic	IV. Early Equations and Expressions, Measurement, and Geometry
Subtopic	A. Early Equations and Expressions

13. The first and fourth explanations do not provide evidence of a mathematically valid strategy for comparing $\frac{1}{3}$ and $\frac{7}{8}$, but the second and third explanations do. In the first explanation, the student only compares the numerators of the fractions, which is not a valid strategy because it does not take into account the effect of the denominator on the size of the pieces. In the second explanation, the student compares both fractions to the benchmark fraction $\frac{1}{2}$, which is a valid strategy since $\frac{1}{3}$ is less than $\frac{1}{2}$ and $\frac{7}{8}$ is greater than $\frac{1}{2}$. In the third explanation, the student uses multiplicative reasoning to find a common numerator, and then the student compares the fractions by reasoning about the sizes of the unit fractions $\frac{1}{8}$ and $\frac{1}{21}$. This is a valid strategy. In the fourth explanation, the student uses additive reasoning to try to find a fraction equivalent to $\frac{1}{3}$ that has a denominator of 8, but $\frac{6}{8}$ is not equivalent to $\frac{1}{3}$, so this strategy is not valid.

Content Knowledge Type	Specialized Content Knowledge
Task of Teaching Mathematics	2. Evaluating mathematical explanations for their validity, generalizability, explanatory power, and/or completeness
Topic	III. Fractions, Operations with Fractions, and Ratios

Suggested citation:

Phelps, G., Steinberg, J., Leusner, D., Minsky, J., Castellano, K. & McCulla, L. (2020). PRAXIS® *content knowledge for teaching: Initial reliability and validity results for elementary reading language arts and mathematics* (Research Report No. RR-20-15). Educational Testing Service. https://doi.org/10.1002/ets2.12295

Action Editor: Elizabeth Stone

Reviewers: Brent Bridgeman and Gary Sykes

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/