No matter the domain, licensure assessments typically measure the knowledge, skills, and abilities (KSAs) required for performing a job, rather than actual performance on particular job tasks or practices (Wang, Schnipke, & Witt, 2005). It is important that licensure assessments test job-related content, and job-analytic evidence supports that aim (Raymond, 2001). The specifications for a licensure test outline the content of the assessment and the KSAs that should be measured therein (Raymond, 1996) and are a crucial component of validity evidence (Ebel & Frisbie, 1991). Thus, to develop test specifications that are closely tied to performance in a particular profession, it is important to identify the KSAs required for performing critical job practices.

One job-analytic strategy commonly employed to describe the content domain for a licensure test requires a panel of subject matter experts (SMEs) to create a list of KSAs necessary for the effective completion of job tasks (Rosenfeld & Tannenbaum, 1991; Tannenbaum & Wesley, 1993; Wang et al., 2005). It is important that this panel of SMEs be diverse and representative of a variety of work backgrounds and job positions within a particular field (Raymond, 2001; Raymond & Luecht, 2013). For example, in the domain of teaching, it may be desirable to collect information from both teachers and teacher educators. SME background characteristics (e.g., race, ethnicity, gender, urban/rural setting, geographic location) should be taken into account when identifying and choosing qualified experts for this purpose (Clauser et al., 2006; Tannenbaum & Wesley, 1993). It is crucial that the final group of SMEs reflects perspectives from different demographic groups, as respondents with different characteristics may vary in the type of communities they serve and may encounter different kinds of work-related issues (Raymond, 2001).

A common second step to this approach is to survey a large sample of qualified practitioners in a given profession to verify the importance of the list of KSAs created by the panel of SMEs (Rosenfeld & Tannenbaum, 1991; Tannenbaum & Wesley, 1993). The SMEs surveyed are often asked to rate the KSAs regarding their relevance and importance (Kane, Kingsbury, Colton, & Estes, 1989; Raymond, 2005; Tannenbaum & Wesley, 1993), and sometimes, the frequency with which particular practices are performed (Wang et al., 2005). These ratings can then be used to create empirically derived specifications for a licensure test or to provide evidence to support test specifications derived using the input of the aforementioned panel of experts. For teacher licensure, judgments regarding relevance, importance, and frequency should be similar across different educator roles, grade levels taught, geographic regions, and races/ethnicities and should indicate that the KSAs and job practices included in the survey are judged as being of sufficient relevance, importance, and frequency for inclusion in the licensure exam across subgroups.

The survey methodology tends to be beneficial, as a large sample of experts can provide information regarding numerous knowledge or skill statements across multiple locations in a way that is very efficient (Raymond, 2001, 2005). Research suggests that adequate generalizability can be garnered from samples of 200 to 400 respondents (Kane, Miller, Trine, Becker, & Carson, 1995). A sample of at least this size, compared to smaller samples, improves the precision of statistics and analyses, reduces bias, and improves the generalizability of findings to the broader population of workers in a particular occupation (Raymond, 2001). Another advantage of using a structured survey methodology is that it allows for statistical analyses that allow practitioners to draw meaningful conclusions (Raymond & Luecht, 2013).

However, it is worth noting that survey methodology is not without its drawbacks. First, it may be difficult to gather complex information using a structured survey (Raymond & Luecht, 2013). Second, as a survey does not allow for an opportunity to interact with respondents, instructions and questions must be very clear to avoid discrepancies in interpretation (Raymond, 2005). Finally, web-based surveys may be associated with lower response rates (Cook, Heath, & Thompson, 2000) and produce samples that are not necessarily representative of the target group.

Once survey information has been collected, this information can be used to support and verify test specifications. Information that is ultimately included in a licensure test should be deemed both relevant and important for performing a job by the sample of SMEs surveyed (Tannenbaum & Rosenfeld, 1994). If a particular KSA is not considered relevant or important for a beginning teacher to be able to safely and effectively practice, that KSA should not be assessed within a licensure test. Thus, data collected from the survey may be also be used to modify test specifications as necessary.

The Current Study

In this study, we examined evidence regarding the content-related validity for the science domain of a licensure exam for elementary school teachers. This exam is called the ETS *PRAXIS* Elementary Education: Content Knowledge for Teaching (CKT) assessment and measures subject-specific content knowledge, with a focus on specialized content knowledge used in elementary teaching, across four domains—English language arts, mathematics, science, and social studies.[1] The assessment must have valid test specifications for defining the content domain (American Educational Research Association et al., 2014).

We built from prior work (see Mikeska, Kurzum, Steinberg, & Xu, 2018) detailing the identification, selection, and compilation of science content knowledge topics and science teaching practices germane to a beginning elementary school science teacher's ability to effectively educate students. As this work is described in great detail in Mikeska et al. (2018), we present only a summary of the process used to define the science content knowledge domain.

Developing the Test Specifications for the ETS *PRAXIS*® Elementary Education: Content Knowledge for Teaching Science Licensure Exam

The CKT science assessment is designed to assess elementary teacher candidates on the knowledge that is most critical to support student learning in elementary science. As mentioned earlier, the first step in this process is to create a set of test specifications that clearly identify the KSAs considered relevant or important for a beginning elementary teacher to be able to effectively teach science at this level. For this assessment, our research and development team defined these KSAs in two ways. First, we identified the science content—what we call the student-level content domain—that is foundational to the elementary science curriculum and puts students' future academic success at risk when it is not understood well. Second, we identified the science-specific teaching practices—or what others sometimes refer to as the tasks of teaching science—that support student learning of the student-level content domain. We did this because the tasks of teaching are at the intersection of the two parts that form the practice-based content knowledge required to support student learning of the content domain. The questions on this assessment measure the practice-based content knowledge related to a particular concept (e.g., forces and motion) to carry out a particular science teaching practice (e.g., choosing which science ideas or instructional activities are most closely related to a particular instructional goal). Below we describe the process by which we identified the student-level content domain and the teaching practices for elementary science.

Defining the Student-Level Content Domain

For this assessment, the student-level content domain is defined by a set of performance expectations in the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). Each performance expectation addresses at least one disciplinary core idea, one science or engineering practice, and one crosscutting concept.

For example, in the NGSS under the Matter and Its Interactions category, one performance expectation states that students who demonstrate understanding can plan and conduct an investigation to describe and clearly differentiate kinds of materials by their observable properties (see https://www.nextgenscience.org/pe/2-ps1-1-matter-and-its-interactions). As such, each performance expectation specifies the ways in which students are expected to use knowledge of core ideas and concepts to engage in particular scientific practices. Since the NGSS had been recently authored and reviewed by a national panel of science education and content experts with the explicit purpose of identifying the most critical aspects of science disciplinary content knowledge at the K-12 level, we decided to leverage these core science ideas as the basis for the student-level content domain for the CKT science licensure assessment. Figure 1 provides an example of one of the 12 science content areas specified for the student-level content domain and the associated performance expectations (see Appendix A for descriptions of all content areas). The performance expectations noted in Figure 1 come directly from the NGSS (NGSS Lead States, 2013).

Defining the Science Teaching Practices

In order to define the practice-based content knowledge required to teach the student-level content domain, we also had to identify the critical practices that elementary science teachers engage in as they work with students and curriculum and provide instruction. To do so, we reviewed a compilation of resources—including empirical research literature about instructional teaching practices in science and relevant standards—such as the Interstate New Teacher Assessment and Support Consortium (InTASC) Standards (Council of Chief State School Officers, 2011), the CAEP Standards (Council for the Accreditation of Educator Preparation, 2016), and the NGSS (NGSS Lead States, 2013). Based on this review, we proposed a set of key instructional practices that are critical for beginning elementary science teachers to be able to engage in from the first day on the job and that have been either nominated or shown to hold the most promise for improving student outcomes. Table 1 provides some examples of science teaching practices and related assessment examples that beginning elementary science teachers are required to engage in when they enter the science classroom (see Appendix B for all science teaching practices).

Physical Science (PS3): Energy

This topic area focuses on energy—its definition, conservation, and transfer, and how it is captured, stored, and released by physical and chemical processes.

Evidence is used to construct explanations about the relationship between the speed and energy of an object and how energy is transferred from place to place by means such as sound, light, heat, and electrical currents. Sunlight warms the surface of Earth, and solutions can be designed to solve the problem of reducing that warming effect where needed (e.g., umbrellas and awnings).

This topic area also addresses the transfer in energy due to the change in speed or direction that occurs when objects interact. Models can be used to predict the outcomes of certain collisions between objects such as pool balls, marbles, and toy cars. Devices are designed that convert one form of energy (e.g., light energy) to another (e.g., electrical energy). Models, such as energy and matter diagrams and food webs, describe how energy in food sources used by animals for movement, growth, and repair came from the sun and was captured by plants for use and storage.

Definition of Energy, Conservation of Energy and Energy Transfer

4-PS3-1. Use evidence to construct an explanation relating the speed of an object to the energy of that object.

4-PS3-2. Make observations to provide evidence that energy can be transferred from place to place by sound, light, heat, and electric currents.

4-PS3-3. Ask questions and predict outcomes about the changes in energy that occur when objects collide.

4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another.

Relationship Between Energy and Forces

4-PS3-3. Ask questions and predict outcomes about the changes in energy that occur when objects collide.

Energy in Chemical Processes and Everyday Life

5-PS3-1. Use models to describe that energy in animals' food (used for body repair, growth, motion, and to maintain body warmth) was once energy from the sun.

Figure 1. Example of student-level content domain specifications for one content area.

Adapted in part from *The PRAXIS® Study Companion. Elementary Education: Content Knowledge for Teaching* (pp. 111), by Educational Testing Service, 2017, Princeton, NJ, Author. Copyright 2017 Educational Testing Service. Adapted in part from *Next Generation Science Standards: For States, by State*s, by NGSS Lead States, 2013, Washington, DC, The National Academies Press. Copyright 2013 Achieve.

Table 1. Science Teaching Practices and Assessment Examples for Scientific Instructional Goals, Big Ideas, and Topics

Science teaching practices	Assessment examples
Selecting or sequencing age-appropriate, grade-level instructional goals or big ideas for a topic	Given grade-level guides for a topic, district curriculum for the topic, state level standards for science, and/or Next Generation Science Standards, select and justify appropriate goals for that topic in this grade level.
Evaluating and selecting science ideas for their relevance to a particular instructional goal	Given a specific instructional goal, determine which science idea is most closely connected to this goal.
Identifying the big idea or instructional goal of an instructional activity	Given a specific activity, identify the underlying big idea or instructional goal.

Note. Adapted from *The PRAXIS® Study Companion. Elementary Education: Content Knowledge for Teaching* (p. 99), by Educational Testing Service, 2017, Princeton, NJ, Author. Copyright 2017 Educational Testing Service.

Role of the National Advisory Committee

For the CKT assessment in science, we recruited and convened a national advisory committee (NAC) to review the test specifications, or blueprints, for inclusion of the most essential student-level content domain and science teaching practices. In order to maximize diversity of perspectives on the NAC, we selected experienced elementary science teachers, elementary science teacher educators, and elementary science education researchers from across various regions of the United States. We also made sure to include those who had expertise across the various science content areas (e.g., life science, physical science, earth and space science, engineering) and across the elementary grades. The NAC comprised 13 members (11 females, two males), which included six elementary science teacher faculty members and seven elementary science teachers.

The NAC members analyzed the two main parts of the CKT science test specifications to confirm the importance of the KSAs being assessed on this new elementary teacher licensure assessment. In addition, they reviewed the blueprints to ensure that the descriptions were clear and to recommend relative weights for each content category. The NAC members focused on determining (a) whether the student-level content domain specified was important for elementary students to master and, therefore, important for a beginning elementary teacher to know on the first day of instruction and (b) whether the science teaching practices specified important instructional practices elementary science teachers need to engage in from their first day on the job to support student mastery of the student-level content domain.

In reviewing the student-level content domain, the NAC decided to accept all 77 performance expectations from the NGSS across the K−5 grade-level span. There were only some very minor changes to some of the language suggested—mainly for additional clarification—in this part of the test specifications. The main focus in the NAC's feedback about the student-level content domain targeted the disciplinary core ideas descriptions for each of the 12 content areas (see the written description in Figure 1 for PS3: Energy). The NAC members requested that these descriptions, which were generated by the Educational Testing Service (ETS) development team, be written at a finer grain size and include more details about the specific content covered within each area. In terms of the science teaching practices, the NAC members ranked the majority of these science-specific instructional practices as "somewhat important" or "important" for beginning elementary teachers to effectively facilitate student mastery of the student domain. Overall, the NAC's decisions resulted in the addition, removal, and merging of a handful of science teaching practices. For example, the NAC decided that beginning elementary science teachers should be able to critique student-generated explanations and be able to select scientific explanations that are accurate and accessible to students; they decided to keep both of these practices. However, the NAC members did not think that it was essential for beginning elementary science teachers to be able to construct their own scientific explanations from the first day on the job, and this science teaching practice was removed from the test specifications. Two categories were merged together (the scientific models and representations categories) and other practices were added to the list resulting in a final list of 27 science teaching practices (see Appendix B for a list of these practices) organized into seven different categories.

Method

Sample Recruitment and Selection Process

For this study, we aimed to sample elementary school science teachers and college faculty who prepare elementary school science teachers, and screened interested participants according to an eligibility survey. To be eligible to participate in this study, respondents had to be licensed elementary school science teachers or college faculty currently preparing elementary school teacher candidates to teach science. In the first wave of recruitment, we contacted 785 potential participants, including both faculty members at teacher preparation programs and elementary school science teachers who had previously participated in studies at ETS related to teaching. However, as we were unable to reach a target sample of at least 200 respondents that is necessary for achieving adequate levels of generalizability to the larger population of elementary science educators (Kane et al., 1995), we opted to recruit additional respondents via electronic newsletters sent to members of professional associations for elementary school science teachers and faculty members who prepare elementary school science teacher candidates (e.g., the Association for Science Teacher Education, the National Science Teachers' Association). In this second wave of recruitment, a total of 7,246 association members were invited to participate in the study via a newsletter.

Sample

We aimed to sample approximately twice as many teachers as faculty members to obtain a significant number of teachers currently teaching lower (kindergarten through Grade 3) and upper (Grades 4–6) elementary classes. Both public- and private-school teachers were included in the sample. The final sample included 203 participants (146 teachers and 57 faculty members).

In the first wave of recruitment in which 785 participants were contacted, 169 (21.5%) participated in the study. Ultimately, 34 respondents of the 7,246 association members contacted in the second wave of recruitment (< 1%) participated. After combining the two groups, the overall sample size was 203, reflecting a 2.5% response rate for the survey. While the overall response rate for the survey was low, as shown in Table 2, the resulting sample mostly reflected the composition of the national population of science teachers when compared to the National Center for Education Statistics (NCES) 2011–2012 School and Staffing Survey (SASS; Goldring, Gray, & Bitterman, 2013), with the exception of some undersampling of Hispanic educators (3% in the sample compared to 6% nationally). The sample of educators adequately represented the percentage of Asian science educators (3% in the sample compared to 3% nationally) and slightly overrepresented the percentage of Black science educators (7% in the sample compared to 6% nationally) compared to the latest SASS results. The sample approximately mirrored elementary school teachers nationally in terms of years of experience, with approximately 33% of teachers with 10 or fewer years of experience and the remaining two thirds with 11 or more years of experience, and gender, with approximately 85% of elementary school teachers being female (Goldring et al., 2013).

Table 2. Summary of Frequencies for Background Information Survey for Teachers and Faculty

Item	Teachers (*N* = 146)	Faculty (*N* = 57)
Gender		
Female	129 (88%)	44 (77%)
Male	16 (11%)	12 (21%)
Missing/prefer not to answer	1 (1%)	1 (2%)
Race/ethnicity		
American Indian or Alaska Native	1 (1%)	1 (2%)
Asian or Asian American	6 (4%)	1 (2%)
Black or African American	12 (8%)	2 (4%)
Hispanic/Latino	6 (4%)	0 (0%)
Native Hawaiian or Other Pacific Islander	0 (0%)	0 (0%)
White	114 (78%)	49 (86%)
Two or more races	3 (2%)	1 (2%)
Other/prefer not to answer/missing	4 (3%)	3 (5%)
Geographic region		
Northeast	34 (23%)	12 (22%)
Midwest	30 (20%)	17 (31%)
South	55 (38%)	17 (31%)
West	27 (19%)	9 (16%)
Current teaching assignment		
Lower (Grades K–3)	38 (26%)	-
Upper (Grades 4–6)	76 (52%)	-
Other	32 (22%)	-
Years of teaching experience		
0–2 Years	5 (3%)	-
3–5 Years	13 (9%)	-
6–10 Years	31 (21%)	-
11–20 Years	72 (49%)	-
21 years or more	25 (17%)	-
Mentored or supervised student teachers		
Yes	-	46 (81%)
No	-	11 (19%)
School or institution location		
Urban	50 (34%)	-
Suburban	67 (46%)	-
Rural	29 (20%)	-
Minority-serving institution		
Yes	-	5 (9%)
No	-	50 (88%)
Designation not available	-	2 (4%)

Content Validity Survey

We collected relevance and importance judgments for each of 12 science content knowledge topics and 27 science teaching practices. Two content-related validity questions were posed to participants for each science content topic and teaching practice:

- Is knowing how to teach this area of content, or the practice, relevant to a beginning elementary school science teacher's ability to be a safe and effective educator?
- If knowing how to teach this area of content, or the practice, is relevant, how important is it to a beginning elementary school science teacher's ability to be a safe and effective educator?

For the 27 science teaching practices, a third question was posed:

- How frequently is this practice applied by beginning elementary school teachers when teaching science?

If participants judged a science content topic or science teaching practice as relevant (indicated by Yes), they then rated the importance of the topic or practice using a 6-point Likert scale, with anchors that ranged from 1 (*not at all important*) to 6 (*extremely important*) Thus, importance ratings were only collected from respondents who judged the science content topic or science teaching practice as relevant. For the science teaching practices only, if participants judged a science teaching practice as relevant, they then rated the frequency with which the practice is applied by beginning elementary school teachers when teaching science, after making judgments regarding the practice's importance using a 6-point Likert scale, with anchors that ranged from 1 (*never*) to 6 (*very frequently*). Participants were paid $50 via a virtual gift card in exchange for their participation in the study. The survey, conducted online, took approximately one hour to complete.

Analysis

We report two types of analyses. The first type represents patterns in average relevance, importance, and frequency judgment ratings across the overall sample and by subgroups among teachers following the rationale provided by Tannenbaum and Rosenfeld (1994). The content knowledge topic areas and science teaching practices should be deemed relevant, important, and frequently applied by the participants surveyed to support their inclusion on the licensure test (Tannenbaum & Rosenfeld, 1994). Comparisons were made for different groups (e.g., educator role [teacher, faculty], grade level bands [K−3, 4−6], race/ethnicity [White, Non-White], geographic region [Northeast, Midwest, South, West]) using effect sizes in which the mean difference between two groups was divided by a combination of group sample sizes and standard deviations. This was done to highlight notable areas in which differences between subgroups were observed. In the case of race/ethnicity, White teachers were used as the reference group; for geographic region, Northeastern teachers were used as the reference group (see Martin-Raugh et al., 2016). Effect sizes are often categorized by magnitude in terms of absolute value (Cohen, 1988): negligible (< 0.20), low (0.20 − < 0.50), medium (0.50 − < 0.80), and high (≥ 0.80). While non-negligible effect sizes are categorized as low, following Fan (2001), for the purposes of this paper, we therefore report effect sizes that are at least medium in size to highlight notable differences in judgments across different groups. For clarification purposes, when discussing teachers and faculty together, we refer to this as the overall sample. When discussing relevant subgroups of teachers together, we refer to this as the combined teacher sample. The second type of analysis indexes agreement for relevance, importance, and frequency judgments using intraclass correlations (ICC[2]; Shrout & Fleiss, 1979).

Results

Relevance Judgments

Across the 12 content knowledge (CK) topic areas and the 27 science teaching practices (STPs), more than 75% of the overall sample agreed that most of these are all relevant for effective practice for elementary school teachers teaching science. We computed intraclass correlations (ICCs; Shrout & Fleiss, 1979) to estimate interrater reliability across the overall sample's judgments. These estimates indicate the consistency in the ratings across different individuals, where greater consistency provides greater confidence in the accuracy of the judgments (Landers, 2015). The ICC(2) (Shrout & Fleiss, 1979) indicating agreement among the overall sample regarding their relevance ratings across the 12 topic areas and 27 teaching practices were .85 (95% CI [.81, .88]) and .85 (95% CI [.82, .88]), respectively, suggesting agreement was excellent (Cicchetti, 1994).

All CK areas were rated as relevant by more than 75% of the faculty, while 10 of the 12 CK areas were rated as relevant by more than 75% of the teachers. The two CK areas that were judged as relevant by less than 75% of teachers and had among the largest differences compared to faculty were: CK 6 (Heredity: Inheritance and Variation of Traits; judged relevant by 74% of teachers and 95% of faculty) and CK 11 (Waves and Their Application in Technologies for Information Transfer; judged relevant by 71% of teachers and 86% of faculty). Additionally, the relevance judgments for teachers and faculty differed by more than 10% for CK 7 (Biological Evolution: Unity and Diversity; judged relevant by 79% of teachers and 96% of faculty). Overall, relevance judgments support the content validity of the CK areas and STPs included in the assessment.

Importance Judgments

Tables 3 and 4 summarize average educators' importance judgments with respect to each CK areas and STPs. Importance judgments ranged from 1 (*not at all important*) to 6 (*extremely important*). Results are provided for teachers, faculty, and the overall sample. It is important to note that respondents only answered these questions if they found the practices to be relevant; therefore sample sizes for individual practices may vary. Agreement across the individuals sampled regarding importance judgments was excellent (Cicchetti, 1994). The ICC(2) indicating agreement among the overall sample regarding their importance ratings across the 12 CK topic areas and 27 practices were .88 (95% CI [.85, .90]) and .91 (95% CI [.90, .93]), respectively.

Table 3. Summary of Average Importance Judgments for Content Knowledge (CK) Areas for Teachers, Faculty, and Overall

Item	Teachers	Faculty	Overall
CK 1	4.92 (0.95)	4.93 (0.86)	4.92 (0.92)
CK 2	4.85 (0.92)	4.77 (0.96)	4.83 (0.93)
CK 3	4.98 (0.90)	4.93 (0.97)	4.97 (0.92)
CK 4	4.98 (0.80)	5.09 (0.84)	5.01 (0.81)
CK 5	4.86 (0.90)	4.96 (0.82)	4.89 (0.88)
CK 6	4.50 (0.98)	4.67 (0.95)	4.56 (0.97)
CK 7	4.60 (0.98)	4.82 (1.02)	4.67 (0.99)
CK 8	4.82 (0.99)	4.90 (1.01)	4.84 (1.00)
CK 9	4.81 (0.87)	4.82 (1.02)	4.81 (0.92)
CK 10	4.67 (1.07)	4.83 (0.91)	4.72 (1.03)
CK 11	4.28 (1.17)	4.14 (1.08)	4.24 (1.14)
CK 12	5.22 (1.04)	5.32 (0.85)	5.25 (0.98)
Minimum	4.28	4.14	4.24
Maximum	5.22	5.32	5.25
Sample size	103−143	49−57	152-200

Note. Importance scale: 1 (*not at all important*), 2 (*of little importance*), 3 (*of some importance*), 4 (*moderately important*), 5 (*very important*), 6 (*extremely important*).

Table 8. Summary of Average Importance Judgments for Science Teaching Practices (STPs) by Race/Ethnicity and Combined

Item	White	Non-White	Combined
STP 1	5.44 (0.74)	5.18 (0.90)	5.39 (0.78)
STP 2	5.41 (0.76)	5.29 (0.66)	5.39 (0.74)
STP 3	5.17 (0.83)	4.82 (0.86)	5.10 (0.85)
STP 4	5.12 (0.78)	5.08 (0.64)	5.11 (0.75)
STP 5	5.19 (0.83)	5.08 (0.86)	5.17 (0.84)
STP 6	4.72 (0.93)	4.70 (0.82)	4.72 (0.90)
STP 7	5.01 (0.81)	4.68 (0.86)	4.94 (0.83)
STP 8	4.80 (0.90)	4.64 (0.81)	4.77 (0.89)
STP 9	4.66 (0.90)	4.92 (0.95)	4.71 (0.92)
STP 10	5.12 (0.85)	5.22 (0.58)	5.14 (0.80)
STP 11	5.13 (0.96)	5.11 (0.70)	5.12 (0.91)
STP 12	5.23 (0.84)	5.04 (0.74)	5.19 (0.82)
STP 13	5.38 (0.80)	5.37 (0.56)	5.38 (0.75)
STP 14	4.86 (0.99)	4.84 (0.75)	4.86 (0.94)
STP 15	4.86 (0.97)	5.12 (0.82)	4.91 (0.94)
STP 16	4.56 (0.94)	4.76 (0.89)	4.59 (0.93)
STP 17	5.29 (0.87)	5.19 (0.74)	5.27 (0.84)
STP 18	5.23 (0.83)	5.19 (0.83)	5.22 (0.83)
STP 19	4.87 (0.86)	4.92 (0.88)	4.88 (0.86)
STP 20	4.93 (1.04)	5.12 (0.88)	4.96 (1.01)
STP 21	4.68 (0.93)	4.65 (0.78)	4.68 (0.90)
STP 22	4.94 (0.90)	5.04 (0.81)	4.96 (0.88)
STP 23	4.87 (0.85)	4.69 (0.97)	4.83 (0.87)
STP 24	4.99 (0.90)	4.92 (0.74)	4.98 (0.87)
STP 25	5.05 (0.79)	4.85 (0.78)	5.01 (0.79)
STP 26	4.86 (0.86)	4.72 (0.79)	4.83 (0.85)
STP 27	4.78 (0.85)	4.75 (0.85)	4.77 (0.85)
Minimum	4.56	4.64	4.59
Maximum	5.44	5.37	5.39
Sample size	99–112	21–28	120–140

Note. Importance scale: 1 (*not at all important*), 2 (*of little importance*), 3 (*of some importance*), 4 (*moderately important*), 5 (*very important*), 6 (*extremely important*).

Importance judgments disaggregated by teacher region (Northeast, Midwest, South, and West) are depicted in Tables 9 and 10. No matter the teacher region, on average, all CK areas and STPs were judged to be at least moderately important across groups, supporting their inclusion in the teacher licensure examination for elementary school science teachers.

Using Northeastern teachers as a reference group, there were no effect sizes above 0.50 in CK among areas with other regions, and the only effect sizes above 0.50 in STPs among areas in average judgments were with STP 12 (Choosing resources that support the selection of accurate, valid, and age-appropriate goals for science learning) by Southern teachers (diff. = 0.48; ES = 0.57); and STP 19 (Supporting and critiquing students' participation in and use of verbal and written scientific discourse and argumentation) by Western teachers (diff. = 0.39; ES = 0.50).

Table 9. Summary of Average Importance Judgments for Content Knowledge (CK) Areas by Geographic Region and Combined

Item	Northeast	Midwest	South	West	Combined
CK 1	4.80 (0.92)	4.85 (1.16)	5.02 (0.87)	4.91 (0.92)	4.92 (0.95)
CK 2	4.74 (1.12)	4.60 (1.00)	4.96 (0.81)	5.00 (0.72)	4.85 (0.92)
CK 3	4.91 (0.86)	4.85 (0.92)	5.04 (0.92)	5.13 (0.92)	4.98 (0.90)
CK 4	4.65 (0.84)	5.12 (0.88)	5.08 (0.72)	5.04 (0.77)	4.98 (0.80)
CK 5	4.81 (0.91)	4.96 (0.81)	4.83 (0.98)	4.88 (0.83)	4.86 (0.90)
CK 6	4.38 (1.01)	4.41 (0.85)	4.59 (1.07)	4.56 (0.89)	4.50 (0.98)
CK 7	4.41 (1.02)	4.74 (0.92)	4.73 (1.07)	4.42 (0.69)	4.60 (0.98)
CK 8	4.62 (1.18)	4.70 (1.06)	5.02 (0.82)	4.78 (0.95)	4.82 (0.99)
CK 9	4.69 (0.93)	4.78 (0.93)	4.98 (0.80)	4.63 (0.88)	4.81 (0.87)
CK 10	4.63 (1.18)	4.72 (0.94)	4.79 (1.06)	4.41 (1.10)	4.67 (1.07)
CK 11	4.36 (1.22)	4.50 (1.04)	4.18 (1.21)	4.19 (1.17)	4.28 (1.17)
CK 12	5.03 (1.09)	5.24 (0.95)	5.27 (1.13)	5.32 (0.85)	5.22 (1.04)
Minimum	4.36	4.41	4.18	4.19	4.28
Maximum	5.03	5.24	5.27	5.32	5.22
Sample size	24–34	18–29	44–55	16–25	103–143

Note. Importance scale: 1 (*not at all important*), 2 (*of little importance*), 3 (*of some importance*), 4 (*moderately important*), 5 (*very important*), 6 (*extremely important*).

Table 10. Summary of Average Importance Judgments for Science Teaching Practices (STPs) by Geographic Region and Combined

Item	Northeast	Midwest	South	West	Combined
STP 1	5.26 (0.83)	5.45 (0.63)	5.49 (0.77)	5.23 (0.91)	5.38 (0.78)
STP 2	5.41 (0.56)	5.52 (0.57)	5.34 (0.94)	5.26 (0.76)	5.38 (0.76)
STP 3	4.94 (0.89)	4.86 (0.83)	5.30 (0.85)	5.19 (0.69)	5.11 (0.84)
STP 4	4.94 (0.94)	5.23 (0.57)	5.19 (0.81)	5.00 (0.69)	5.11 (0.75)
STP 5	5.15 (0.91)	5.17 (0.83)	5.21 (0.84)	5.18 (0.73)	5.18 (0.83)
STP 6	4.75 (0.95)	4.82 (1.02)	4.69 (0.94)	4.61 (0.58)	4.72 (0.90)
STP 7	4.88 (0.81)	4.86 (0.83)	5.04 (0.91)	4.96 (0.65)	4.95 (0.82)
STP 8	4.81 (0.83)	4.79 (0.82)	4.78 (1.05)	4.74 (0.62)	4.78 (0.88)
STP 9	4.72 (0.85)	4.64 (0.91)	4.78 (0.97)	4.70 (0.95)	4.72 (0.92)
STP 10	5.21 (0.65)	5.13 (0.73)	5.02 (0.96)	5.15 (0.88)	5.11 (0.83)
STP 11	4.97 (1.13)	5.10 (0.80)	5.24 (0.85)	5.15 (0.83)	5.13 (0.91)
STP 12	4.91 (0.93)	5.23 (0.63)	5.40 (0.79)	5.15 (0.83)	5.20 (0.82)
STP 13	5.31 (0.59)	5.39 (0.88)	5.47 (0.76)	5.23 (0.76)	5.37 (0.75)
STP 14	4.90 (0.72)	4.54 (1.07)	4.90 (1.02)	5.13 (0.76)	4.86 (0.94)
STP 15	4.90 (0.71)	4.59 (1.02)	5.02 (1.01)	5.13 (0.85)	4.92 (0.93)
STP 16	4.68 (0.86)	4.48 (0.96)	4.57 (1.06)	4.75 (0.64)	4.61 (0.93)
STP 17	5.18 (0.81)	5.20 (1.03)	5.36 (0.88)	5.26 (0.53)	5.27 (0.84)
STP 18	5.00 (0.92)	5.20 (0.81)	5.33 (0.86)	5.26 (0.66)	5.21 (0.83)
STP 19	4.83 (0.89)	4.71 (0.90)	4.85 (0.90)	5.22 (0.60)	4.88 (0.86)
STP 20	5.03 (0.90)	4.70 (1.09)	5.06 (1.11)	5.00 (0.76)	4.96 (1.00)
STP 21	4.72 (0.84)	4.62 (0.94)	4.65 (0.99)	4.78 (0.74)	4.68 (0.90)
STP 22	5.03 (0.87)	4.90 (0.86)	4.94 (0.96)	5.04 (0.75)	4.97 (0.88)
STP 23	4.73 (0.91)	4.93 (0.86)	4.82 (0.93)	4.92 (0.70)	4.84 (0.87)
STP 24	4.88 (0.91)	5.14 (0.76)	4.94 (0.91)	5.08 (0.84)	4.99 (0.86)
STP 25	4.91 (0.82)	5.00 (0.64)	4.96 (0.85)	5.23 (0.76)	5.01 (0.79)
STP 26	4.85 (0.71)	4.67 (0.80)	4.92 (0.94)	4.84 (0.94)	4.83 (0.86)
STP 27	4.72 (1.00)	4.79 (0.73)	4.75 (0.89)	4.90 (0.70)	4.78 (0.85)
Minimum	4.68	4.48	4.57	4.61	4.61
Maximum	5.41	5.52	5.49	5.26	5.38
Sample size	28–34	25–30	48–55	20–27	122–144

Note. Importance scale: 1 (*not at all important*), 2 (*of little importance*), 3 (*of some importance*), 4 (*moderately important*), 5 (*very important*), 6 (*extremely important*).

Frequency Judgments

Participants were asked how frequently STPs were applied by beginning elementary school teachers when teaching science. As with judgments of importance, participants were only asked to provide judgments of frequency for practices deemed as being relevant to a beginning teacher's ability to be a safe and effective educator. Ratings were applied using a 6-point scale from 1 (*never*) to 6 (*very frequently*). Table 11 summarizes average frequency judgments for teachers, faculty, and for the overall sample. These ranged from 4.22 to 5.16 for teachers, 3.57 to 5.02 for faculty, and 4.04 to 5.12 for the overall sample. The ICC(2) indexing agreement among the overall sample regarding their frequency ratings for the 27 practices was .93 (95% CI [.92, .94]), indicating excellent agreement (Cicchetti, 1994).

Table 11. Summary of Average Frequency Judgments for Science Teaching Practices (STPs) for Teachers, Faculty, and Overall

Item	Teachers	Faculty	Overall
STP 1	4.97 (0.99)	4.88 (0.97)	4.95 (0.98)
STP 2	5.16 (0.82)	5.02 (0.90)	5.12 (0.84)
STP 3	4.70 (0.87)	4.40 (1.08)	4.62 (0.95)
STP 4	4.87 (0.85)	4.91 (0.87)	4.88 (0.85)
STP 5	4.76 (0.94)	4.34 (1.12)	4.64 (1.01)
STP 6	4.32 (0.96)	3.81 (1.08)	4.17 (1.02)
STP 7	4.56 (1.00)	4.21 (1.10)	4.46 (1.04)
STP 8	4.24 (1.03)	3.65 (1.15)	4.07 (1.10)
STP 9	4.41 (0.96)	3.70 (0.92)	4.21 (1.00)
STP 10	4.59 (0.92)	4.34 (1.07)	4.52 (0.97)
STP 11	4.65 (1.07)	4.31 (1.23)	4.56 (1.12)
STP 12	4.86 (0.96)	4.55 (1.03)	4.77 (0.99)
STP 13	4.77 (1.07)	4.34 (1.16)	4.64 (1.11)
STP 14	4.36 (1.04)	3.80 (1.18)	4.21 (1.11)
STP 15	4.38 (0.98)	3.98 (1.17)	4.27 (1.05)
STP 16	4.33 (1.06)	3.65 (1.20)	4.13 (1.14)
STP 17	5.06 (0.90)	4.51 (1.05)	4.90 (0.97)
STP 18	5.06 (0.90)	4.52 (0.91)	4.91 (0.93)
STP 19	4.37 (1.02)	3.96 (1.26)	4.25 (1.11)
STP 20	4.47 (1.11)	4.36 (1.11)	4.44 (1.11)
STP 21	4.32 (0.98)	3.88 (1.10)	4.19 (1.03)
STP 22	4.58 (1.04)	4.25 (1.03)	4.48 (1.05)
STP 23	4.58 (0.93)	4.13 (1.10)	4.45 (1.00)
STP 24	4.57 (1.06)	3.84 (1.21)	4.36 (1.15)
STP 25	4.52 (0.90)	4.07 (1.13)	4.39 (0.99)
STP 26	4.43 (0.96)	3.73 (1.27)	4.24 (1.10)
STP 27	4.22 (0.98)	3.57 (1.19)	4.04 (1.08)
Minimum	4.22	3.57	4.04
Maximum	5.16	5.02	5.12
Sample size	122−144	49−57	171−201

Note. Frequency scale: 1 (*never*), 2 (*very rarely*), 3 (*rarely*), 4 (*occasionally*), 5 (*frequently*), 6 (*very frequently*).

Overall, all STPs were judged to be at least occasionally performed. Below, we describe observed differences in frequency judgments across roles to provide a more nuanced analysis of how frequency judgments may differ. All STPs, with the exception of STP 4 (Choosing which science ideas or instructional activities are most closely related to a particular instructional goal), were judged by teachers as more frequently applied when compared to faculty. In examining the absolute difference in judgments, nine STPs exceeded 0.50, including: STP 8 (Critiquing scientific procedures, data, observations, or results for their quality, accuracy, or appropriateness), STP 9 (Evaluating and selecting media for engaging students in virtual investigations not possible in firsthand situations), STP 14 (Selecting diagnostic items and eliciting student thinking about scientific ideas and practices to identify common student misconceptions and the basis for those misconceptions), STP 16 (Identifying the connections between students' talk and work, and scientists' talk and work), STP 17 (Selecting scientific language that is precise, accurate, and grade-appropriate and illustrates key scientific concepts), STP 18 (Anticipating scientific language and vocabulary that may be difficult for students), STP 24 (Engaging students in using, modifying, creating, and critiquing scientific models and representations that are matched to instructional goals), STP 26 (Generating or selecting diagnostic questions to evaluate student understanding of specific models and representations, and STP 27 (Evaluating student ideas about what makes for good scientific models and representations).

Table 12 presents average frequency judgments comparing lower elementary school teachers (kindergarten to Grade 3) and upper elementary school teachers (Grades 4–6) across the STPs. It is worth noting that participants were asked to consider the frequency with which practices were applied across the elementary school grade span. As can be seen, all but five of the STPs were judged to be applied more frequently by lower elementary school teachers than by upper elementary school teachers. The average frequency judgments fell within similar ranges for lower elementary school teachers (4.22 to 5.19) and for upper elementary school teachers (4.18 to 5.16). Effect sizes were below 0.50 for all STPs, suggesting that differences in teacher judgments across different grade levels were small, which further supports the inclusion of the STPs on the licensure exam.

Average frequency judgments disaggregated by teacher race/ethnicity and combined are presented in Table 13. As mentioned earlier, due to the small subgroup sample sizes, analyses were conducted using White and non-White categorizations. Results should be interpreted with caution due to the large difference in sample sizes (N = 99–112 for White teachers and N = 21–28 for non-White teachers). Regardless of educator race/ethnicity, on average, all STPs were judged to be occasionally or frequently applied across groups, supporting their inclusion in the teacher licensure examination for elementary school science teachers. We highlight observed differences in frequency judgments across educator races/ethnicities below to provide a more in-depth analysis of how frequency of practice application may vary across groups.

Average frequency judgments ranged from 4.20 to 5.17 for White teachers, 4.00 to 5.21 for non-White teachers thus representing slightly greater variation, and 4.21 to 5.18 for the combined sample. Across STPs, a roughly equal proportion were rated to be more likely applied by each group with STP 16 (Identifying the connections between students' talk and work and scientists' talk and work) and STP 22 (Selecting explanations of scientific phenomena that are accurate and accessible to students) having equal average frequency judgments. Effect sizes were below 0.50 for all STPs.

Table 12. Summary of Average Frequency Judgments for Science Teaching Practices (STPs) by Current Grade Level Taught

Item	Lower (K–3)	Upper (4–6)	Difference
STP 1	5.11 (0.86)	4.89 (1.07)	0.21 (0.21)
STP 2	5.19 (0.78)	5.16 (0.83)	0.03 (0.03)
STP 3	4.81 (0.92)	4.64 (0.85)	0.17 (0.19)
STP 4	4.89 (0.76)	4.80 (0.91)	0.09 (0.10)
STP 5	4.74 (0.82)	4.66 (0.96)	0.08 (0.09)
STP 6	4.41 (0.92)	4.40 (0.92)	0.01 (0.01)
STP 7	4.82 (0.93)	4.45 (1.04)	0.36 (0.36)
STP 8	4.43 (0.92)	4.19 (1.05)	0.24 (0.24)
STP 9	4.43 (1.12)	4.47 (0.86)	0.04 (0.04)
STP 10	4.76 (0.76)	4.49 (1.00)	0.27 (0.29)
STP 11	4.76 (1.04)	4.54 (1.05)	0.22 (0.21)
STP 12	5.11 (0.91)	4.76 (1.00)	0.35 (0.36)
STP 13	4.79 (1.07)	4.71 (1.01)	0.08 (0.08)
STP 14	4.39 (1.00)	4.30 (1.03)	0.09 (0.09)
STP 15	4.55 (0.89)	4.28 (0.97)	0.27 (0.28)
STP 16	4.69 (1.11)	4.18 (1.04)	0.51 (0.48)
STP 17	5.00 (1.00)	5.05 (0.89)	0.05 (0.06)
STP 18	5.17 (0.94)	5.03 (0.86)	0.14 (0.16)
STP 19	4.32 (1.17)	4.43 (0.92)	0.10 (0.10)
STP 20	4.53 (1.30)	4.48 (1.03)	0.05 (0.04)
STP 21	4.22 (1.04)	4.31 (0.93)	0.10 (0.10)
STP 22	4.60 (1.12)	4.57 (1.00)	0.03 (0.03)
STP 23	4.72 (1.11)	4.47 (0.85)	0.25 (0.26)
STP 24	4.64 (1.07)	4.56 (1.01)	0.08 (0.07)
STP 25	4.43 (0.95)	4.58 (0.88)	0.15 (0.17)
STP 26	4.45 (0.97)	4.41 (0.94)	0.04 (0.05)
STP 27	4.53 (1.16)	4.18 (0.74)	0.35 (0.39)
Minimum	4.22	4.18	0.01
Maximum	5.19	5.16	0.51
Sample size	29–38	66–75	

Note. Frequency scale: 1 (*never*), 2 (*very rarely*), 3 (*rarely*), 4 (*occasionally*), 5 (*frequently*), 6 (*very frequently*). Consistent with prior reporting in RM-16-08, RM-16-11, and RM-16-10, difference values are presented in the third column.

Table 13. Summary of Average Frequency Judgments for Science Teaching Practices (STPs) by Race/Ethnicity and Combined

Item	White	Non-White	Combined
STP 1	4.98 (1.02)	5.04 (0.88)	4.99 (0.99)
STP 2	5.17 (0.83)	5.21 (0.79)	5.18 (0.82)
STP 3	4.73 (0.84)	4.64 (1.03)	4.71 (0.88)
STP 4	4.86 (0.87)	4.96 (0.79)	4.88 (0.85)
STP 5	4.76 (0.97)	4.80 (0.76)	4.77 (0.93)
STP 6	4.29 (0.97)	4.48 (0.98)	4.33 (0.97)
STP 7	4.55 (1.03)	4.57 (0.96)	4.56 (1.01)
STP 8	4.20 (1.08)	4.32 (0.90)	4.22 (1.04)
STP 9	4.39 (0.92)	4.44 (1.19)	4.40 (0.97)
STP 10	4.58 (0.96)	4.74 (0.71)	4.61 (0.92)
STP 11	4.64 (1.10)	4.67 (1.00)	4.65 (1.08)
STP 12	4.90 (0.96)	4.71 (1.01)	4.86 (0.97)
STP 13	4.73 (1.14)	4.89 (0.80)	4.76 (1.08)
STP 14	4.32 (1.10)	4.48 (0.82)	4.35 (1.05)
STP 15	4.33 (1.03)	4.54 (0.81)	4.38 (0.99)
STP 16	4.33 (1.05)	4.33 (1.15)	4.33 (1.06)
STP 17	5.11 (0.92)	4.85 (0.86)	5.06 (0.92)
STP 18	5.12 (0.89)	4.93 (0.96)	5.08 (0.90)
STP 19	4.42 (1.01)	4.13 (1.08)	4.36 (1.03)
STP 20	4.48 (1.13)	4.40 (1.12)	4.47 (1.13)
STP 21	4.37 (0.99)	4.00 (0.95)	4.30 (0.99)
STP 22	4.58 (1.09)	4.58 (0.88)	4.58 (1.05)
STP 23	4.64 (0.95)	4.38 (0.85)	4.59 (0.94)
STP 24	4.60 (1.11)	4.35 (0.85)	4.55 (1.06)
STP 25	4.57 (0.91)	4.35 (0.89)	4.53 (0.91)
STP 26	4.45 (1.01)	4.32 (0.80)	4.42 (0.97)
STP 27	4.23 (0.97)	4.13 (1.08)	4.21 (0.99)
Minimum	4.20	4.00	4.21
Maximum	5.17	5.21	5.18
Sample size	99–112	21–28	120–140

Note. Frequency scale: 1 (*never*), 2 (*very rarely*), 3 (*rarely*), 4 (*occasionally*), 5 (*frequently*), 6 (*very frequently*).

Table 14 reports average frequency ratings by teachers' geographic region (Northeast, Midwest, South, West) and combined. Using Northeast teachers as a reference group (see Martin-Raugh et al., 2016), the greatest difference in average frequency judgments was found for STP 12 (Choosing resources that support the selection of accurate, valid, and age-appropriate goals for science learning) favoring Midwest teachers (diff. = 0.63; ES = 0.64) as well as teachers in the South (diff. = 0.50; ES = 0.51).

Table 14. Summary of Average Frequency Judgments for Science Teaching Practices (STPs) by Geographic Region

Item	Northeast	Midwest	South	West	Combined
STP 1	4.82 (1.17)	5.31 (0.76)	4.98 (0.93)	4.77 (1.03)	4.97 (0.99)
STP 2	5.18 (0.72)	5.45 (0.63)	5.13 (0.90)	4.89 (0.89)	5.16 (0.82)
STP 3	4.56 (0.86)	4.83 (0.85)	4.79 (0.88)	4.58 (0.90)	4.70 (0.87)
STP 4	4.59 (0.95)	5.13 (0.68)	4.98 (0.84)	4.65 (0.80)	4.87 (0.85)
STP 5	4.58 (0.97)	4.93 (0.69)	4.85 (1.05)	4.59 (0.91)	4.76 (0.94)
STP 6	4.34 (1.12)	4.61 (0.83)	4.18 (0.97)	4.22 (0.80)	4.32 (0.96)
STP 7	4.53 (1.13)	4.59 (0.73)	4.63 (1.07)	4.41 (0.97)	4.56 (1.00)
STP 8	4.16 (1.16)	4.34 (0.90)	4.20 (1.18)	4.30 (0.63)	4.24 (1.03)
STP 9	4.34 (0.97)	4.50 (0.88)	4.47 (1.01)	4.30 (0.99)	4.41 (0.96)
STP 10	4.76 (1.00)	4.57 (0.73)	4.57 (1.00)	4.46 (0.86)	4.59 (0.92)
STP 11	4.45 (1.12)	4.63 (0.93)	4.83 (1.09)	4.54 (1.10)	4.65 (1.07)
STP 12	4.50 (1.08)	5.13 (0.86)	5.00 (0.92)	4.73 (0.87)	4.86 (0.96)
STP 13	4.56 (1.11)	4.79 (0.99)	5.00 (1.04)	4.54 (1.10)	4.77 (1.07)
STP 14	4.24 (1.12)	4.00 (0.98)	4.56 (1.01)	4.52 (0.99)	4.36 (1.04)
STP 15	4.23 (1.01)	4.28 (0.88)	4.43 (1.08)	4.58 (0.88)	4.38 (0.98)
STP 16	4.18 (1.19)	4.48 (0.77)	4.31 (1.12)	4.40 (1.10)	4.33 (1.06)
STP 17	4.94 (1.00)	5.00 (0.95)	5.26 (0.84)	4.85 (0.82)	5.06 (0.90)
STP 18	4.94 (0.95)	5.10 (0.76)	5.04 (1.03)	5.22 (0.70)	5.06 (0.90)
STP 19	4.55 (1.12)	4.29 (0.85)	4.19 (1.07)	4.61 (0.94)	4.37 (1.02)
STP 20	4.59 (1.01)	4.33 (1.06)	4.40 (1.27)	4.64 (0.95)	4.47 (1.11)
STP 21	4.38 (1.01)	4.41 (0.73)	4.12 (1.11)	4.52 (0.90)	4.32 (0.98)
STP 22	4.48 (1.21)	4.55 (0.95)	4.63 (1.12)	4.63 (0.77)	4.58 (1.04)
STP 23	4.39 (0.86)	4.68 (0.86)	4.67 (1.09)	4.56 (0.71)	4.58 (0.93)
STP 24	4.38 (1.01)	4.75 (0.93)	4.51 (1.14)	4.73 (1.08)	4.57 (1.06)
STP 25	4.47 (0.98)	4.63 (0.67)	4.43 (0.91)	4.62 (1.02)	4.52 (0.90)
STP 26	4.27 (1.13)	4.43 (0.73)	4.50 (0.99)	4.52 (0.96)	4.43 (0.96)
STP 27	4.14 (1.09)	4.48 (0.69)	4.08 (1.11)	4.33 (0.80)	4.22 (0.98)
Minimum	4.14	4.00	4.08	4.22	4.22
Maximum	5.18	5.45	5.26	5.22	5.16
Sample size	28−34	25−30	48−55	20−27	122−144

Note. Frequency scale: 1 (*never*), 2 (*very rarely*), 3 (*rarely*), 4 (*occasionally*), 5 (*frequently*), 6 (*very frequently*).

Discussion

The goal of this investigation was to examine evidence regarding the content-related validity of the content-knowledge section of a teacher licensure exam designed for elementary school teachers in the subject of science. Judgments gathered from both practicing elementary school science teachers and college faculty members who prepare science teachers support the inclusion of the 12 topics and 27 science teaching practices evaluated in this study as part of an elementary school teacher licensure examination. The overall sample indicated that on average, all CK areas and STPs are relevant and important. Of the 27 STPs, three STPs were judged as being performed somewhere between frequently and very frequently, and the others fell somewhere between occasionally and frequently. Establishing that all content assessed in a licensure examination is germane to the profession is an important first step in gathering requisite content-related validity evidence (Raymond & Neustel, 2006). In this investigation, more than three quarters of the overall sample judged 10 of the 12 content topics and all 27 practices to be relevant, supporting the inclusion of the CK areas and STPs on the teacher licensure examination.

We also indexed agreement in relevance, importance, and frequency ratings across participants to establish the extent to which there was consistency in their judgments. Across relevance, importance, and frequency judgments for both the 12 CK areas and 27 STPs, ICC(2) (Shrout & Fleiss, 1979) values were quite high, ranging from .85 to .93. Therefore, agreement regarding relevance, importance, and frequency judgments can be considered excellent (Cicchetti, 1994).

Kane (1982) argued that the importance ratings collected from expert survey data should be especially emphasized when examining the content validity of a licensure examination, as this information is particularly consequential in that it pertains directly to the protection of students from incompetent teachers. In professional credentialing work, importance ratings are collected more often than information for any other dimension (e.g., relevance, frequency; Newman, Slaughter, & Taranath, 1999). The average importance judgment for the 12 topics was at least 4.24 and for the 27 practices was at least 4.51 on a 6-point scale (with 4 being *moderately important*), indicating that all CK areas and STPs are considered to be at least moderately important for beginning elementary school science teachers. Tannenbaum and Rosenfeld (1994) recommended that an average importance judgment of 3.5 on a 5-point scale is sufficient to determine importance for licensure. Transforming this recommendation to a 6-point scale would result in a threshold of 4.2, suggesting that nearly all topics and practices included in this survey are of sufficient importance for inclusion in a licensure examination. If focusing only on the judgments provided by practitioners (current elementary school teachers), the same conclusion would be reached. Despite the fact that in some instances there are medium-sized differences in mean judgments across different groups of practitioners, analyses across different educator roles, grade levels taught, geographic regions, and races/ethnicities showed that most of the CK areas and STPs included in the survey were judged of sufficient importance for inclusion in a licensure exam. One exception was CK 11 (Waves and Their Application in Technologies for Information Transfer), which received average importance ratings of 4.18 and 4.19 across participants in the South and West, respectively. Mean importance ratings for faculty members, specifically, for CK

11 (Waves and Their Application in Technologies for Information Transfer) and STP 9 (Evaluating and selecting media for engaging students in virtual investigations not possible in firsthand situations) also fell under this threshold, at 4.14 and 4.11, respectively.

Frequency judgments provide useful information for content validation, as professional licensure assessments should place greater emphasis on practices that are performed more often (Raymond & Neustel, 2006). Thus, respondents in this study judged how frequently a practice is applied by beginning elementary school teachers when teaching science. Overall, the average frequency judgments for the 27 science teaching practices across both practicing teachers and faculty members who prepare teachers support the prevalence of these science teaching practices for beginning elementary school science teachers.

These findings add to the growing body of empirical literature identifying the ways in which science teachers use their subject-matter knowledge in a wide range of teaching practices across varied contexts. Most importantly, the results from this content validity survey suggest a set of core teaching practices for beginning elementary science teachers, which builds on and extends earlier work identifying a core set of teaching practices for secondary science teachers (Kloser, 2014). Not only do these results support the validation process for this licensure assessment, but narrowing in on a core set of teaching practices most important for beginning elementary science teachers can also support teacher educators in determining how best to direct their efforts in the limited time they have available during teacher preparation. In addition, this focus on a limited set of 27 science teaching practices can support future efforts to measure elementary science teaching effectiveness and provide areas ripe for targeted, formative feedback to preservice teachers as they learn how to teach science.

The accumulation of content validity evidence is of critical importance in the construction and evaluation of licensure assessments (Kane, 2004; Sireci & Sukin, 2013). In this study, we have collected evidence from teachers and faculty members in the form of relevance, importance, and frequency judgments about science content and practices critical for effective practice for beginning elementary school science teachers. Our results provide support for the inclusion of the science content areas and practices that are focal aspects of the ETS *PRAXIS* Elementary Education: Content Knowledge for Teaching (CKT) assessment in science.

Limitations

This research is not without limitations. One shortcoming of the current study is that our response rate can be considered low. However, our sample largely mirrored the composition of the national population of science teachers when compared to the NCES 2011–2012 SASS (Goldring et al., 2013), with the exception of some undersampling of Hispanic educators. Nonetheless, caution should be exercised in generalizing from the sample of individuals surveyed in this study to the broader population of educators or to subgroups of interest.

It should also be noted that although importance judgments represent a crucial piece of content validity evidence for licensure examinations, importance judgments can be considered multidimensional, and, as is the case with all expert judgments, subjective (Harvey, 1991; Raymond, 2001). For instance, a practice may be considered important either because it is performed with great frequency or because inadequate mastery of the task could result in serious negative consequences. Teachers and faculty members may view the frequency with which practices may be performed differently. For example, although the survey question clearly asks about beginning elementary school teachers more generally, each individual's judgment may be impacted by how frequently he or she performs that practice. This may have been one of the reasons several differences in the perception of the frequency with which STPs are applied were observed between the teachers and faculty members surveyed.

Directions for Future Research

Future research may explore the establishment of direct links between test specifications derived from this work and specific items from the assessment, as individual items should pertain to one or more test specifications (Tannenbaum & Rosenfeld, 1994). The alignment between standards and assessment is critical for high-stakes assessments (Webb, Herman, & Witt, 2007) such as licensure tests. Additional research exploring the extent to which there is alignment between what is assessed on teacher licensure examinations and standards for teaching should be explored. Moreover, further studies examining washback effects (Shohamy, 1992) that licensure examination content can have on what teacher preparation programs choose to focus on in their curriculum are warranted.

Finally, future research could be conducted to better understand the relative importance of each of the science CK areas and STPs for the beginning elementary science teacher. For a licensure assessment, it is important to focus on the most important content that beginning teachers will be likely to teach. While the results suggest that all of the CK areas and STPs were deemed important and relevant, it is unclear whether certain ones should be prioritized differentially in the overall test specifications. Research to more clearly differentiate the science CK areas and STPs in terms of importance would be useful to test developers as they create test items that are relevant to the job domain. Follow-up focus groups with a broader representation of elementary science teachers and teacher educators may also be useful in further distinguishing the relative importance across these various areas.

References

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

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clauser, B. E., Margolis, M. J., & Case, S. M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 701-731). Washington, DC: American Council on Education.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, *60,* 821–836. https://doi.org/10.1177/00131640021970934

Council for the Accreditation of Educator Preparation. (2016). *Introduction: The CAEP Standards*. Retrieved from http://www.caepnet.org/standards/introduction

Council of Chief State School Officers. (2011, April). *Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards: A resource for state dialogue.* Washington, DC: Author.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Educational Testing Service. (2017). *The PRAXIS® study companion. Elementary education: Content knowledge for teaching.* Princeton, NJ: Author.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research*, *94*, 275–282. https://doi.org/10.1080/00220670109598763

Ferguson, R. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 273–317). Washington, DC: Brookings Institution.

Goldhaber, D. (2002). The mystery of good teaching: Surveying the evidence on student achievement and teachers' characteristics. *Education Next, 2,* 50–55.

Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal, 47*, 218–251. https://doi.org/10.3102/0002831209348970

Goldring, R., Gray, L., & Bitterman, A. (2013). *Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2011–12 Schools and Staffing Survey*. *First Look* (NCES Report 2013-314). Washington, DC: National Center for Education Statistics.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 1051–1078). https://doi.org/10.1016/S1574-0692(06)02018-6

Harvey, R. J. (1991). Job analysis. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 71–163). Palo Alto, CA: Consulting Psychologists Press.

Kane, M. T. (1982). A sampling model for validity. *Applied Psychological Measurement, 6*, 125–160. https://doi.org/10.1177/014662168200600201

Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2,* 135–170. https://doi.org/10.1207/s15366359mea0203_1

Kane, M. T., Kingsbury, C., Colton, D., & Estes, C. (1989). Combining data on criticality and frequency in developing plans for licensure and certification examinations. *Journal of Educational Measurement, 26,* 17–27. https://doi.org/10.1111/j.1745-3984.1989.tb00315.x

Kane, M. T., Miller, T., Trine, M., Becker, C., & Carson, K. (1995). The precision of practice analysis results in the professions. *Evaluation and the Health Professions, 18,* 29–50. https://doi.org/10.1177/016327879501800103

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

Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower, 2,* 1–4. https://doi.org/10.15200/winn.143518.81744

Martin-Raugh, M. P., Reese, C. M., Howell, H., Tannenbaum, R. J., Steinberg, J. H., & Xu, J. (2016). *Investigating the relevance and importance of high-leverage practices for beginning elementary school teachers* (Research Memorandum No. RM-16-11). Princeton, NJ: Educational Testing Service.

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

Newman, L. S., Slaughter, R. C., & Taranath, S. N. (1999, April). *The selection and use of rating scales in task surveys: A review of current job analysis practice*. Paper presented at the meeting of the National Council on Measurement in Education, Montreal, Canada.

NGGS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

Raymond, M. R. (1996). Establishing weights for test plans for licensure and certification examinations. *Applied Measurement in Education, 9,* 237–256. https://doi.org/10.1207/s15324818ame0903_3

Raymond, M. R. (2001). Job analysis and the specification of content for licensure and certification examinations. *Applied Measurement in Education, 14,* 369–415. https://doi.org/10.1207/S15324818AME1404_4

Raymond, M. R. (2005). An NCME instructional module on developing and administering practice analysis questionnaires. *Educational Measurement: Issues and Practice, 24*(2), 29–42. https://doi.org/10.1111/j.1745-3992.2005.00009.x

Raymond, M. R., & Luecht, R. M. (2013). Licensure and certification testing. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology: Vol. 3. Testing and assessment in school psychology and education* (pp. 391–414). https://doi.org/10.1037/14049-019

Raymond, M. R., & Neustel, S. (2006). Determining the content of credentialing examinations. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 181−224). Mahwah, NJ: Erlbaum.

Rosenfeld, M., & Tannenbaum, R. J. (1991). *Identification of a core of important enabling skills for the NTE Successor Stage I Examination* (Research Report No. RR-91-37). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1991.tb01404.x

Sanders, W., Wright, P., Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11,* 57–67. https://doi.org/10.1023/A:1007999204543

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, *76*, 513−521. https://doi.org/10.1111/j.1540-4781.1992.tb05402.x

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420‒428. https://doi.org/10.1037/0033-2909.86.2.420

Sireci, S. G., & Sukin, T. (2013). Test validity. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology: Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 61–84). Washington, DC: American Psychological Association. https://doi.org/10.1037/14047-004

Smith, I. L., & Hambleton, R. K. (1990). Content validity studies of licensing examinations. *Educational Measurement: Issues & Practice, 9*(4), 7–10. https://doi.org/10.1111/j.1745-3992.1990.tb00385.x

Tannenbaum, R. J., & Rosenfeld, M. (1994). Job analysis for teacher competency testing: Identification of basic skills important for all entry-level teachers. *Educational and Psychological Measurement, 54*, 199–211. https://doi.org/10.1177/0013164494054001026

Tannenbaum, R. J., & Wesley, S. (1993). Agreement between committee-based and field-based job analyses: A study in the context of licensure testing. *Journal of Applied Psychology, 78*, 975–980. https://doi.org/10.1037/0021-9010.78.6.975

Wang, N., Schnipke, D., & Witt, E. A. (2005). Use of knowledge, skill, and ability statements in developing licensure and certification examinations. *Educational Measurement: Issues and Practice, 24*(1), 15−22. https://doi.org/10.1111/j.1745-3992.2005.00003.x

Webb, N. M., Herman, J. L., & Webb, N. L. (2007). Alignment of mathematics state-level standards and assessments: The role of reviewer agreement. *Educational Measurement: Issues and Practice*, *26*(1), 17–29. https://doi.org/10.1111/j.1745-3992.2007.00091.x

Appendix A. Science Content Knowledge (CK) Areas[2]

CK 1. Earth's Place in the Universe

This topic area focuses on observable motions of the Sun, Moon, and stars as well as quick and slow events on Earth and the resulting patterns that occur.

By collecting observations of the Sun, Moon, and stars, predictable patterns, such as the Sun's rising and setting and stars being more visible at night, are described and used to answer scientific questions. Graphical displays of collected data reveal these patterns. This topic area targets explanations (e.g., Earth rotates about an axis with a fixed orientation and orbits around the Sun) for seasonal patterns of sunrise and sunset, the seasonal appearance of some stars, daily changes in the relative amount of light, and the length and direction of shadows.

This topic area also uses local, regional, and global patterns of rock formations to construct and support explanations about the changes in landscapes over time. These changes are a results of events that occur very rapidly, such as earthquakes, and others that occur so slowly that a person would not directly observe the change as it happens, such as erosion of rocks.

CK 2. Earth's Systems

This topic area focuses on Earth's materials and systems, such as plate tectonics, the influence of water, weather and climate, and biogeology.

Investigations determine the effects of weathering or the rate of erosion by the motion of water, ice, wind, and living things on the shape of Earth's materials (e.g., rocks, soils, and sediment). This topic area targets the distribution patterns of volcanoes and earthquakes, which can be described by examining maps or graphical displays of Earth's features, such as mountains, volcanoes, and ocean ridges and trenches. Maps of natural features, and graphs and maps showing the amount and location of salt water or freshwater in Earth's reservoirs (lakes, streams, oceans) can be made to better understand the nature of Earth's land features and water resources, respectively. Design solutions that slow or prevent wind or water from changing the shape of the land are generated and evaluated. Models and maps are used to describe ways that Earth systems (geosphere, biosphere, hydrosphere, atmosphere) affect Earth's surface materials and processes.

This topic area also includes the use of observations about local weather conditions to make claims about patterns, such as the typical weather conditions expected at different times (seasons) and different areas (climates).

CK 3. Earth and Human Activity

This topic area focuses on the relationships between living things and the natural resources they depend on as well as the impact of natural hazards.

Simple models can be used to present relationships between plants and animals and their habitats. This topic area targets an understanding that human activity, such as changes to landforms, diversion of water, and the addition of substances to the air, can negatively affect the world around us. Information collected on how the use of energy and fuels derived from natural resources affects the environment is used to make informed decisions about the use of both renewable and nonrenewable natural resources, and ways to reduce the impact of human activities on environments.

This topic area also addresses how forecasts can be used to prepare for severe weather events. Examples include (a) making claims about the merits of a design to reduce the impact of a weather-related hazard, (b) describing evidence about the design solution, and (c) evaluating evidence to determine if the solution adequately addresses the effects of the weather hazard. This process can also be applied to other natural hazards, such as earthquakes, tsunamis, and volcanic eruptions.

CK 4. From Molecules to Organisms: Structures and Processes

This topic area focuses on how plants and animals use their internal and external structures for survival, growth, behavior, reproduction, and processing information. Observations from studying how organisms use their body parts in different ways and for different purposes can be used to design a device to solve a specific human problem, such as designing clothing for protection or warmth. These observations can support an argument that these internal and external structures function and interact to support survival, growth, behavior, and reproduction.

This topic area explores and identifies the similarities and differences in the lifecycles of plants and animals. Models describe lifecycles of organisms and promote the understanding of the relationships among the components of the lifecycles. The models can make predictions as to what would happen if components of the lifecycle were altered.

This topic area also includes claims about how organisms obtain the energy and materials needed for survival. Plants need air, water, and light to produce plant matter and to grow. Animals eat plants, other animals, or both in order to obtain the materials and energy they need for growth, survival, and reproduction. Observations, models and other sources of evidence can be used to identify these patterns and relationships and support these claims. Models promote an understanding of the sense receptors animals use to receive different types of information from the environment. This information is processed by the brain and leads to appropriate actions, such as vocalizations, feeding activities, and protective reactions.

CK 5. Ecosystems: Interactions, Energy, and Dynamics

This topic area focuses on ecosystems—interdependent relationships, movement and cycling of matter, transfer of energy, and social interactions.

Investigations can be planned and conducted to determine that plants need light and water to grow. This topic area explores ways that interactions help organisms survive. Models, such as using pollen sticks to mimic the fuzzy bodies of bees, show how animals facilitate seed dispersal or plant pollination. Models describe the movement of materials in a system that allow species to meet their needs and identify the relevant components of the system (plants, animals, decomposers, matter, other environmental factors), and the role of each component. The models can be used to determine how changes such as the effect of a newly introduced species or a change in the environment affect the system.

This topic area also targets evidence, data, or models to support the claim that some animals form groups, and that being a member of a group helps each member survive. For example, groups experience more success in defending themselves and can make faster, more effective adjustments to harmful changes in their ecosystem than animals acting alone.

CK 6. Heredity: Inheritance and Variation of Traits

This topic area focuses on the conditions that influence characteristics (traits), including both inherited and environmental factors.

The claim that young plants and animals are similar to their parents—leaves from the same kind of plant are the same shape but can differ in size; a particular breed of dog looks like its parents but is not exactly the same – can be articulated and supported using data collected by comparing traits of plant and animal parents and offspring. This data can be used to identify relevant patterns of similarities and differences between parents and offspring in plants and animals that indicate traits are inherited and can vary.

Similarly, this topic area addresses evidence that can be collected to support the claim that many inherited traits can be influenced by the environment. For example, normally tall plants grown with insufficient water are stunted, and a pet dog that is given too much food and little exercise may become overweight.

CK 7. Biological Evolution: Unity and Diversity

This topic area focuses on evolution and the variability of organisms, over time, within a population, and in different habitats.

Fossil data associated with various geologic layers at a site provide evidence of environments that used to exist long ago, such as a tropical forest or an ocean. Fossil evidence can be used to describe ancient environments that supported species no longer present in the current environment and to construct arguments about why these extinct species could not survive in a changed environment.

This topic area targets using evidence about the variation of characteristics within a population, such as protective coloration among insects or the presence of thorns among plants, to develop an understanding that some individuals are better suited to survive and reproduce in their environment than others. This evidence supports the observations that (a) each environment has its own mix of organisms (diversity), (b) specific traits would give an organism a survival advantage in a given environment, and (c) a specific change in the environment would lead to a change in the relative advantage of different traits.

This topic area also addresses the effect of human activity on the environment and involves analyzing solutions to mitigate the influence of a selected activity. Examples include (a) mitigating pollution from waste disposal by recycling or composting and (b) reducing the diversion of water for agriculture through various water conservation policies.

CK 8. Matter and Its Interactions

This topic area focuses on matter—its structures and properties, the physical and chemical changes that can occur to it, and the particles that make it up. Investigations can be carried out to observe, measure, and identify various materials according to their properties, such as reflectiveness, color, and hardness. This topic area targets reversible and irreversible changes that happen to materials from heating, cooling, and mixing of substances. Evidence can be used to categorize changes such as those observed when (a) heating butter or an egg, or (b) freezing water or a leaf.

Investigations address whether substances that interact in a chemical reaction become new substances with different properties. Evidence from these investigations can support an argument that the overall weight of the materials is conserved regardless of what change occurs, including the apparent disappearance of materials. This topic area also addresses the idea that objects are made of small pieces that can be taken apart and recombined to form a new object and the idea that all matter consists of particles that are so small as to be invisible. Students are expected to develop a particle model that could explain why adding an invisible gas to a balloon increases its volume, or sugar added to a glass of water seems to disappear.

CK 9. Motion and Stability: Forces and Interactions

This topic area focuses on forces, motion, and the interaction of objects—pushes and pulls, balanced and unbalanced forces, and patterns of motion.

Investigations determine the effects of varying strengths and directions of pushes and pulls on the motion of objects, such as stopping a rolling ball with a barrier, using a ramp to increase the speed of a marble, and two toy cars colliding with each other. The forces exerted on an object can change its speed and/or direction of motion, keep it at rest, or start it in motion. An object's pattern of motion can be used to predict future motion, such as the motion of a pushed swing.

This topic area also addresses the cause and effect of interactions between two objects that are not in contact with each other. This might involve electrical forces between a charged comb and someone's hair or magnetic forces between a magnet and some nails. Evidence is collected to argue that the gravitational force of Earth always pulls objects downward toward the center of the planet. Examples include: (a) dropping objects from the second story of a building, and (b) observing the movement of a ball on an inclined ramp.

CK 10. Energy

This topic area focuses on energy—its definition, conservation, and transfer, and how it is captured, stored, and released by physical and chemical processes.

Evidence is used to construct explanations about the relationship between the speed and energy of an object and how energy is transferred from place to place by means such as sound, light, heat, and electrical currents. Sunlight warms the surface of Earth, and solutions can be designed to solve the problem of reducing that warming effect where needed (e.g., umbrellas and awnings).

This topic area also addresses the transfer in energy due to the change in speed or direction that occurs when objects interact. Models can be used to predict the outcomes of certain collisions between objects such as pool balls, marbles, and toy cars. Devices are designed that convert one form of energy (e.g., light energy) to another (e.g., electrical energy). Models such as energy and matter diagrams and food webs describe how energy in food sources used by animals for movement, growth, and repair came from the Sun and was captured by plants for use and storage.

CK 11. Waves and Their Application in Technologies for Information Transfer

This topic area focuses on observing the behavior of waves, studied through activities relating to light, sound, and waves in water.

This topic area addresses the observation that light travels in a straight line and that objects can only be seen when illuminated or when they emit light. Observations could be made of objects in the dark being illuminated by lights, such as a person entering a dark room and lighting a candle. Investigations explore the effect of placing various types of objects in the path of a light source. Opaque, transparent, and translucent objects allow all or some of the light through, and mirrors reflect the light. A model shows why only light from objects that enters the eyes can be seen, explaining, for example, why an object can be seen through a window, but not through a brick wall, or why a mirror can allow objects to be seen around corners or over fences.

This topic area explores the nature of mechanical waves. Investigations determine that vibrating materials, such as a tuning fork or a guitar string, produce sound waves. These waves can make other objects vibrate, as demonstrated by placing a vibrating tuning fork into a container of water. The physical waves formed on the surface of the water can be characterized by amplitude (wave height) and wavelength (distance between successive wave peaks).

This topic area also studies how information can be carried by waves. Wave characteristics lead to the design of devices that use a pattern of light or sound to send information over a distance, such as Morse code.

CK 12. Engineering Design

This topic area focuses on introducing students in the earliest grades to the term *problem* as indicating a situation that people want to change.

Students can use tools and materials to solve simple problems, use different representations to convey solutions, and compare different solutions to a problem and determine which is best. Students in all grade levels are not expected to come up with original solutions; instead, the emphasis is on thinking through the needs or goals that need to be met and determining which solutions best meet those needs and goals.

At the upper elementary grades, this topic area engages students in more formalized problem solving. Students define a problem using criteria for success and constraints or limits of possible solutions. Students research and consider multiple possible solutions to a given problem. Generating and testing solutions also becomes more rigorous as the students learn to optimize solutions by revising and retesting to obtain the best possible design.

Appendix B. Science Teaching Practices (STPs)

STP	Description
STP 1	Selecting or sequencing age-appropriate, grade-level instructional goals or big ideas for a topic
STP 2	Identifying the big idea or instructional goal of an instructional activity
STP 3	Linking science ideas to one another and to particular activities, models, and representations within and across lessons
STP 4	Choosing which science ideas or instructional activities are most closely related to a particular instructional goal
STP 5	Selecting investigations or demonstrations that facilitate understanding of disciplinary core ideas, scientific practices, or crosscutting concepts
STP 6	Evaluating investigation questions for quality
STP 7	Determining the variables, techniques, or tools that are appropriate for use by students to address a specific investigation question
STP 8	Critiquing scientific procedures, data, observations, or results for their quality, accuracy, or appropriateness
STP 9	Evaluating and selecting media for engaging students in virtual investigations not possible in firsthand situations
STP 10	Supporting students in generating questions for investigation or identifying patterns in data and observations
STP 11	Evaluating instructional materials and other resources for their ability to sufficiently address scientific concepts; engage students with relevant phenomena; develop and use scientific ideas; promote students' thinking about phenomena, experiences and knowledge; provide a sense of purpose; take account of students' ideas; and assess student progress
STP 12	Choosing resources that support the selection of accurate, valid, and age-appropriate goals for science learning
STP 13	Analyzing student ideas for common misconceptions regarding intended scientific learning
STP 14	Selecting diagnostic items and eliciting student thinking about scientific ideas and practices to identify common student misconceptions and the basis for those misconceptions
STP 15	Developing or selecting instructional moves, approaches, or representations that provide evidence about common student misconceptions and help students move toward a better understanding of the idea, concept, or practice
STP 16	Identifying the connections between students' talk and work and scientists' talk and work
STP 17	Selecting scientific language that is precise, accurate, and grade-appropriate and illustrates key scientific concepts
STP 18	Anticipating scientific language and vocabulary that may be difficult for students
STP 19	Supporting and critiquing students' participation in and use of verbal and written scientific discourse and argumentation
STP 20	Modeling the use of appropriate verbal and written scientific language in critiquing arguments or explanations, in describing observations, or in using evidence to support a claim
STP 21	Critiquing student-generated explanations and descriptions for their generalizability, accuracy, precision, or consistency with scientific evidence
STP 22	Selecting explanations of scientific phenomena that are accurate and accessible to students
STP 23	Evaluating or selecting scientific models and representations that predict or explain scientific phenomena or address instructional goals
STP 24	Engaging students in using, modifying, creating, and critiquing scientific models and representations that are matched to instructional goals
STP 25	Evaluating student models or representations for evidence of scientific understanding
STP 26	Generating or selecting diagnostic questions to evaluate student understanding of specific models and representations
STP 27	Evaluating student ideas about what makes for good scientific models and representations

Note. Adapted from "Assessing Elementary Teachers' Content Knowledge for Teaching Science for the ETS® Educator Series: Pilot Results" by Jamie N. Mikeska, Christopher Kurzum, Jonathan H. Steinberg, and Jun Zu, 2018, Research Report No. RR-18-20, p. 7, Table 3. Copyright 2018 by Educational Testing Service.

Notes

[1] The English language arts, mathematics, and science components of this new licensure assessment are currently available. The plan is for the social studies component to be available in fall 2019.

[2] Adapted from *The PRAXIS® Study Companion. Elementary Education: Content Knowledge for Teaching* (pp. 106–112), by Educational Testing Service, 2017, Princeton, NJ, Author. Copyright 2017 Educational Testing Service.