Received: 16 November 2021 | Revised: 15 November 2022 | Accepted: 21 November 2022

DOI: 10.1002/sce.21779

RESEARCH ARTICLE

Science Education WILEY

Examining elementary science teachers' responses to assessments tasks designed to measure their content knowledge for teaching about matter and its interactions

Jamie N. Mikeska[1] | Dante Cisterna[1] | Heena Lakhani[2] | Allison K. Bookbinder[3] | David L. Myers[4] | Luronne Vaval[3]

[1]K-12 Teaching, Learning, and Assessment Center, ETS, Lawrence Township, New Jersey, USA

[2]College of Education, University of Washington, Seattle, Washington, USA

[3]Department of Mathematics, Science, and Technology, Teachers College, Columbia University, New York, New York, USA

[4]College of Education, University of Georgia, Athens, Georgia, USA

Correspondence
Jamie N. Mikeska, ETS, Department of Mathematics, Science, and Technology, Lawrence Township, NJ, USA.
Email: jmikeska@ets.org

Abstract

Despite the importance of developing elementary science teachers' content knowledge for teaching (CKT), there are limited assessments that have been designed to measure the full breadth of their CKT at scale. Our overall research project addressed this gap by developing an online assessment to measure elementary preservice teachers' CKT about matter and its interactions. This study, which was part of our larger project, reports on findings from one component of the item development process examining the construct validity of 118 different CKT about matter assessment items. In this study, 86 elementary teachers participated in cognitive interviews to examine: (a) the knowledge and reasoning they used when responding to these CKT about matter assessment items and (b) the nature of the content challenges and the content teaching challenges they encountered. Findings showed that over 80% of participant interview responses indicated that the CKT about matter items functioned as hypothesized, providing evidence to support future use of these items on a large-scale assessment and in studies of science teachers' CKT. When responding to the items, participants showed evidence of four main challenges with the science content: (a) using scientific concepts to reason about science tasks, (b) using adequate evidence to reason about science phenomenon, (c) drawing upon examples of scientific phenomena, and (d) drawing upon science vocabulary. Findings also showed that participants experienced challenges regarding the following content teaching aspects when responding to these items: (a) connecting to key scientific concepts involved in the work of teaching science, (b) attending to instructional goal(s), and (c) recognizing features of grade-level appropriateness. Implications for using CKT items as part of large-scale science assessment systems and identifying areas to target in elementary science teachers' CKT development are addressed.

KEYWORDS

assessments, content knowledge for teaching, elementary science, matter and its interactions, pedagogical content knowledge

High-quality science instruction requires that science teachers leverage their content knowledge for teaching (CKT) in the work of teaching (National Research Council, 2012; NGSS Lead States, 2013). CKT refers to the professional knowledge that teachers draw upon as they engage in the work of teaching in a specific discipline (Loewenberg Ball et al., 2008). The main focus here is on the application—or use—of their knowledge in the actual tasks of teaching—for example, the knowledge that teachers use when they prepare for, enact, and reflect on their instruction with students. CKT includes both subject matter knowledge, as well as other forms of practice-based knowledge that are directly tied to the work of teaching—what is commonly referred to as teachers' specialized and pedagogical content knowledge. In the area of science this includes knowledge about how students think about, engage in, and learn about specific scientific practices and concepts, as well as knowledge about various instructional strategies and tools science teachers can draw upon to develop students' learning (Carlson et al., 2019; Mikeska et al., 2020; Schneider & Plasman, 2011).

In science education, research has indicated that CKT, particularly science teachers' pedagogical content knowledge, can be quite nuanced and variable across subjects, topics, concepts, and even knowledge components (e.g., knowledge of student ideas vs. knowledge of instructional strategies) (Bertram & Loughran, 2012; Hanuscin et al., 2018; Henze & van Driel, 2015; Loughran et al., 2004; Mikeska et al., 2020; Park & Suh, 2015). Research has suggested that CKT is important for science teachers as they engage in various science teaching practices (STPs), such as eliciting, interpreting, and using students' scientific ideas and selecting scientific models, investigations, and demonstrations to address student learning goals. Overall, research findings have suggested that teachers with more advanced and robust CKT in specific science areas tend to have higher instructional quality, which directly impacts student learning (Davis et al., 2006; Kam Ho Chan & Hume, 2019; Roth et al., 2011; Sadler et al., 2013; Schneider & Plasman, 2011).

Despite the importance of developing science teachers' CKT, there are limited assessments that have been designed to measure their CKT at scale (Minner et al., 2012; National Academies of Sciences Engineering and Medicine, 2015; Wilson, 2016); the ones that do exist tend to focus on measuring only one aspect of science teachers' CKT, such as their subject matter knowledge (Chen et al., 2020; Sadler et al., 2013; Smith, 2010) or pedagogical content knowledge (Bertram & Loughran, 2012; Henze & van Driel, 2015; Park & Suh, 2015;

Roth et al., 2011). In addition, most use approaches that preclude large-scale use because they include constructed response formats such as in-depth interviews or open-ended survey questions. This gap in the science assessment landscape hinders researchers, teacher educators, school district leaders, and professional development facilitators from efficiently pinpointing the strengths and areas of need for individual or groups of science teachers and limits the field's ability to assess science teachers' CKT across content areas, sites, or longitudinally across the professional continuum (National Research Council, 2013; Wilson et al., 2019).

1 | RESEARCH QUESTIONS

Our larger research project was designed to address this gap by developing an online assessment to measure the full breadth of elementary preservice teachers' CKT in one high-leverage science content area: matter and its interactions. This assessment focuses on the specialized knowledge that elementary teachers use to address the concepts and practices embedded in the set of eight performance expectations identified in the area about Matter and its Interactions (PS1) within the Next Generation Science Standards (NGSS) at the elementary level. Our overall goal was to create a valid and reliable CKT instrument that could be used within elementary science methods courses to gauge elementary preservice teachers' CKT about matter and its interactions.

To do so, our larger research project used the principles of evidence-centered design (Mislevy & Riconscente, 2006) along with a process closely modeled after the item development work on the Measures of Effective Teaching project (Phelps et al., 2014) and an earlier National Science Foundation EAGER project (Mikeska et al., 2017) to develop the CKT about matter assessment instrument. This iterative development process required examining individual CKT about matter items and the overall assessment form in terms of response processes, content validity, structural validity, the use of assessment information to provide feedback, and external construct validity. Other publications and presentations have addressed some of these criteria to assess the adequacy of this newly developed instrument to make valid inferences about preservice teachers' CKT about matter (Castellano & Mikeska, 2022; Cisterna et al., 2022; Mikeska et al., 2020). The study reported in this manuscript focuses on our examination of preservice teachers' responses when selecting their answer choice for individual CKT about matter items; this examination targets construct validity to ensure that the CKT about matter items are measuring preservice teachers' CKT about matter.

The main purpose of the proof-of-concept study reported in this manuscript was to determine whether our team could develop CKT about matter items that could be dichotomously scored and later incorporated into a CKT assessment to measure potential changes in elementary preservice teachers' CKT about matter across one semester. In this study, 86 elementary teachers (79 preservice and 7 in-service teachers) participated in cognitive interviews to examine the knowledge and reasoning they used when responding to 118 different CKT assessment items about matter and its interactions. The purpose of the cognitive interviews was to elicit CKT from the target population in response to the CKT about matter items, to provide information for refining the CKT about matter items for large scale use, and to examine the nature of the challenges that elementary science teachers encountered when leveraging their CKT in this content area. The present study addresses the following research questions:

- **Research Question 1**: To what extent do elementary teachers use the intended knowledge and reasoning when responding to items designed to assess their CKT about matter?
  - o Research Question 1a: When answering the items correctly, what proportion of teacher responses used the intended knowledge and reasoning?
  - o Research Question 1b: When answering the items incorrectly, what proportion of teacher responses did not use the intended knowledge and reasoning?

- **Research Question 2**: When elementary science teachers did not use the intended knowledge and reasoning when responding to these items, what patterns in their knowledge and reasoning were observed?

- Research Question 2a: What is the nature of the content challenges these elementary teachers encountered when responding to these CKT about matter items?
- Research Question 2b: What is the nature of the content teaching challenges these elementary teachers encountered when responding to these CKT about matter items?

The first set of research questions seeks to understand, overall, how the developed CKT about matter items meet the assessment intent and justification. This analysis provides evidence to determine whether the items function as intended and adequately assess the teachers' CKT in this science area. The second set of research questions explores in more detail how participants reasoned about the assessment intent and justification, which provides a way to make inferences about participants' CKT. These analyses provide a more contextualized perspective of participants' CKT in this area and provide a deeper understanding into the patterns that were observed in teachers' reasoning.

This manuscript begins with a background section explicating the current research base in three areas: (a) challenges new science teachers encounter when preparing for, engaging in, and reflecting on NGSS-aligned instruction, (b) teaching and learning about matter and its interactions, and (c) CKT about science and its measurement. We then provide details about the study's methodology, including the development of the CKT about matter assessment items used in this study. We end with the study's findings and a discussion of implications for using CKT items as part of large-scale assessment systems and identifying areas to target in elementary science teachers' CKT development.

2 | BACKGROUND

2.1 | Challenges experienced by new science teachers

In this paper, we strive to seek a balance in how we talk about the continuous journey of learning to teach science. We acknowledge the rich array of knowledge, experiences, and practices teachers bring to teaching science (Gray et al., 2022). For example, recently Cisterna et al. (2022) explored preservice teachers' perceptions of the relevance of CKT item scenarios and found that elementary preservice teachers relied on a broad repertoire of prior experiences and knowledge when explaining how and why these scenarios related to the work of teaching elementary science. At the same time, we acknowledge that the practice of teaching science is complex and involves multiple, coordinated learning opportunities along the preservice to in-service teaching continuum. Thus, we do not view the challenges that teachers face from a deficit or problem-focused lens. Rather, we view them as opportunities to identify areas for continued support as teachers develop and grow in their understanding of science content and science instruction, which can inform efforts to develop well-prepared elementary science teachers. Science teachers, especially those who are learning to and have recently become new classroom teachers, need to develop the knowledge and skills to plan for, enact, and reflect on their science instruction with their students. As noted in Davis et al.'s (2006) comprehensive and widely cited literature review on this topic, the challenges that new science teachers face, which they identify as preservice teachers in teacher education or alternative certification programs and those who are in their first five years of teaching, can be described in five key themes.

The first theme focuses on the processes by which new science teachers build a strong knowledge base regarding their understanding of scientific concepts and the way that scientific knowledge is constructed in specific disciplines. In general, studies show that new science teachers, especially preservice teachers at the elementary level, need opportunities to develop sophisticated understandings of science content, scientific inquiry, and the nature of science, although there is variability among and within individuals as they develop these understandings (Davis et al., 2006). Most importantly, new science teachers' understandings can improve over time in response to high-quality learning opportunities and through further teaching. For example, first-year elementary teachers can exhibit robust ways of engaging in inquiry-based approaches with their students (Avraamidou & Zembal-Saul, 2010). Thus, teachers can be "well-started" in their abilities to engage their students in scientific inquiry and continue their development through multiple learning experiences, including their in-classroom teaching experiences (Avraamidou & Zembal-Saul, 2010).

The second theme focuses on how teachers develop understanding about recognizing students as learners and understanding how students learn and develop. In general, studies have shown that new science teachers value the importance of paying attention to students and their ideas and that preservice teachers want to understand their students in many ways, including their interests and experiences (Mikeska et al., 2009). However, they need support to understand the complexity or sophistication of students' ideas and determine how to respond productively during instruction to those ideas (Forbes et al., 2015). As Mikeska et al. (2009) note, the goal is to help teachers refine their focus so that they can better understand the intellectual resources that students bring. Drawing teachers' attention to students' thinking at various points during science instruction can help teachers learn to consider and use these ideas productively when teaching.

The third theme targets the challenges that new science teachers encounter when planning and enacting instruction, which includes selecting and using topic- and subject-specific instructional activities, representations, teaching strategies, and assessment approaches to support student learning. Research has shown that teachers can enact high-quality, inquiry-based practices (Avraamidou & Zembal-Saul, 2010), have rich ways of understanding students' ideas and sense-making (Mikeska et al., 2009), and can recognize the importance of specific aspects of the work of teaching science (WOTS) even if they have not directly experienced them yet (Cisterna et al., 2022). However, findings generally on this theme "…illustrate a mismatch between teachers' ideas and practices—their ideas about instruction seem generally to be more sophisticated and innovative than are their actual practices" (Davis et al., 2006, p. 621). Thus, developing new science teachers' abilities to engage in ambitious STPs that align with the vision of the science framework and NGSS involves attending to teachers' ongoing learning trajectory through science teacher education and professional development (National Research Council, 2012; Windschitl et al., 2020).

The final two themes focus on the challenges new science teachers encounter as they work to create productive science learning environments and become effective science teachers who are professionals and reflective practitioners. In terms of creating productive learning environments, findings have shown that new science teachers sometimes face classroom management difficulties, which can make it challenging for them to engage their students in reform-oriented science instruction. In terms of becoming effective and reflective practitioners, the current research suggests the importance of support from colleagues, opportunities to reflect critically on their science teaching, and learning experiences that develop their self-efficacy and science teacher identity (Richmond & Manokore, 2011).

Since the publication and release of the National Research Council's Framework for Science Teaching (2012) and the NGSS (NGSS Lead States, 2013), research also has indicated that science teachers across the teaching continuum have encountered challenges in planning for and productively engaging K-12 students in three-dimensional science learning. Addressing this challenge requires that teachers develop a deep understanding of the vision of high-quality science instruction and its coherence to support productive implementation (Penuel et al., 2015; Pruitt, 2014). Developing CKT is one of the steps to implement the NGSS and address challenges such as knowing how and being able to: productively engage students in explaining real-world phenomena, appropriately scaffold students in three-dimensional learning, successfully build from students' previous learning experiences, and achieve strong coherence across students' learning opportunities (Bybee & Chopyak, 2017; Gale et al., 2019; Ryan et al., 2017; Tyler et al., 2020). Teachers also must consider how to best align these NGSS-focused science learning opportunities to the learning standards in other disciplines, such as English language arts and mathematics, and how to address the language demands that specific student populations, such as English language learners, will encounter during NGSS-aligned instruction (Lee et al., 2013; Lee, 2017; Tolbert et al., 2014). These challenges can surface no matter what the science instructional focus is or what curriculum materials teachers are using to address the NGSS, although research has indicated that learning opportunities in teacher education and professional development can help teachers address these challenges and learn to engage in more NGSS-aligned instructional practice (Hanuscin & Zangori, 2016; Isabelle, 2017; Kang et al., 2019; Sinapuelas et al., 2018; Tuttle et al., 2016).

To address these challenges, science teacher educators, researchers, and professional development leaders have offered several learning opportunities within preservice and in-service teacher education coursework and professional development. Findings that are most relevant for our study suggest that, to become effective science teachers (Windschitl et al., 2020), preservice teachers, especially those at the elementary level, require targeted opportunities to develop their understanding of scientific concepts (Del Pozo, 2001; Ginns & Waters, 1995), their knowledge of strategies to promote student learning, and their ability to engage in high-quality science instructional practices. Typically, these learning experiences occur as part of elementary science methods courses in teacher education programs and through their fieldwork experiences under the mentorship of experienced elementary science teachers. In addition, such experiences tend to focus on the development of elementary preservice teachers' understandings and teaching abilities in specific science content areas or topics, as research has foregrounded the importance of the topic-specific nature of the professional knowledge that science teachers leverage during instruction (Gess-Newsome, 2015; Sadler et al., 2013). Using assessment instruments, like the CKT about matter instrument we developed as part of our larger research project, can assist teacher educators and professional development leaders in determining the assets that preservice elementary teachers bring to these learning opportunities and in identifying the potential areas for continued growth that would be helpful to address in their future learning.

2.2 | Teaching and learning about matter

One of the core disciplinary areas in the NGSS (NGSS Lead States, 2013) focuses on matter and its interactions and emphasizes the structure and properties of matter and chemical reactions across the K-12 instructional progression (Krajcik & Merrit, 2012). Developing CKT about a specific science area, such as matter and its interactions, is essential for being able to implement high-quality instruction addressing the NGSS performance expectations. Researchers and scientists have frequently argued that in-depth knowledge of matter concepts is critically important to develop students' scientific understanding (Harrison & Treagust, 2003). For example, the NGSS note how K-12 students' ideas tend to progress in this topic area from understanding that “matter exists as different substances that have different observable properties” to understanding that “matter exists as particles that are too small to see” to then understanding that matter is comprised of atoms and molecules that can be rearranged; this knowledge can be used to explain and predict various scientific phenomena, such as states of matter, conservation of matter, phase changes, and chemical changes (NGSS Lead States, 2013 , p. 7). As such, much research in this area has targeted examining the nature of K-12 students' ideas and understandings about various matter concepts including the composition and structure of matter (Gómez, et al., 2006; Hadenfeldt et al., 2016; Harrison & Treagust, 2003; Nakhleh & Samarapungavan, 1999; Nakhleh et al., 2005; Talanquer, 2009) and physical and chemical properties and change (Taber & García-Franco, 2010; Varelas et al., 2006).

One of the most comprehensive research studies in this area examined how students progress in their understanding of four key aspects of matter—conservation, physical properties and change, chemical properties and change, and structure and composition—using data from the Third International Mathematics and Science Study (Liu & Lesniak, 2005). Findings illustrated the complexity of these matter aspects as all four of them were closely related and overlapped with each other across the elementary and secondary grade levels. For example, Liu and Lesniak's study suggested that in order for students to be able to explain how changes to matter occur using the particulate model, students need to draw upon their understanding of various concepts including what is involved in changes to matter, how and why matter is conserved, and an understanding of the properties of matter. These findings illustrate how different concepts are leveraged to make sense of a phenomenon. Furthermore, Merritt et al. (2008) pointed out that understanding the particulate nature of matter is one of the most difficult concepts for students, especially because traditional curriculum materials tend to introduce this idea without making connections between the particulate model of matter and specific real-world phenomena.

Other studies have used different approaches, such as interviews, open-ended written surveys, or responses to class assignments, to investigate the nature of students' matter conceptions. For example, Krnel et al. (2003) conducted interviews with children from ages 3 to 13 to understand how and why they classified multiple sets of objects and matter. Findings showed that as students progressed in age they were more likely to classify matter using intensive properties while younger children tended to use a combination of both extensive and intensive properties. Likewise, Taber and García-Franco (2010) conducted interviews with 11–16-year-old students where the students were asked to describe and explain scientific phenomena from their everyday lives (e.g., how ice floats on water; how dye spreads out in water). Interview responses were used to examine how students activated and used their prior knowledge and experiences related to the particulate model of matter. The work of Merritt et al. (2008) showed that, with appropriate instruction, students can refine their initial ideas about matter as a combination of particles and continuous substance to then understand that all substances are made up of particles and explain how they behave in a particular state of matter. Collectively these studies suggest that K-12 students vary in their understanding of matter concepts, their understanding across matter concepts is strongly related, and they can leverage their prior knowledge and experiences to reason about relevant matter-related scientific phenomena, although they sometimes may show difficulties in being able to consistently explain scientific phenomena using matter concepts.

Research has also targeted examining how science teachers, especially preservice and novice teachers, understand similar matter concepts (Aydeniz et al., 2017; Bayuni et al., 2018; del Pozo, 2001; Ginns & Watters, 1995; Kahveci, 2009; Kokkotas et al., 1998; Özalp & Kahveci, 2015; Smith & Plumley, 2016). In one study (Valanides, 2000), researchers interviewed elementary teachers about changes in the macroscopic and microscopic properties of different substances due to dissolution, filtering, or heating. Findings indicated that these teachers struggled to understand and use the particulate nature of matter to explain and connect specific macroscopic and microscopic properties and events in the context of these phenomena. Bayuni et al. (2018) found that many elementary preservice teachers have difficulty explaining the properties of gases and particular changes of state, such as between liquids and gases. Aydeniz et al. (2017) showed that many preservice teachers held intuitive ideas about the particulate nature of matter, such as thinking that particles of vapor do not have weight. In a different study, Del Pozo (2001) had elementary preservice teachers create conceptual maps showing their understanding of the connection between concepts related to the composition of matter. They found that most teachers either had correct, yet incomplete, relationships represented in their conceptual maps or had very limited to no knowledge of relationships between these matter concepts. These studies suggest that challenges in understanding, applying, and connecting matter concepts are well-known challenges for science teachers, especially prospective ones, and can be similar to some of the challenges that K-12 learners face. However, it is important to note that these studies tended to focus more on describing teachers' difficulties rather than discussing alternative conceptions or emerging reasoning patterns that they can use to generate explanations about matter and its interactions.

While less common, some studies have investigated the instructional approaches that science teachers use when teaching about matter and its interactions and the challenges they encounter when doing so (Kang, 2007; Lee et al., 1993). Other studies have examined the professional knowledge that teachers have access to and use to address matter concepts (Hanuscin et al., 2018; Mikeska et al., 2020; Smith et al., 2017). In relation to the learning goal of students developing a particulate model of matter, Hanuscin et al. (2018) found that in-service teachers often designed instructional activities that included representations of particle models already created (as a final model) and displayed a limited number of assessment strategies to identify student ideas about the behavior of matter particles. Similarly, Smith et al. (2017) noted that elementary teachers rarely enacted these topics in ways that reflect the three-dimensional vision of science learning described in the NGSS. For example, teachers tended to focus more on introducing the particulate model of matter to their students than on using evidence from different phenomena to develop such a model to explain different properties. In the context of professional development programs, Kruse et al. (2020) documented that in-service teachers are able to refine their initial ideas about matter-related concepts into more accurate understandings but they face challenges representing or explaining matter-related concepts in different ways. Other practitioner-focused literature has highlighted lesson activities, instructional sequences, and materials that teachers can use to help students develop their understanding of matter concepts and specific phenomena (Jackson, 2009; Lee et al., 2014; Lott & Jensen, 2012; Lott & Wallin, 2012; Royce, 2012; Troncale, 2016; Vowell & Phillips, 2015; Weishaar, 2011), as well as specific formative assessment probes and activities that teachers can use to elicit students' thinking about matter (Eberle & Keeley, 2008; Keeley, 2016; Keeley et al., 2007; Palmeri et al., 2008). Overall, these studies suggest there has been limited focus on: (a) exploring the reasoning underlying teachers' instructional decisions about teaching matter or (b) in developing and using assessments that can be administered and scored efficiently for large scale use to measure the full range of professional knowledge that elementary science teachers use when planning for, enacting, and reflecting on their instruction in this high-leverage content area.

2.3 | CKT science and its measurement

Content knowledge for teaching, or CKT, refers to the usable knowledge that teachers draw upon as they engage in the work of teaching (Etkina et al., 2018; Ball et al., 2008; National Academies of Sciences Engineering and Medicine, 2015; National Research Council, 2012). CKT includes both teachers' subject matter knowledge and their pedagogical content knowledge. For example, when addressing the NGSS performance expectations in this content area, elementary science teachers need to draw upon their understanding of key concepts and ideas related to matter and its interactions, such as their knowledge about ways to describe and classify matter and why matter is conserved in various situations, and their understanding of crosscutting concepts and scientific practices that students engage in to develop their understanding in this area, such as understanding how to support students in using standard units to measure and describe physical quantities and how students can graph and measure quantities like weight to provide evidence that the total weight of matter is conserved in various changes. Elementary science teachers also need to be able to use their knowledge about different activities and investigations that are useful to help students develop their own understanding about properties of matter and leverage the varied previous experiences and conceptions students bring with them to their learning in this content area. In the context of authentic, classroom-based settings, teachers use their knowledge about the science content and content teaching, which targets the three dimensions of the NGSS performance expectations, to make decisions about planning or enacting science instruction about matter and its interactions.

Many scholars have investigated teachers' CKT—including their subject matter knowledge and/or pedagogical content knowledge—across content areas. Research studies have shown that teachers' CKT is related to their ability to engage in critical teaching practices, such as eliciting and interpreting students' ideas, selecting and implementing instructional activities to address specific learning goals, and critiquing student-generated explanations (Baumert et al., 2010; Hill et al., 2007, 2008; Kloser, 2014; National Research Council, 2012; Schneider & Plasman, 2011; Windschitl et al., 2012). For example, when addressing a NGSS performance expectation, such as "Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved" (NGSS Lead States, 2013; 5-PS1-2), teachers need to draw upon both their subject matter knowledge and their pedagogical content knowledge and apply it to the work of teaching. This CKT includes their understanding of how matter is conserved in various situations (e.g., phase changes, dissolving, mixing), the ways in which students can produce data to justify their claims about matter conservation, how students can use standard units to measure and describe physical quantities like weight, and their understanding about scientific investigations, scientific representations, and computational analyses they could use to support students in meeting this performance expectation. This CKT is then used as they engage in specific instructional practices, such as selecting which activities to include in an instructional sequence to address this performance expectation, making in-the-moment decisions about how to elicit and probe student thinking around specific phenomenon illustrating the conservation of matter, and determining what feedback to provide students about the varied representations they create to explain their thinking about the conservation of matter for specific phenomenon.

In our study, each of the CKT about matter items was designed to assess the usable knowledge, including both teachers' subject matter knowledge and pedagogical content knowledge, that elementary science teachers need to leverage when addressing one of the NGSS performance expectations in the area of matter and its interactions at the elementary level. Teachers' ability to engage productively in these instructional practices also impacts student learning outcomes. Without access to relevant and adequate CKT, research has shown that teachers tend to struggle to engage in specific instructional practices and are less likely to positively impact student learning (Davis et al., 2006; Schneider & Plasman, 2011).

Much research has been conducted to understand how teachers develop and use their CKT, especially in response to learning opportunities they engage in within teacher education courses and professional development. More recently, researchers have focused on how to best develop formative and summative assessments of teachers' CKT to better understand what and how teachers learn (Etkina et al., 2018; Mikeska et al., 2018). This work on developing, validating, and using measures of teachers' CKT has been most prominent in mathematics. In science teacher education, there has been a concerted effort more recently to develop assessments of science teachers' pedagogical content knowledge, especially their knowledge of student ideas and their knowledge of instructional strategies to address the student-level science standards. For example, Roth and colleagues (Roth et al., 2011) developed an assessment where teachers analyze video clips of science instruction and write open-ended responses about what they notice regarding the science content, student ideas, and instructional strategies used in those clips. Others have used and developed interviews, observational protocols, rubrics, performance assessments, and intake surveys to assess aspects of science teachers' CKT (Bertram & Loughran, 2012; Hanuscin et al., 2018; Henze & van Driel, 2015; Park & Oliver, 2008; Park & Suh, 2015).

Other work has focused on developing more scalable assessments of science teachers' subject matter knowledge. One novel approach used a set of misconception-focused multiple-choice assessment items where each item addressed a disciplinary core idea from one of the NGSS; each item had five answer choices—one was the correct answer and one of the remaining four incorrect answers represented a common student misconception (Chen et al., 2020). Both high school science teachers and their students answered the 29-item biology assessment, although the teachers also selected which of the incorrect answers they thought represented the most common student misconception. Findings indicated that teachers who had stronger subject matter knowledge also had students who performed better on the assessment. In addition, teachers who were better able to identify the most common student misconception had even better student performance. Researchers in another study—Horizon Research, Inc.'s Assessing the Impact of the MSPs: K-8 Science (AIM)—developed a suite of online assessments to measure teacher knowledge in particular science content areas, one of which was focused on assessing teacher knowledge about properties and changes in matter. While all items were multiple-choice items and could be administered and scored efficiently, these items were designed to measure one main aspect of teachers' professional knowledge—their knowledge of science content and use of that knowledge to analyze/diagnose student thinking or make instructional decisions.

In general, most of the currently developed assessments only focus on one or two aspects of science teachers' CKT, such as their subject matter knowledge or a specific component of their pedagogical content knowledge (e.g., knowledge of student misconceptions; knowledge of instructional strategies). In addition, due to their specific approaches, most measures require an extensive amount of time to both administer and score. To address this gap, our study focused on developing assessment items to measure elementary teachers' CKT in one high-leverage content area—matter and its interactions—and used teachers' think-aloud responses when interacting with the CKT about matter assessment items to examine their reasoning and the nature of the content challenges and content teaching challenges they encounter when leveraging their professional knowledge to address common scenarios elementary teachers face when teaching about matter. A better understanding of the challenges that elementary teachers face when using their CKT to engage in the WOTS for specific science content areas can support science teacher educators and professional development facilitators in determining how best to design productive learning experiences. The CKT assessment can be used to measure changes to preservice teachers' CKT before and after instruction and provide information to teacher educators about which specific aspects to focus on when building preservice teachers' CKT in this area.

3 | METHODS

3.1 | CKT about matter item development

Our team, which included 18 external consultants with collective expertise in elementary science teaching, elementary science teacher education, and content about matter and its interactions, developed a set of CKT about matter assessment items using a multistep design process (Castellano & Mikeska, 2022). As shown in Figure 1, each of the CKT about matter items developed is conceptualized to live at the intersection of the science content—in this case one of five different matter topics (e.g., properties of matter, conservation of matter, etc.)—and one of seven WOTS instructional tools categories. The five matter topics were identified and defined based on a review of the specifications from the relevant NGSS performance expectations about matter and its interactions (PS.1), a review of research about student and teacher ideas about matter-related topics (e.g., Krajcik & Merritt, 2012), and curriculum materials focused on those topics (e.g., Kessler & Galvan, 2007). The WOTS framework (Mikeska et al., 2018) identifies 27 STPs that are critical for beginning elementary science teachers to be able to engage in from their first day on the job. This framework was developed in a process co-led by the first author, in which a national panel of elementary science teachers and elementary science teacher educators defined 27 STPs that are organized into seven categories by the types of instructional tools (e.g., scientific models and representations; scientific investigations; etc.) that elementary science teachers work with as they prepare for, enact, and reflect on their

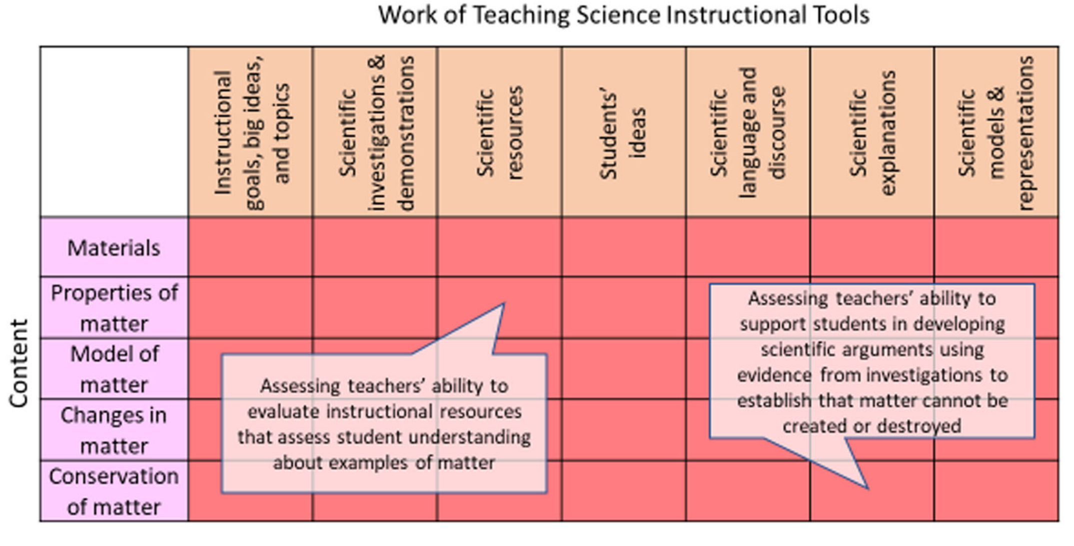

FIGURE 1 Content knowledge for teaching (CKT) matter item matrix

elementary science instruction. Each WOTS instructional tool category includes between three and seven STPs that comprise the work of teaching elementary science; these STPs are ones that require elementary science teachers to use their CKT to engage in them successfully. Table 1 provides a list of the seven WOTS instructional tools categories along with one example of a STP within each category.

Similar to CKT assessment items designed on other projects (Hill et al., 2008; Mikeska et al., 2018; Phelps et al., 2020), all the CKT about matter items our team designed are scenario-based. Each item starts with a task scenario that situates the elementary teacher within an instructional setting and provides key details about the students, curriculum, or instructional context. Each scenario was designed to tackle a specific teaching challenge that elementary science teachers face when preparing for, engaging in, or reflecting on NGSS-aligned science instruction in topics about matter and its interactions. Furthermore, each CKT about matter item was created to address the CKT that preservice teachers need to leverage when addressing one of the NGSS performance expectations within a specific content category about matter and its interactions (e.g., changes in matter) and one specific STP from the WOTS framework. It is important to note that each CKT about matter item was designed to address one specific moment—what we think of a micro moment—of the work of teaching elementary science that teachers engage in within a larger instructional unit. While the full unit likely addresses one or more NGSS performance expectations, the micro moment described in the opening scenario of each CKT about matter item only provides a brief snapshot into the hypothetical teachers' instructional decision-making during planning, enactment, or reflection. In other words, the item addresses the CKT needed to address a very specific aspect of an NGSS PE within a larger instructional unit. As such, the opening instructional scenario only provides the context needed to situate the test taker, in this case the preservice teacher, into the WOTS that is occurring at one small moment in time when a teacher is planning for, engaging in, or reflecting on their science instruction. One limitation of this assessment approach is that it makes it difficult to adequately capture the full complexity of the WOTS across the instructional unit and illustrate how each instructional snapshot addresses the full breadth of one or more NGSS performance expectations.

Since our goal was to develop a CKT about matter instrument that could be used on a large scale and administered and scored efficiently, all the CKT about matter items were designed as discrete, automatically-scorable items. Across

TABLE 1 Work of teaching science framework (Mikeska et al., 2018)

Instructional tools	Examples of instructional practices
1. Scientific Instructional Goals, Big Ideas, and Topics	Choosing which science ideas or instructional activities are most closely related to a particular instructional goal
2. Scientific Resources (texts, curriculum materials, etc.)	Evaluating instructional materials for their ability to address scientific concepts; engage students with relevant phenomena; promote students' scientific thinking; and assess student progress
3. Scientific Models and Representations	Evaluating or selecting scientific models and representations that predict or explain scientific phenomena or address instructional goals
4. Student Ideas	Analyzing student ideas for common misconceptions regarding intended scientific learning
5. Scientific Language, Discourse, and Vocabulary	Anticipating scientific language and vocabulary that may be difficult for students
6. Scientific Explanations	Critiquing student-generated explanations or descriptions for their accuracy, precision, or consistency with scientific evidence
7. Scientific Investigations and Demonstrations	Selecting investigations or demonstrations that facilitate understanding of disciplinary core ideas, scientific practices, or cross-cutting concepts

the CKT about matter items, our development team used a combination of traditional multiple-choice single select items and a variety of technology-enhanced item types, including multiple-choice multiple select, grid items, inline choice items, and matching items. All items were connected to one CKT scenario. Some of the teaching scenarios in these CKT about matter items incorporated different stimuli, such as students' written work, transcripts of students' conversations, video clips, and graphics.

Figure 2 shows an example of one CKT about matter item designed to assess the CKT that elementary science teachers leverage when identifying scientific resources to use for teaching about properties of matter. This item is aligned with the NGSS content category of properties of matter and its measurements and the WOTS instructional tool category of scientific resources. Specially, this CKT item engages the test-taker, in this case a preservice teacher, in evaluating instructional materials and other resources for their ability to sufficiently address scientific concepts, which is one of the 27 critical STPs identified in the WOTS framework. It is important to note that this CKT about matter item is not designed to measure elementary teachers' understanding of a particular NGSS performance expectation. Instead, this item reflects a “micro moment” that assess one important component of the CKT that elementary teachers need to leverage when addressing a particular NGSS performance expectation, in this case 5-PS1-3, which states that students should “Make observations and measurements to identify materials based on their properties.” The instructional scenario used in this CKT about matter item was purposefully constructed to illustrate how Ms. Wu is planning this formative assessment in the context of addressing this NGSS performance expectation with her students. Each of the CKT about matter items developed is linked to one of the eight NGSS performance expectations about matter and its interactions at the K-5 level.

Ms. Wu is preparing a formative assessment for a third-grade unit on matter. She wants to find out if her students understand that matter includes things beyond objects and materials that they can see, feel, measure, and weigh. Ms. Wu locates four resources and each resource includes a list of four different examples for students to consider.
Which of the following is the most useful resource for Ms. Wu's purpose?
1. Resource A: a rock, a wooden board, a steel rod, a plastic ball 2. Resource B: shade, light, sound, heat 3. Resource C: takes up space, has weight, is visible, has color 4. Resource D: rock, dust, ant, air
Rationale for Item: *In this item, the opening scenario informs the teacher that there is a third-grade teacher (Ms. Wu) who wants to develop a formative assessment to determine if her students understand that matter includes things beyond objects and materials that they can see, feel, measure, and weigh. The teacher in this scenario has located four different resources – each one gives four different examples for students to consider – and now must select the resource that is most useful for addressing this purpose. To answer correctly, the teacher must draw upon his or her knowledge about what matter is, knowledge about student difficulties with matter concepts, and knowledge about how examples can be used to address specific purposes. A teacher who has this CKT would be able to identify that all the examples in Resource D are examples of matter because they all have mass and take up space, yet some of them are ones that students may find difficult to recognize as matter, such as gases or matter that is too small to see with our eyes, which addresses Ms. Wu's instructional purpose.*

FIGURE 2 Content knowledge for teaching (CKT) matter item example with accompanying item rationale

In this CKT about matter item, a teacher (Ms. Wu) is preparing a formative assessment for her third-grade students to determine if they understand that matter includes things beyond objects and materials that they can see, feel, measure, and weight. In doing so, Ms. Wu located four different resources, with each one including a list of four different examples for her students to consider. The teacher who responds to this CKT about matter item must determine which of the four sets of resources—resource A, B, C, or D—is the most useful one to address Ms. Wu's purpose. To answer correctly, the teacher responding to this item must draw upon their knowledge about what matter is, knowledge about student difficulties with matter concepts (e.g., understandings about how students may measure and describe properties of matter), and knowledge about how examples can be best used to address specific purposes. For this item, a teacher who has this CKT would be able to identify that all the examples in Resource D are examples of matter because they all have mass and take up space, yet some of them are ones that students may find difficult to recognize as matter, such as gases or matter that is too small to see with our eyes, which addresses Ms. Wu's instructional purpose. Thus, this item assesses CKT that elementary science teachers need to leverage when determining which resources to use to elicit students' initial ideas about types of and properties of matter, which is an important component of an instructional sequence that would be useful for addressing the previously mentioned NGSS performance expectation.

Each of the CKT about matter items our team developed are set up in a similar fashion—with an opening scenario, question, and set of options to select or respond to—although not all are multiple-choice single selection item formats. Examples of CKT about matter items using different item formats and addressing different STPs within the seven WOTS instructional tool categories and five matter topics can be found at https://cktscience.org/assessments/example-items/. As part of the team's development process, which included examination of the items by preservice and in-service elementary teachers and science content experts, we explored the cognitive process used by study participants as they each responded to a set of 8–10 CKT about matter items (the total study included 118 different CKT about matter items). For those who showed evidence of difficulty in responding to specific CKT about matter items or justifying their choices, we further examined their reasoning and the nature of the content challenges and the content teaching challenges they encountered.

3.2 | Sample

This study's sample included 79 preservice elementary teachers (PSETs) and seven in-service elementary science teachers who participated in the cognitive interviews. For the PSET study participants, we recruited PSETs from different higher education institutions, such as research universities, regional universities, and small liberal arts colleges across the United States, and made efforts to recruit PSETs from various geographic settings (i.e., institutions located in urban, suburban, and rural areas). Our team disseminated a recruitment flyer through professional associations' mailing lists that are commonly used by science teacher educators and researchers in elementary science education such as the Association for Science Teacher Education and the National Association for Research in Science Teaching. Furthermore, we actively disseminated the recruitment flyer through the research team members' contact lists. One hundred-twenty PSETs expressed interest in participating in the study and responded to an online survey and provided consent for participation. We used two criteria to initially screen the PSETs: (1) being currently enrolled in a teacher preparation program to teach elementary education and (2) being certified or in the process of certification to teach science in the elementary grades. We selected the pool of 79 PSETs for the cognitive interviews based on their time availability to participate in the interviews and aimed for selecting participants with diverse characteristics and backgrounds. Table 2 provides demographic information of the participant PSETs. Similar to the population of elementary teachers in the United States (Banilower et al., 2018), the majority of the participants self-identified as female (96%) and White (78%).

Based on the recommendation of the project's advisory board, we conducted 10 interviews with seven in-service teachers with relevant background on elementary science teaching (three teachers were interviewed twice

TABLE 2 Preservice teacher background characteristics

Demographic variable	Preservice teacher participants *n* = 79 Frequency (%)
Gender	
Female	76 (96)
Male	3 (4)
Ethnicity	
Asian or Asian American	3 (4)
Hispanic or Latino/a	6 (7)
White or Caucasian	61 (78)
Multiracial (more than one category)	3 (4)
No response/prefer not answer	6 (7)
Undergraduate major	
Elementary education	46 (58)
General science	2 (3)
Dual degree: elementary education and science	2 (3)
Dual degree: elementary education and other programs[a]	6 (7)
Early childhood education	8 (10)
Special education	3 (4)
Other[b]	10 (12)
No response	2 (3)
Undergraduate program year	
Freshman/sophomore	3 (4)
Junior	14 (18)
Senior	46 (58)
Graduate/fifth year	11 (14)
No response	5 (6)
Undergraduate program location	
Northeast	15 (19)
Midwest	14 (18)
South	16 (20)
West	32 (40)
No response	2 (3)
Undergraduate program setting	
Urban	9 (11)
Rural	2 (3)

TABLE 2 (Continued)

Demographic variable	Preservice teacher participants n = 79 Frequency (%)
Suburban	66 (83)
No response	2 (3)

[a]Includes programs such as psychology, special education, Spanish, and finance.

[b]Includes programs such as psychology, political communications, philosophy, politics and law, Middle East studies, language, literacy and culture, human development, and education policy.

using two different interview forms). The goal of the in-service teacher interviews was to provide complementary evidence to support the validity argument for the CKT items being developed, regarding content specifications and perceptions of importance and appropriateness for elementary teaching. Comparing responses across the PSETs and in-service teachers was not a focus of this study. We recruited experienced in-service teachers with strong experience in elementary science teaching about matter and its interactions, as noted by their relevant and outstanding training and professional development experiences in this area. For example, two teachers participated in a 2-week practicum-based summer institute focused on learning content and pedagogy about matter and its interactions, one teacher participated in an elementary science leadership group that consisted of teachers from multiple districts, and one teacher participated in an NGSS writing team for elementary levels. Five in-service teachers were female and two were male and their teaching experience in elementary science ranged from six to 24 years.

3.3 | Data collection

Similar to an earlier study (Howell et al., 2013), our team conducted cognitive interviews using concurrent think aloud methodology (Ericsson & Simon, 1984) to examine how participants responded to individual CKT about matter items. This methodology provided a mechanism to elicit participants' knowledge and reasoning as they responded to a subset of the CKT about matter items (8–10 CKT items per interview) and provided a window into the decision processes they made and the CKT they leveraged as they worked through each item. After responding to each CKT about matter item and thinking aloud to explain the decisions and knowledge they were using to do so, a member of our research team asked follow-up probing questions to elicit additional feedback about the item's content, including whether it was clear, comprehensible, and connected to an elementary STP and, if not, the reasons for any concerns. For each item, our goal was to gather empirical evidence on the CKT they leveraged—or did not leverage—when responding to the item, why they selected and eliminated particular answer options, and what challenges, if at all, they encountered when responding to the item.

A structured interview protocol was used with each participant. The protocol included an initial section that described the key interview activities, the interview goals, an explanation of the think aloud approach, expectations for the participants, and participant consent. The second section of the interview included two practice CKT items focused on a different science topic to help participants become familiar with the think aloud approach. The interviewer used the first practice CKT item to model the think aloud approach and the second CKT item served as a practice one for the participant. The third section of the interview protocol included 8–10 CKT about matter items. Participants were asked to read aloud each item and provide their responses and thoughts while responding to the item. After each item, participants responded to a set of questions aimed to gather further evidence of participant thinking. For example, the interviewer asked about the perceived goal of the item, item clarity issues that were noticeable for participants, and, if needed, prompted for further explanations about the rationale for the participant's answer selection. We also included: (a) one or two item specific questions that explored participants' CKT in relation to the item scenario, (b) a question that asked participants to recall a connection, if at all, between the item scenario and their own teaching experience (or hypothetically, for other teachers), and (c) a question about the perceived importance of the item scenario for the work of elementary science teachers. Findings related to PSETs' perceptions of these two latter interview questions can be found in Cisterna et al. (2022). Most interviews ranged between 90 and 120 min and all interviews used one of the interview forms (8–10 distinct CKT about matter items per form) and were audio recorded.

During each interview, the research team member conducting the interview recorded observational notes based on the participant's think aloud responses and answers to the follow-up probing questions. After each interview, the research team member completed a written summary summarizing the CKT used to respond to each item, the participant's perception of the importance of the CKT about matter item and connection to elementary science teaching, and the challenges the participant experienced when responding to the item. In addition, each interview was transcribed, and the interview transcript was imported into a qualitative data analysis program, Dedoose, for analysis purposes. Table 3 indicates the total number of interviews conducted and CKT about matter items used per item batch[1] and across participants.

3.4 | Data analysis

To answer research question (RQ) 1, which focused on the extent to which participants used the intended knowledge and reasoning when responding to these CKT about matter items, our team first developed rationales—or what we called “minimal conformity rules”—for each item detailing the specific CKT about matter that we hypothesized the participant needed to leverage when responding accurately to the item. These minimal conformity rules are like claims that one makes about the specific CKT that each item is measuring. For example, for the CKT about matter item shown in Figure 2, in their think-aloud response a participant answering correctly (selecting option D) should indicate the following: (a) Resource D includes examples of matter because they all have mass and take up space, (b) that students would have difficulty recognizing gases or matter that is too small to see with our eyes as examples of matter, and (c) understand that “usefulness” means addressing Ms. Wu's instructional purpose for formative assessment of students' ideas about what matter is. These three components—knowledge about what matter is, knowledge about student difficulties with matter concepts, and knowledge about how examples can be best used to address specific purposes—collectively comprise the CKT about matter that a participant who correctly

TABLE 3 Content knowledge for teaching (CKT) about matter item interviews by batch and forms

			Participants	
Interview batch	CKT items	Number of forms	Preservice teachers	In-service teachers
Batch 1	24	3	17	0
Batch 2	36	4	24	4
Batch 3	33	4	24	4
Batch 4	25	3	18	2

Note: Our team developed the CKT about matter items in batches for feasibility purposes, as many of our external consultants were full time science teachers or science educational faculty members or researchers. Each batch focused on one or more of the five matter topics (e.g., conservation of matter, changes in matter), although the last batch cut across topic areas. Before these interviews, some CKT about matter items were dropped from further use due to flaws noted in earlier team reviews. Overall, our team initially drafted over 200 CKT about matter items, with 118 of those items proceeding to the cognitive interview and expert content review stage of the item development process.

selects the intended answer key (option D) should have access to and use when responding to this CKT about matter item.

Our team developed these minimal conformity rules, or claims, for each of the 118 CKT about matter items in this study. If an item had multiple answer keys (e.g., multiple-choice multiple select items could have two or three answer keys), then we developed a minimal conformity rule for each answer key. These minimal conformity rules specified the CKT about matter that each item (or answer key) was hypothesized to assess. If a participant selected the intended answer, then we would expect that their think-aloud response would indicate that their reasoning met the minimal conformity rule. However, if a participant selected an inaccurate answer, then we anticipated that their think-aloud response would show one or more gaps in their CKT about matter and would not fully meet the minimal conformity rule.

Each participant's response to each CKT about matter item (total of 1371 individual, item-level responses across the 118 CKT about matter items and 86 participants) was coded in multiple ways. First, participants' answers to all items were coded as correct, incorrect, or other answer. If the participants selected the intended responses(s), as identified in the item metadata, then the response(s) was coded as correct. If their selection did not match the intended answer, then it was coded as "incorrect." If the participant decided not to answer the item or if multiple answers were selected (even if one of the answers was correct), then it was coded as "other answer." If the participant mentioned having difficulty selecting the correct answer, then an "expressed uncertainty" code was also applied to the response.

Second, participants' responses were also coded for whether they met the item's minimal conformity rule(s). Participant responses were coded as "reasoning/justification conformed to rationale" if they met all the criteria in the minimal conformity rule that was specified for the item (or answer choice) and the answer was accurate. If the participant response did not meet all the criteria in the minimal conformity rule for the item (or answer choice), then their response received the "reasoning/justification diverges from the rationale" code. Additional aspects of their reasoning were also noted, such as if they had "incorrect content present with minimal conformity" (e.g., the respondents had misunderstandings about the science content present in the think aloud) or if they "changed earlier answer to conform" (e.g., the respondents changed their earlier answers so that it is accurate). We calculated the total number and percentage of participant responses in each of four categories: (1) responses that indicated the correct answer and aligned with the minimal conformity rule (conformed to rationale), (2) responses that indicated an incorrect answer and did not align with the minimal conformity rule (diverged from rationale), (3) responses that indicated a correct answer, but did not align with the minimal conformity rule, and (4) responses that indicated an incorrect answer, but did align with the minimal conformity rule. Responses in the first two categories (correct and met minimal conformity rule; incorrect and did not meet minimal conformity rule) suggested that the CKT about matter item was functioning as hypothesized. That is, participants who illustrated evidence of using the relevant CKT about matter did respond accurately and those who showed gaps in their CKT about matter for a specific item (or answer choice) failed to respond accurately. Alternatively, the final two categories (correct and did not meet minimal conformity rule; incorrect and met minimal conformity rule) potentially indicated a flaw in the CKT about matter item or its rationale for the hypothesized CKT about matter being measured.

The CKT about matter cognitive interview team met every week to discuss the processes of conducting interviews, data reduction, and coding. During this endeavor, we conducted several procedures to ensure quality data reduction and coding. First, we implemented training sessions on the coding protocol and the coding procedures. Second, when team members found issues regarding coding, these were discussed during the weekly meetings. Third, after coding each interview batch, one team member reviewed the coding in Dedoose, an online qualitative data analysis program, to ensure that all the codes were entered in the qualitative analysis software and memos had been written for specific code applications.

To achieve a shared understanding of the coding procedures, one team member implemented three sessions of coding calibration and reconciliation. In those sessions, team members coded the same set of participant responses—usually two CKT items per session and three participant responses per CKT item. Team members coded the responses individually and, during the calibration meetings, explained the rationale for their coding, so the team was able to find a shared and common coding for those responses. Overall, this process helped our team members develop a better understanding of the coding protocol and how to code interview responses consistently. It also provided team members with conceptual and analytical tools for the analysis. While coming to consensus during these coding exercises helped to maintain the team's shared understanding of the coding scheme, we recognize that calculating interrater reliability would have been advantageous to include, although we did not use that approach in this study.

To answer RQ 2, for any participant responses that did not meet the item-specific (or answer-specific) minimal conformity rule, their responses were also coded for the reason their answer justification diverged from the rationale. There was a total of eight possible codes that could be applied as the reason why the participant response did not meet the minimal conformity rule. Each code identified a specific challenge that these participants illustrated when responding to the CKT about matter items and were adapted from previous research examining teachers' challenges responding to CKT assessment items (Howell et al., 2013). One or more codes could be applied as there could be more than one reason a participant's response did not meet the minimal conformity rule for the specific CKT about matter item (or answer choice).

Two codes—“incorrect content” and “explicit guessing”—were linked to challenges with the content. The code “incorrect content” was applied to capture participant responses that diverged from the rationale because it was clearly demonstrated in their think aloud that the participant showed gaps in their understanding of the subject matter that was required to answer the item. The code “explicit guessing” was applied to capture participant responses that diverged from the rationale because the participant explicitly guessed their answer to the item due to limited subject matter knowledge.

Two codes—“works on a different science teaching practice/answering a different question” and “does not attend to a critical aspect of the item”—were linked to challenges with aspects of content teaching. The code “works on a different science teaching practice/answering a different question” was applied to capture participant responses that diverged from the rationale because the participant's interpretation of the item caused them to work on a different STP than the intended STP represented in the item, as noted by the item writer in the item metadata. Participants could have worked on a different STP because they ignored or overlooked it or redefined the STP in the sense that they worked on a slightly different task than what the item intended. The code “does not attend to a critical aspect of the item” was applied to capture participant responses that diverged from the rationale because they did not attend to key information about aspects of the teaching scenario directly stated or implied in the item prompt.

Sometimes participants did not provide sufficient reasoning in their think-aloud responses to understand the full range of CKT they were leveraging in their item response. Two codes related to insufficient reasoning evidence. The code “reasoning through contrast with non-selected” was applied to capture participant responses that diverged from the rationale because their answer justification showed evidence of reasoning using the process of elimination. In these instances, the participant provided clear reasoning and justification for why they eliminated nonselected options but did not provide any clear justification for why the selected option(s) is the answer. The code “justification is not a justification” was applied to capture participant responses that diverged from the rationale because they provided an answer in their think aloud without providing any reasoning or explanation about why they selected an answer or eliminated other options.

Finally, the code “misreading the item” was applied to capture participant responses that diverged from the rationale because there was clear evidence in their think aloud that they misread a word or symbol in the item (e.g., reading a division symbol as a plus sign). This is not to be confused for misunderstanding the item. In addition, sometimes participants included arguments that were compelling and defensible, but did not include all components of the minimal conformity rule. In the first instance, the participant's response was correct, but the justification did not include all components of the minimal conformity rule and included no evidence of incorrect content knowledge. In the second instance, the participant's response was incorrect, but their answer justification included some of the components of the minimal conformity rule as well as other elements that were compelling and defensible. The code "defensible non-rationale argument" was used to capture these types of participant responses and a memo was applied to the response to describe which parts of the minimal conformity rule was missing or which elements of their reasoning were compelling and defensible. Usually responses coded with "defensible non-rationale argument" meant that the item was not functioning as hypothesized and the item or the item rationale needed to be refined to ensure that the identified CKT about matter needed to be used to respond accurately to the item. After coding participants' responses for these challenges, we then calculated the total number of responses in each coding category and used the descriptive frequencies within and across categories to discern patterns in the reasons participants did not meet the minimal conformity rule.

To address RQ 2a and 2b, we focused on participants' responses that illustrated specific challenges—either where the response evidenced incorrect content knowledge (first coding category) or where they exhibited challenges associated with aspects of content teaching, in particular, challenges with addressing critical aspects of the science teaching scenario represented in the item (fourth coding category). We selected these two specific challenges as these two reasons were the ones most frequently represented in the dataset for why participants' responses did not meet the minimal conformity rule. Since the number of participants that experienced these challenges was variable across items, we decided to focus our analysis on those CKT items in which more than 50% of participants exhibited responses associated with one or both content or content teaching challenges. This decision led us to narrow down the pool to 172 responses for analyses. Three researchers examined the content challenges associated with 34 CKT about matter items[2] (127 participant responses) and the content teaching challenges associated with 10 CKT about matter items (45 responses) to describe the nature and type of respective content and content teaching challenges these participants showed when responding to these CKT about matter items. We conducted a thematic analysis to document and identify patterns in the nature of the challenges (Braun & Clarke, 2006). This analysis involved generating initial codes and then searching for, reviewing, and defining themes based on this coding. To do so, we iteratively examined the types of challenges through successive rounds of revisions to refine the category and generate descriptions for these science content and science teaching challenges. We organized the findings according to these themes, which served to illustrate the nature and extent of these challenges. Table 4 provides the themes and descriptions for the science content and science content teaching challenges observed in the participants' responses.

4 | RESULTS

4.1 | RQ 1: Use of intended knowledge and reasoning when responding to CKT about matter items

Table 5 shows the extent to which the study participants' think aloud responses indicated the correct or incorrect answer when responding to these CKT about matter items and the extent to which their responses did or did not conform to the item rationale, as detailed in the minimal conformity rule for each item (or answer choice, if the item had multiple correct answer choices). Of the 1371 individual item-level responses analyzed across the 86 study participants and 118 CKT about matter items, 80.8% (n = 1108 responses) of them provided evidence that the CKT about matter item was functioning as intended. That is, 56.7% (n = 777) of participant responses provided the correct answer and showed evidence of using the relevant CKT about matter in their think-aloud response, as hypothesized in the item rationale. Similarly, 24.1% (n = 331) of participant responses selected an incorrect answer when responding to the CKT about matter items and they did not show they had access to and could use the relevant CKT about matter in their think aloud response; this suggested they had a gap in their CKT about matter and, as hypothesized, they were unable to respond accurately to the CKT about matter item. Both kinds of responses—correct answer and reasoning that conformed to the item rationale or incorrect answer and reasoning

TABLE 4 Themes and descriptions of science content and science content teaching challenges

Type of challenge	Theme	Description
Science content	Understanding of science concepts	Challenges around reasoning about and applying science concepts
	Identifying examples	Lack of familiarity with examples commonly used in science teaching
	Leveraging evidence in reasoning	Challenges with using appropriate scientific evidence, particularly observations or data, to support scientific reasoning
	Understanding of vocabulary	Uncertainty in the meaning of specific scientific terminology
Science content teaching	Attending to the goals set forth in the items	Challenges with determining how scenario content relates to specific instructional goals
	Identifying connections with scientific content	Uncertainty in recognizing how a specific instructional situation or task represented particular science concepts
	Recognizing grade-level appropriateness of the content or task	Challenges with identifying activities, content, or vocabulary appropriate for students' grade-level

TABLE 5 Participants' use of intended reasoning and knowledge when responding to content knowledge for teaching (CKT) about matter items

		Reasoning conforms to rationale	
		Yes	No
Item batch	Answer accuracy	*n* (%)	*n* (%)
Batch 1	Correct	128 (57.9)	41 (18.6)
	Incorrect	4 (1.8)	48 (21.7)
Batch 2	Correct	214 (53.4)	65 (16.2)
	Incorrect	6 (1.5)	116 (28.9)
Batch 3	Correct	240 (53.5)	85 (18.9)
	Incorrect	5 (1.1)	119 (26.5)
Batch 4	Correct	195 (65)	50 (16.7)
	Incorrect	7 (2.3)	48 (16)
Overall	Correct	777 (56.7)	241 (17.6)
	Incorrect	22 (1.6)	331 (24.1)

Note: *n* refers to the number of participant responses. Between four to six participants responded to each CKT about matter item. If a CKT about matter item had multiple answer keys (e.g., multiple choice multiple selection items could have two or three answer keys), then each answer key counted as a separate "response" for coding purposes to determine if the participant answered correctly and provided a reasoning that did or did not conform to the expected rationale.

that did not conform to the item rationale—provide construct validity evidence that these scenario-based CKT about matter items are assessing the relevant CKT in this topic area. This overall pattern of most participants answering correctly who used the intended CKT and those answering incorrectly showing gaps in their CKT was consistent within and across the four CKT item batches.

4.2 | RQ 2: Nature of challenges when responding to CKT about matter items

Across the 1371 individual item-level participant responses, slightly more than a third of them (41.7%, *n* = 572 responses) did not conform to the item rationale. That is, participants' think-aloud responses showed gaps in their CKT about matter and, therefore, their response was coded as not meeting the minimal conformity rule for that item (or answer choice, if the CKT about matter item had multiple correct answer choices). As shown in Table 6, many responses that did not conform to the item rationale provided evidence in their think aloud of a defensible justification, suggesting that revisions to the CKT about matter item or rationale would be warranted before future use for a subset of these CKT about matter items. Typically results showed that participants would answer the CKT about matter item correctly but did not address all aspects of the item's minimal conformity rule. Usually this meant that the CKT about matter item was measuring a narrower swath of CKT about matter than originally hypothesized, which led to revisions in the item rationale.

Most importantly, findings showed that when these elementary teachers did not provide the intended knowledge and reasoning in their think-aloud responses, they were most likely to encounter challenges in two key areas. The first major area involved difficulty with using the required science content to respond accurately and adequately to the CKT about matter item (231 of 572 responses that did not conform to the rationale showed evidence of incorrect content). The second major area involved difficulty with content teaching aspects, particularly challenges in attending to a critical science teaching aspect required to respond appropriately to the CKT about matter item (103 of 572 responses that did not conform to the rationale did not attend to a critical aspect of the CKT about matter item).

4.2.1 | Nature of the content challenges

For science content challenges (RQ 2a), the main themes that emerged from participant responses were difficulty in: (a) understanding concepts, (b) identifying examples, (c) leveraging evidence in reasoning, and (d) understanding

TABLE 6 Nature of challenges participants encountered when responding to content knowledge for teaching (CKT) about matter items

	Coding category	Batch 1 *n* (%)	Batch 2 *n* (%)	Batch 3 *n* (%)	Batch 4 *n* (%)	Overall *n* (%)
Challenges with content	Incorrect content	36 (40.4)	91 (50.3)	80 (39.2)	24 (24.5)	231 (40.4)
	Explicit guessing	1 (<1)	1 (<1)	6 (2.9)	0 (0)	8 (1.4)
Challenges with content teaching	Works on different task	6 (6.7)	2 (<1)	3 (1.5)	0 (0)	11 (1.9)
	Does not attend to critical aspect	29 (32.6)	22 (12.2)	32 (15.7)	20 (20.4)	103 (18.0)
Challenges with providing sufficient reasoning	Reasoning through contrast with nonselected items	9 (10.1)	8 (4.4)	4 (2.0)	2 (2.0)	23 (4.0)
	Justification is not a justification	6 (6.7)	12 (6.6)	25 (12.3)	17 (17.3)	60 (10.5)
Error	Misreading the item	0 (0)	2 (<1)	5 (2.5)	0 (0)	7 (1.2)
Suggests revisions to items	Defensible justification	19 (21.3)	87 (48.1)	71 (34.8)	53 (54.1)	230 (40.2)

Note: *n* refers to the number of participant responses that did not meet the minimal conformity rule and showed evidence of the specific challenge. The total *n* in each column is based on the number of participant responses that did not meet the minimal conformity rule: 89 responses (Batch 1), 181 responses (Batch 2), 204 responses (Batch 3), 98 responses (Batch 4), and 572 responses (overall). The percentages in each column do not add to 100% because participant responses may have shown multiple challenges when they did not meet the minimal conformity rule.

scientific vocabulary. Table 7 shows the percentage of responses for each of these themes, sample participant responses, and the nature of the incorrect science content. Most of the participants' responses related to challenges with understanding science concepts (59.8%), 22.8% of responses were about challenges with identifying examples, 18.1% of responses were about challenges with leveraging evidence in reasoning, and 5.5% were about challenges related to science vocabulary.

4.2.1.1 | *Understanding of science concepts*

Given that most of the responses in the challenges with content category were coded as challenges with understanding science concepts, all the examples that met our criteria (as outlined in the methods section), are further broken down in Table 8. Thus, Table 8 represents the 59.8% of the responses from Table 7 that are coded as "understanding of science concepts," and organizes them by the following dimensions: matter topic area, item context, participants' responses to the associated item, and nature of the incorrect content.

One CKT about matter item asked participants to reason about the outcome of an investigation in which the ingredients of a cake are mixed and then heat is added. Specifically, they were asked to provide the rationale that explains that the formation of a new substance (the cake) is due to a chemical change. In their response, a participant reasoned, "I was thinking that the molecules of each specific [ingredient] aren't changing, but mixing with others, so I thought that I would want my students to know that they're just getting a new substance and not a new molecule" (Item 108, Participant 32). While the participant acknowledges that a new substance is formed, they do not mention that the formation of a new substance is the result of atoms rearranging to form a new molecule. Thus, the participant did not make the connection that molecular changes (e.g., the microscopic account) are associated with the formation of a new substance (e.g., the macroscopic).

In another item, participants were asked to reason about students' models that reflect continuous or particulate views of matter. In their reasoning, the participants expressed varying degrees of understanding of these two views. One participant says, "So the continuous views the one where they have a good understanding of matter and then the particulate view would be the students who just at that moment may have come up with something. Or maybe they've always had it and it's just not correct" (Item 135, Participant 47). Thus, this participant's understanding of continuous and particulate are not about the nature of matter, rather they are about whether the student's understanding is complete (continuous) or not (particulate). Analogously, another participant states, "…they hold a continuous view of matter. So maybe that means they know that this exists for everything rather than just one thing. I think that's what that means" (Item 135, Participant 48). This participant is interpreting continuous to mean a consistent understanding across all concepts. Another participant states that, "I'm thinking that a particulate view would be like that like it has a bunch of different parts." Thus, this participant thinks that the particulate view means that something has many different parts to it without referring to the particulate composition of matter. Thus, these participants are using everyday understandings of continuous and particulate in their reasoning about this item, and using them in ways that differ from canonical ideas about matter.

In another item, participants were asked to reason about how to demonstrate conservation of matter in a task in which students add ice to water in a cup, and then graph the changes in weight over time. The participants each had different reasonings as to why this would not support conservation of matter principles. One preservice teacher was confused about why one would add substances from two different phases, saying "That's a little confusing because you're adding a solid to a liquid. So, you're mixing two different states in one state and it's melting. So that's kind of hard to see, you have to measure the water, I don't think that's a direct approach, that's not a very clear approach to seeing it." Thus, this participant did not understand the purpose of mixing the solid and liquid, and furthermore, was not sure how it would support in developing students' understanding of the conservation of matter. Other participants did not understand the purpose of measuring the weight changes, saying "I didn't really understand the whole change in weight versus time if the ice cube melts." And another said "I'm not really sure because I'm not really sure that you're doing weight over time. I guess I don't know enough about that to know that." Another participant thought that the weight would change over time as the ice melts, and thus, thought

TABLE 7 Science content challenges

Science content challenges	Percent of responses (out of $n = 127$)	Item alignment	Sample participant response	Nature of incorrect content
Understanding of science concepts	59.8%	Content Topic: Model of matter WOTS Teaching Practice: Critiquing student generated explanations and descriptions for their generalizability, accuracy, precision, or consistency with scientific evidence	"I don't know if particles expand when they turn into solid." (Item 124, Participant 57)	Unsure if particles change as they undergo a phase change
Identifying examples	22.8%	Content Topic: Changes in matter WOTS Teaching Practice: Choosing resources that support the selection of accurate, valid and age-appropriate goals for science learning	"I don't necessarily know what happens when you mix iron filings with baking soda." (Item 96, Participant 30)	Unsure of the outcome of mixing two substances
Leveraging evidence in reasoning	18.1%	Content Topic: Changes in matter WOTS Teaching Practice: Critiquing student generated explanations and descriptions for their generalizability, accuracy, precision, or consistency with scientific evidence	"So I'm wondering, my first instinct is that this might be an incorrect one, just because bubbles, at least from what I remember from science, don't necessarily mean that the mixture is changing." (Item 71, Participant 39)	Unable to use evidence of the formation of gas as evidence of a chemical change
Understanding of vocabulary	5.5%	Content Topic: Scientific instructional goals, big ideas, and topics WOTS Teaching Practice: Linking science ideas to one another and to particular activities, models, and representations within and across lessons	"I really don't know what the word 'precipitate' means." (Item 92, Participant 25)	Unsure of specific science vocabulary

Note: Each participant response can receive one or more of the "science content challenges" codes.

TABLE 8 Examples of challenges with understanding of science concepts

Matter topic area	Item context	Item alignment	Expected justification	Sample participant response	Nature of incorrect content
Properties of matter	Conducting an investigation to determine whether a substance is formed when water and salt are mixed and the mixture is observed after a few days	Content Topic: Properties of matter WOTS Teaching Practice: Engaging students in using, modifying, creating, and critiquing scientific models and representations that are matched to instructional goals	Mixing water and salt together does not form a new substance; after a few days, the water evaporates and leaves the salt behind	"That this is going to be the A, different substance and that they actually did do something new, having the crystals becoming, having the salt then turn into crystals, they're going to weigh more because it solidified into a different mass… It would definitely help them because they're going to see that the weight has changed, and a reaction has happened." (Item 200, Participant 67)	Participant thinks that mixing salt and water together results in a chemical reaction and the formation of a new substance.
Changes in matter	Conducting an investigation to determine whether the mixing of two or more substances results in a new substance	Content Topic: Changes in matter WOTS Teaching Practice: Selecting scientific language that is precise, accurate, and grade level appropriate	After mixing the ingredients of a cake mix and adding heat, a new substance was created (the cake). This new substance is a chemical change.	"I was thinking that the molecules of each specific aren't changing, but mixing with others, so I thought that I would want my students to know that they're just getting a new substance and not a new molecule." (Item 108, Participant 32)	Participant does not understand that a new substance is the result of a molecular change, and the new substance is the product of the mixing.
Changes in matter	Identifying examples that involve changes in states of matter	Content Topic: Changes in matter WOTS Teaching Practice: Selecting diagnostic items and eliciting student thinking about scientific ideas and	Reasoning that the following observation is a change in states of matter: "drops of water on the outside of the glass"	"…it's already water and I don't know…if it's the ice melting that's causing the drops of water on the outside of the glass so I'm gonna put that	Participant is unsure of how water condenses on the outside of the glass.

TABLE 8 (Continued)

Matter topic area	Item context	Item alignment	Expected justification	Sample participant response	Nature of incorrect content
		practices to identify common misconceptions and the basis for those misconceptions		one aside for now." (Item 113, Participant 42)	
Conservation of matter	Providing evidence (through measuring and graphing) that weight is conserved when heating substances	Content Topic: Conservation of matter WOTS Teaching Practice: 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	The weight and the amount of matter will not change as the ice cube melts.	"Allow students to add an ice cube to a cup of water, and graph the change in weight versus time as the ice cube melts… That one would be more measuring how the weight changes over time, and we're looking to see conservation of matter." (Item 150, Participant 49)	Participant thinks that the ice melting in water over time will lead to an increase in weight, thus not understanding the conservation of matter in the context of a phase change.
Model of matter	Identifying which students hold a continuous view of matter rather than a particulate view based on students' models	Content Topic: Model of matter WOTS Teaching Practice: Critiquing student generated explanations and descriptions for their generalizability, accuracy, precision, or consistency with scientific evidence	Identifying aspects of students' models that indicate a continuous view of matter (e.g., there is always something in between particles)	"So the continuous views the one where they have a good understanding of matter and then the particulate view would be the students who just at that moment may have come up with something. Or maybe they've always had it and it's just not correct." (Item 135, Participant 47)	Participant thinks that a continuous view of matter constitutes a good understanding rather than what matter looks like on a microscopic scale.
Model of matter	Critiquing models of particles too small to be seen	Content Topic: Model of matter	The particles of matter do not change their size	"Then as the butter started to cool down, the particles expanded like you see in	Participant thinks that particles of matter might expand when a substance is

(Continues)

TABLE 8 (Continued)

Matter topic area	Item context	Item alignment	Expected justification	Sample participant response	Nature of incorrect content
		WOTS Teaching Practice: Critiquing student generated explanations and descriptions for their generalizability, accuracy, precision, or consistency with scientific evidence	when a substance is solidified.	Anika's model. They were closer together in the dish. That's the part I'm a little confused about, because I don't know if particles expand when they turn into solid." (Item 124, Participant 57)	melted, instead of increasing their movement and separation.
Properties of matter	Conducting investigations to determine the chemical identity of three jars of unlabeled white powders (baking soda, salt, and sugar)	Content Topic: Properties of matter and its measurements WOTS Teaching Practice: Selecting investigations or demonstrations that facilitate understanding of disciplinary core ideas, scientific practices, or crosscutting concepts	Conducting multiple tests allows one to determine the identities of the substances	"Yeah, all three powders are going to have different shapes and factors underneath the magnifying glass I think." (Item 51, Participant 22)	Participant thinks that the three white powders can be differentiated by putting them under a microscope and looking at their respective shapes.
Properties of matter	Modeling the structure and properties of matter using data by selecting a representation to illustrate the properties of hardness/softness	Content Topic: Properties of matter and its measurements WOTS Teaching Practice: Evaluating or selecting scientific models and representations that predict or explain scientific phenomena or address instructional goals	Hardness is a continuous property (a spectrum). Therefore, a continuum most accurately represents the hardness property and students could use the continuum to sort items from hardest to softest.	"I don't think this would be a strong way to do it because I think that something can be either hard or soft, it can't be both. A T-chart with hard on one side and soft on the other, I think is a strong way to do it because like I said, it's either one or the other. It can't be both." (Item 42, Participant 13)	Participant thinks that hard and soft are discrete categories, so using a T-chart would be a good representation for sorting items for the property of hardness.

that this task would not contribute to students' understanding of the conservation of matter. The participant noted, "That one would be more measuring how the weight changes over time, and we're looking to see conservation of matter." In this item, the participants' reasoning are insufficient to explain: (a) why ice and water are being added together, (b) whether the weight changes over time, or (c) how this example illustrates the conservation of matter in the context of phase change.

4.2.1.2 | *Identifying examples*

Another content challenge evidenced in participants' responses was their lack of familiarity with examples typically used to illustrate scientific phenomena. For instance, mixing baking soda and vinegar is a chemical reaction commonly used to illustrate changes in matter in which two substances (a solid and a liquid) combine to form a new substance (a gas), although some participants were not familiar with this phenomenon. Similarly, some participants were not familiar with other examples that asked about predicting the outcomes of mixing specific substances. For example, one CKT about matter item included a scenario that asked about the outcome of mixing chocolate powder and milk. Knowledge about common examples associated with a scientific phenomenon and how they can be used instructionally to support students' learning are important aspects of CKT.

Another type of challenge in this area was related to challenges in identifying examples of materials with specific properties. One theme that was observed was not being able to identify which materials had magnetic properties. For example, not all participants were able to identify that metals such as iron, cobalt and nickel are magnetic. Therefore, in the CKT scenario that included selecting magnetic materials for a classroom activity, many participants were unable to identify which materials would be appropriate for an investigation that involved magnetic materials. For example, one participant said, "I wasn't sure about the metal. I'm thinking it's the zinc and the pennies that are not going to be magnetic. I honestly have to do it with a magnet myself before I go into a magnet lesson" (Item 104, Participant EXP 04). This lack of familiarity about which materials have magnetic properties influenced which materials participants selected for use in instructional activities or demonstrations.

In addition, some participants were not able to predict whether the mixing of specific substances would result in a new one, due to their lack of familiarity with the specific examples in the scenarios. For example, when reasoning about what would happen when mixing baking soda with vinegar, a participant said: "I don't know if I think it's going to be a great investigation just because it's kind of out on a limb as to whether a new substance is even going to form" (Item 72, Participant 30). This lack of familiarity with the mixing of specific substances was observed in other scenarios. Several participants mentioned not knowing what will happen when common substances such as vinegar and water, vinegar and oil, vinegar and milk, and iron filings and baking soda are combined. For example, one participant said, "And then the second part of that, mix iron filings and baking soda, I don't really know if that would do anything. It's possible that it would, but I'm really not sure" (Item 96, Participant 31).

One final challenge related to participants' lack of familiarity with examples is knowing the exceptions of the phase behavior trends for common substances like water. For example, when using the particulate model of matter to explain the phase behavior of water as it transitions from a liquid to a solid, participants reasoned that particles in the solid phase would be closer together than the liquid phase. However, participants did not know that this pattern of change did not apply to water molecules, as water molecules spread out upon freezing.

4.2.1.3 | *Leveraging evidence in reasoning*

When responding to some CKT about matter items, participants did not always use the appropriate scientific evidence in their reasoning. The example in Table 6 shows a sample participant's response around a CKT about matter item about which observations can be used to identify chemical changes. The participant says, "…so I'm wondering, my first instinct is that this might be an incorrect one, just because bubbles, at least from what I remember from science, don't necessarily mean that the mixture is changing" (Item 71, Participant 39). In the response, the participant can recognize that bubbles may be evidence of a new substance, but did not use this knowledge for the specific item scenario in which bubbles appear when two substances are mixed as evidence for a chemical change. Thus, this example shows that the participant was unable to identify specific evidence that is indicative of a chemical change in this situation.

In another CKT about matter item, participants are asked how they can assert that a new substance is formed when five grams of baking soda and 50 ml of vinegar are mixed together. In one of the correct options, participants are asked to reason about weighing the materials before and after the investigation.[3] Some participants did not think that weighing the material before and after the investigation would support the assertion that there is a chemical change, "because I think the weight doesn't really determine whether a new substance is formed. You could have just added more of some liquid or something so I just don't think weight is accurate support for a new substance being created" (Item 81, Participant 34).

Last, in another CKT about matter item, participants were given different scenarios in which both reversible and irreversible changes occur and they had to identify the irreversible change in the scenario. For example, one participant noted how "the wick of a birthday candle is lit. The candle burns for a few minutes, allowing students to notice the wax melting and the candle getting shorter before it is blown out." For this item, while some participants were able to claim that this example had an irreversible change, they were not able to leverage the appropriate scientific evidence in their reasoning. For example, one participant said, "melting a candle wax and cooling it down is an irreversible change (the shape changes), they can't really get back to their other shape… To me this just shows that changing, but not that it's reversible. You can't really pack that melted wax back in a candle…to get to the same shape and reburn it. That's exactly an irreversible change" (Item 202, Participant 80). Thus, in their reasoning, the participant is leveraging the change in shape of the wax as the evidence for an irreversible change, rather than the burning of the wick/wax. Thus, the participant's conception of irreversibility was about shape changes that may be difficult to undo, rather than chemical change which is due to the burning.

4.2.1.4 | *Understanding of vocabulary*

While reasoning about the CKT about matter items, some participants expressed that they were uncertain about the meaning of specific scientific terminology. In our current analysis, this only represented 5.5% of our data and occurred specifically with the term "precipitate." For example, in one CKT about matter item, it describes a teacher conducting a demonstration where she pours some vinegar into a glass of milk and the students observe that a precipitate forms in the milk. In their response, one participant says, "I'm kind of wondering what that looks like, so maybe on the outside of the milk…I really don't know what the word 'precipitate' means…" (Item 92, Participant 25). Their reasoning about the demonstration showed how the participant cannot use the term to make sense of what a precipitate could be in the context of the mixing of the vinegar and milk.

4.2.2 | Nature of the content teaching challenges

The other major challenges we observed were regarding participants' challenges with aspects of science teaching. Three themes emerged indicating difficulties with: (a) identifying connections with scientific content (31.1% of responses), (b) attending to the goals as set-forth in the items (40%), and (c) recognizing the grade-level appropriateness of the content (22.2%). Table 9 shows the percentage of responses for each of these themes and key examples of the science content teaching areas implicated for each one.

4.2.2.1 | *Identifying connections with scientific content*

A critical skill for elementary teachers is to identify the features of an instructional representation or activity and know how to use it for teaching about matter. Some participants' responses illustrated challenges related to this skill. For example, a specific item scenario described objects and materials that can be found on a beach. Participants had to identify which student responses demonstrated an understanding of the different properties of

TABLE 9 Science content teaching challenges

Science content teaching challenges	Percent of responses (out of *n* = 45)	Examples of the science content teaching areas implicated
Identifying connections with scientific content	31.1%	Identifying how specific actions are irreversible (e.g., being burnt is irreversible) Recognizing that multiple tests need to be done to identify powders
Attending to the goals as set-forth in the items	40.0%	Attending to specific misconceptions that students may have around specific science content (e.g., location is not a physical property) Attending to concepts of interest in the item scenario (e.g., not recognizing that there are fewer particles in a model than there should be)
Recognizing the grade-level appropriateness of the content or task	22.2%	Recognizing that molecular models are not appropriate for third graders Recognizing grade-level appropriateness of scientific explanations

Note: Each participant response can receive one or more of the "science content teaching challenges" codes.

matter by using them as a sorting criterion for the objects and materials. For example, one participant did not recognize that the "place" where these objects are located is not a property of matter.

Another example of this teaching challenge is related to the use of representations in the item scenario. When making sense of scientific models in the item scenarios, participants needed to identify the components of the model and recognize the relationships between the model components (concepts and processes) and the scientific phenomenon. For example, in one CKT about matter item, participants had to identify the purpose of an activity in which elementary students work with plastic pieces that can be assembled and reassembled to form different objects; this process served as an illustration of the concept that matter is made of small units that can be combined and recombined to form different substances. For this item, one participant (Item 199, Participant 79) mentioned, "They're using pieces from the same structure and kind of building smaller structures. The goal doesn't say anything about that, about taking pieces away from a structure or splitting up a structure." This response suggests that this participant did not understand how the plastic pieces in the structure could be reorganized to make different objects and, therefore, did not connect this model to how it represents the particulate nature of matter.

4.2.2.2 | *Attending to the instructional goals*

To plan and organize instructional activities with students, elementary science teachers need to be able to identify the specific science content that is embedded in the scenario and how it relates to instructional goals. For example, one item scenario asked participants to recognize which student responses had misconceptions about matter. Some participants tended to focus their attention on identifying the properties of matter described in the item scenario. For example, one participant reasoned, "I just said I feel like he wasn't even considering the weight of any of the objects which might make it so D could actually be the answer (Item 10, Participant 18)." This suggests that this participant focused on identifying weight of an object as the property of interest in the scenario, instead of evaluating the student responses in terms of the misconceptions that the students expressed.

Another example we observed was when participants failed to recognize that the number of particles is conserved in a conceptual model. In the scenario of dissolving sugar in water, the number of particles is the same before and after the dissolution. In their reasoning about this scenario, a participant (Item 156, Participant 60) reasoned, "I know that model C is the correct one, but it would be either model B or model A. I feel like, I'm thinking about both of these. Both of these, both have the particles, lots of the sugar. It's not like they just disappeared."

This participant selected the correct response, was able to identify the type of change that is described in the model, and recognized that matter particles stay in the system after the change. However, they did not mention that the number of particles of matter did not change from the dissolution, as evidence of the conservation of matter.

4.2.2.3 | *Recognizing the grade-level appropriateness of the content or task*

Our team developed items that explicitly asked participants whether the level of depth and content in the item scenario was consistent with the NGSS and developmentally appropriate for the grade level described in the scenario. For example, the NGSS recommend that upper elementary students learn that matter is composed of particles, but students are not expected to learn about the atomic and molecular structure of particles. There were several examples of participants not recognizing the grade-level appropriateness in this content area. For example, some participants did not recognize that a simulation that shows the behavior of particles in a chemical reaction is not an appropriate activity for third grade students, given that the particles of matter were represented at the molecular level. Other challenges involved failing to recognize that specific vocabulary is not appropriate for the grade-level or that features in models of matter are not appropriate for elementary students. For example, one participant discussed the characteristics of a chemical reaction without recognizing that the substances involved are beyond the expectations of the grade level. This participant (Item 154, Participant 45) mentioned how the following reaction would be helpful for elementary students to learn about conservation of matter: "You mix the two together and it forms a white precipitate of barium sulfate, and the mass stays the same following the reaction. For that one I think that could still be a good option because it's showing the students that even when you mix it it's the same weight afterwards."

5 | DISCUSSION

Findings from this study suggest important considerations and implications in the following areas: (a) using CKT science items as part of formative and summative assessment systems and (b) identifying areas to target in elementary science teachers' CKT development. We discuss each one in turn.

5.1 | Using CKT science items for varied assessment purposes

Our findings showed that over 80% of participant interview responses indicated that the CKT about matter items functioned as hypothesized. We contend that these findings indicate strong support, in terms of quality and rigor, for the CKT about matter items' construct validity and for the use of these items on future CKT assessments and in studies of science teachers' CKT. This result is an important one as this study illustrates one of the few attempts to create assessment items that can adequately measure both aspects of elementary science teachers' CKT—their subject matter knowledge and their pedagogical content knowledge—in an integrated way and using an approach that can be more efficiently delivered and scored on a large scale. Ensuring that the majority of the developed CKT about matter items do, in fact, meet the hypothesized assessment intent and justification suggests that the items are functioning as intended and adequately assessing the teachers' CKT in this science area.

As noted in previous literature (Bertram & Loughran, 2012; Chen et al., 2020; Henze & van Driel, 2015; Park & Suh, 2015; Roth et al., 2011; Sadler et al., 2013; Smith, 2010), the science education assessment landscape is filled with robust examples of instruments that could be used to measure or examine one or more aspects of science teachers' CKT, but typically not the full breadth of their CKT within one or more science content areas. In addition, the current measures that do exist tend to use methods, such as in-depth interviews or open-ended survey responses, that require an extensive amount of time and human resources to administer and score. Such features preclude them from large scale use across a wide diversity of teacher education programs or professional development contexts. Providing empirical evidence indicating that discrete, automatically-scorable assessment items can be developed and used to measure the full breadth of elementary science teachers' CKT in one high-leverage content area addresses a significant gap in the current research.

To build a well-prepared STEM teacher workforce that is poised to close the STEM achievement gap, the field needs valid assessment tools for monitoring and improving elementary science teachers' learning in content-focused teacher preparation settings (Minner et al., 2012; National Academies of Sciences Engineering and Medicine, 2015; Wilson, 2016). The findings from this study suggest that there is much promise to developing and using these types of scenario-based CKT science items to address this need. The main implication is that this CKT item development approach, which uses teaching scenarios to address the CKT that science teachers leverage at the intersection of NGSS-aligned content topics (e.g., conservation of matter) and STPs (e.g., critiquing student-generated explanations or descriptions for their accuracy, precision, or consistency with scientific evidence), is both feasible and productive. As such, this approach has the potential for replication across other science content areas in future research and development efforts. Most importantly, these items have the potential to be used on a large scale and administered and scored efficiently to provide valid information to those who work with elementary science teachers. Doing so would provide elementary science teacher educators, district leaders, and professional development facilitators with valid assessment tools they can use to make empirically grounded instructional decisions about the nature and extent of the support they provide to develop elementary science teachers' CKT.

In our overall research project, we developed a complete CKT about matter assessment (Castellano & Mikeska, 2022) using a subset of the CKT about matter items from this study. As elementary science teacher educators use the CKT about matter assessment with their PSETs, the results can be used to provide the teacher educators with comprehensive score reports indicating their PSETs' strengths and areas of growth within these WOTS instructional categories and matter topic areas (Cisterna et al., 2021). Providing this fine-grained assessment information across a cohort of PSETs can support teacher educators in identifying specific areas where they can leverage their PSTs' strengths and where their PSETs might need more extensive scaffolded support to develop their CKT about matter.

As such, we see two primary purposes for the use of an efficiently scored assessment measuring elementary preservice teachers' CKT in a specific science topic area—one focuses on a formative use case and the other for summative use. In terms of formative use, the scores on a CKT assessment can provide specific, actionable information and feedback to support instructional decision-making for building elementary teachers' science CKT (Black & Wiliam, 2009). Alternatively, the scores generated from a CKT science assessment can be used to track potential changes in elementary teachers' CKT proficiency and evaluate the relative merit of various instructional interventions or strategies (Black et al., 2010; Harlen, 2007). Collectively, developing a CKT science assessment to address these formative and summative use cases can support future research that enriches the field's understanding of how science teachers build their CKT over time and how teacher educators and professional development leaders interact with, make sense of, and use CKT assessment tools to support their instructional decision making as they work to productively develop elementary teachers' CKT.

5.2 | Developing science teachers' CKT

The CKT about matter items are a comprehensive set of items around a high leverage content area (matter and its interactions). Teachers' reasoning about these items can provide us with a deeper and more complete understanding of where teachers may need support for developing their CKT in this area. Study findings provided a more nuanced examination of the nature of challenges these participants encountered related to the content and content teaching about matter and its interactions when their justifications did not meet the intended response. Participants' responses evidenced four main challenges with the science content: (a) using scientific concepts to reason about science tasks, (b) using adequate evidence to reason about science phenomenon, (c) drawing upon examples of scientific phenomena, and (d) drawing upon science vocabulary. The responses in these areas reveal both assets and opportunities for future learning. As Table 8 illustrates, multiple content topics are implicated from physical and chemical changes to conservation of matter, which can be addressed in teacher preparation. As teachers' CKT in these content topics grow, they can become better able to identify and use evidence in their reasoning of science phenomena (e.g., the formation of bubbles is evidence of a chemical change). High-quality learning opportunities that include instructional materials aimed to develop more sophisticated understanding of matter-related ideas (e.g., Merritt et al., 2008) can be used to support teachers. Likewise, teachers need learning opportunities to learn about and analyze common scientific phenomena, such as the combination of baking soda and vinegar (Liu & Lesniak, 2006) and key vocabulary used in this content area, as they serve as foundational knowledge for making sense of scientific phenomena and reasoning about tasks. The categorization of different types of challenges that teachers face with content could also be used to support the design of instructional materials for use with preservice teachers in teacher education or in-service teachers in professional development settings.

Findings also showed that participants experienced challenges with: (a) connecting to key scientific concepts involved in the WOTS, (b) attending to instructional goal(s), and (c) recognizing features of grade-level appropriateness. These three categories for teacher challenges broadly connect to the knowledge of curriculum dimension described by Hanuscin et al. (2018) for elementary teachers' pedagogical content knowledge about matter. These three categories also highlight teachers' opportunities for future learning. For example, teachers can have a rich understanding of scientific ideas and their students' ideas about a phenomenon (e.g., understanding how specific substances will react and recognizing students' misconceptions about matter, respectively), which can provide a foundation for future learning opportunities. As teachers are provided with opportunities to grow their content knowledge, they will be better able to connect the key scientific concepts they are learning about to their teaching about matter and its interactions (e.g., understanding which physical properties can be used for sorting). Likewise, through these experiences, teachers can become better at honing in on the specific instructional goals for a task and identifying the grade-level appropriate features of a task. As noted in the literature (Davis et al., 2006), elementary science teachers, especially those who are preservice teachers, require scaffolded support and learning opportunities to build a strong understanding of scientific concepts and more sophisticated knowledge on how to apply that knowledge to the work of teaching elementary science. Research also has suggested that preservice science teachers need support to accurately interpret and respond to students' ideas or select and use topic-specific instructional activities or representations to support student learning (Forbes et al., 2015; Schneider & Plasman, 2011). By providing a more nuanced picture of these content and content teaching challenging in one science topic area, study findings build upon and extend the current research.

Results illustrate that elementary science teachers, as suggested in previous research (e.g., Aydeniz et al., 2017; Bayuni et al., 2018; Del Pozo, 2001; Ginns & Watters, 1995; Kokkotas et al., 1998; Smith & Plumley, 2016), sometimes do have similar misunderstandings as their students, have incomplete or intuitive explanations about scientific concepts, and are unable to use evidence to reason about a scientific phenomenon. Similar to K-5 students, some elementary science teachers also have limited experience engaging with scientific phenomena to illustrate and investigate specific scientific topics, which makes it difficult for them to provide such experiences for their own students. Research also describes the difficulties that some teachers face to plan and enact instruction about matter-related topics aligned with the NGSS (Hanuscin et al., 2018; Smith et al., 2017). Second, study findings suggest that targeting connections between scientific concepts and the WOTS, considering how best to address specific instructional goals, and considering the grade-level appropriateness of scientific investigations, demonstrations, models, and representations should be priorities in PSETs' science teaching learning experiences. Yung et al. (2013) suggested that the use of rich task scenarios, similar to the ones presented in this study, have the potential to increase PSETs' familiarity with content and practices related to the work of science teachers. Thus, science teacher educators should not only should target the development of PSETs' understanding of the subject matter, but they also need to address the challenges that PSETs experience with understanding students' ideas and experiences and considering how to use instructional materials and resources in grade-appropriate ways to address specific instructional goals in science.

In summary, study findings underscore the point that elementary science teachers need high-quality learning opportunities to develop their CKT in specific science topic areas (Bertram & Loughran, 2012; Hanuscin et al., 2018; Henze & van Driel, 2015; Park & Suh, 2015). Such findings suggest the importance of carefully constructed learning opportunities tied directly to the gaps or types of challenges illustrated by these kinds of CKT assessment outcomes. These findings identify aspects of elementary science teachers' CKT about matter that would be ripe for addressing in future work. The findings from our study can provide guidance for the design of learning opportunities for teachers, which center specific ways of engaging with science content and science content teaching around the topic of matter and its interactions.

6 | CONCLUSION

Our study provides a nuanced picture of the nature of the content and content teaching challenges elementary teachers encounter when responding to CKT science assessment items in one high-leverage content area: matter and its interactions. Understanding the nature of the challenges elementary science teachers face in this area provides researchers, teacher educators, and professional development facilitators with specific guidance on what to look for, which can inform how they might respond to these different challenges in their work preparing elementary science teachers. Results also point to the productive use of discretely scored CKT science items to assess the CKT that elementary science teachers use in the WOTS. Future research should examine the development of CKT assessment items for other high-leverage content topics and extend the use of such items to examine similar content and content teaching challenges in other science topic areas.

ACKNOWLEDGMENTS

This study was supported by grants from the National Science Foundation (Award Nos. 1813254 and 1814275). The opinions expressed herein are those of the authors and not the funding agency. We are grateful to the elementary teachers who participated in this study and provided feedback to help us learn about and improve our design of assessment items for measuring content knowledge for teaching (CKT) about matter and its interactions. We are also appreciative of the item writers and lead facilitators who collaborated on developing and refining these CKT matter items for use in our overall research project, and for the guidance of our advisory board on this project.
Correspondence concerning this article should be addressed to Jamie Mikeska, Educational Testing Service, 660 Rosedale Rd., Princeton, NJ 08541. Email: jmikeska@ets.org.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jamie N. Mikeska http://orcid.org/0000-0002-8831-2572

Luronne Vaval http://orcid.org/0000-0002-3872-2928

ENDNOTES

[1] Our team developed the CKT about matter items in batches for feasibility purposes, as many of our external consultants were full time science teachers or science educational faculty members or researchers. Each batch focused on one or more of the five matter topics (e.g., conservation of matter, changes in matter), although the last batch cut across topic areas.

[2] Initially there were 37 CKT about matter items in which 50% or more of the participant responses showed difficulty with the content. However, three of these 37 CKT about matter items were eliminated from further analysis because there were item flaws, as determined by the item writers and/or content reviewers.

[3] Based on the guidelines in the Next Generation Science Standards, the scenario in this CKT about matter item does not distinguish between weight and mass and uses the more familiar term–weight–since the scenario is set in a K-5 classroom context.

REFERENCES

Avraamidou, L., & Zembal-Saul, C. (2010). In search of well-started beginning science teachers: Insights from two first-year elementary teachers. *Journal of Research in Science Teaching*, *47*(6), 661–686.

Aydeniz, M., Bilican, K., & Kirbulut, Z. D. (2017). Exploring pre-service elementary science teachers' conceptual understanding of particulate nature of matter through three-tier diagnostic test. *International Journal of Education in Mathematics, Science and Technology*, *5*(3), 221.

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.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180.

Bayuni, T. C., Sopandi, W., & Sujana, A. (2018). May Identification misconception of primary school teacher education students in changes of matters using a five-tier diagnostic test, *Journal of Physics: Conference Series* (Vol. No. 1. 1013, p. 012086). IOP Publishing.

Bertram, A., & Loughran, J. (2012). Science teachers' views on CoRes and PaP-eRs as a framework for articulating and developing pedagogical content knowledge. *Research in Science Education*, *42*(6), 1027–1047.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215–232.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, *21*(1), 5–31.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bybee, R., & Chopyak, C. (2017). Instructional materials and implementation of NGSS: Demand, supply, and strategic opportunities. A report for the Carnegie Corporation of New York.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., & Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education, *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer.

Castellano, K. E., & Mikeska, J. N. (2022). Developing and using a scalable assessment to measure preservice elementary teachers' content knowledge for teaching about matter. [Manuscript submitted for publication]. Data Strategy and Research Technology Center, ETS.

Chen, C., Sonnert, G., Sadler, P. M., & Sunbury, S. (2020). The impact of high school life science teachers' subject matter knowledge and knowledge of student misconceptions on students' learning. *CBE—Life Sciences Education*, *19*(1), ar9.

Cisterna, D., Bookbinder, A. K., Mikeska, J. N., & Lakhani, H. R. (2022). Elementary preservice teachers' perceptions of assessment tasks that measure content knowledge for teaching about matter. *Journal of Science Teacher Education*, *33*(8), 1–28. https://doi.org/10.1080/1046560X.2021.2015831

Cisterna, D., Mikeska, J. N., Castellano, K., & Lentini, J. (2021). April 7-10 Exploring science teacher educators' evaluation of a score report to support content knowledge for teaching. *[Poster presentation]*. National Association for Research in Science Teaching Annual Meeting, Orlando, FL.

Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research*, *76*(4), 607–651.

Eberle, F., & Keeley, P. (2008). Formative assessment probes. *Science and Children*, *45*(5), 50–54.

Etkina, E., Gitomer, D., Iaconangelo, C., Phelps, G., Seeley, L., & Vokos, S. (2018). Design of an assessment to probe teachers' content knowledge for teaching: An example from energy in high school physics. *Physical Review Physics Education Research*, *14*, 010127.

Forbes, C. T., Sabel, J. L., & Biggers, M. (2015). Elementary teachers' use of formative assessment to support students' learning about interactions between the hydrosphere and geosphere. *Journal of Geoscience Education*, *63*(3), 210–221. https://doi.org/10.5408/14-063.1

Gale, J., Koval, J., Ryan, M., Usselman, M., & Wind, S. (2019). Implementing NGSS engineering disciplinary core ideas in middle school science classrooms: Results from the field. *Journal of Pre-College Engineering Education Research (J-PEER)*, *9*(1), 2.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). Routledge.

Ginns, I. S., & Watters, J. J. (1995). An analysis of scientific understandings of preservice elementary teacher education students. *Journal of Research in Science Teaching*, *32*(2), 205–222. https://doi.org/10.1002/tea.3660320209

Gómez, E. J., Benarroch, A., & Marín, N. (2006). Evaluation of the degree of coherence found in students' conceptions concerning the particulate nature of matter. *Journal of Research in Science Teaching*, *43*(6), 577–598. https://doi.org/10.1002/tea.20130

Gray, R., McDonald, S., & Stroupe, D. (2022). What you find depends on how you see: Examining asset and deficit perspectives of preservice science teachers' knowledge and learning. *Studies in Science Education*, *58*(1), 49–80.

Hadenfeldt, J. C., Neumann, K., Bernholt, S., Liu, X., & Parchmann, I. (2016). Students' progression in understanding the matter concept. *Journal of Research in Science Teaching*, *53*(5), 683–708.

Hanuscin, D. L., Cisterna, D., & Lipsitz, K. (2018). Elementary teachers' pedagogical content knowledge for teaching structure and properties of matter. *Journal of Science Teacher Education*, *29*(8), 665–692.

Hanuscin, D. L., & Zangori, L. (2016). Developing practical knowledge of the next generation science standards in elementary science teacher education. *Journal of Science Teacher Education*, *27*(8), 799–818.

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, *16*(2), 207–223.

Harrison, A. G., & Treagust, D. F. (2003). The particulate nature of matter: Challenges in understanding the submicroscopic world. In J. K. Gilbert, O. De Jong, R. Justi, D. F. Treagust, & J. H. Van Driel (Eds.), *Chemical education: Towards research-based practice* (pp. 189–212). Kluwer Academic Publishers.

Henze, I., & Van Driel, J. H. (2015). Toward a more comprehensive way to capture PCK in its complexity, In *Re-examining pedagogical content knowledge in science education* (pp. 120–134). Routledge.

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

Hill, H. C., 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.

Howell, H., Phelps, G., Croft, A. J., Kirui, D., & Gitomer, D. (2013). Cognitive interviews as a tool for investigating the validity of content knowledge for teaching assessments. *ETS Research Report Series*, *2013*, (1), i–97.

Isabelle, A. D. (2017). STEM is elementary: Challenges faced by elementary teachers in the era of the next generation science standards. *The Educational Forum*, *81*(1), 83–91. https://doi.org/10.1080/00131725.2016.1242678

Jackson, J. (2009). H2O and you. *Science Activities: Classroom Projects and Curriculum Ideas*, *46*(1), 3–6. https://doi.org/10.3200/SATS.46.1.3-6

Kahveci, A. (2009). Exploring chemistry teacher candidates' profile characteristics, teaching attitudes and beliefs, and chemistry conceptions. *Chemistry Education Research and Practice*, *10*(2), 109–120.

Kang, E. J. S., McCarthy, M. J., & Donovan, C. (2019). Elementary teachers' enactment of the NGSS science and engineering practices. *Journal of Science Teacher Education*, *30*(7), 788–814.

Kang, N.-H. (2007). Elementary teachers' epistemological and ontological understanding of teaching for conceptual learning. *Journal of Research in Science Teaching*, *44*(9), 1292–1317. https://doi.org/10.1002/tea.20224

Keeley, P. (2016). Uncovering students' concepts of matter. *Science and Children*, *53*(5), 26–28. https://search.proquest.com/docview/1753969315?accountid=458

Keeley, P., Eberle, F., & Tugel, J. (2007). *Uncovering student ideas in science*. National Science Teachers Association.

Kessler, J. H., & Galvan, P. M. (2007). Inquiry in action: Investigating matter through inquiry. Retrieved from http://www.inquiryinaction.org/

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

Kokkotas, P., Vlachos, I., & Koulaidis, V. (1998). Teaching the topic of the particulate nature of matter in prospective teachers' training courses. *International Journal of Science Education*, *20*(3), 291–303.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *The Science Teacher*, *79*(3), 38.

Krnel, D., Gla?ar, S., & Watson, R. (2003). The development of the concept of "matter": A cross-age study of how children classify materials. *Science Education*, *87*(5), 621–639. https://doi.org/10.1002/sce.10080

Lee, E. J., Cite, S., & Hanuscin, D. (2014). Taking the "mystery" out of argumentation. *Science and Children*, *52*(1), 46–52.

Lee, O. (2017). Common core state standards for ELA/literacy and next generation science standards: Convergences and discrepancies using argument as an example. *Educational Researcher*, *46*(2), 90–102.

Lee, O., Eichinger, D. C., Anderson, C. W., Berkheimer, G. D., & Blakeslee, T. D. (1993). Changing middle school students' conceptions of matter and molecules. *Journal of Research in Science Teaching*, *30*(3), 249–270. https://doi.org/10.1002/tea.3660300304

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English language arts and mathematics. *Educational Researcher*, *42*(4), 223–233.

Liu, X., & Lesniak, K. (2006). Progression in children's understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, *43*(3), 320–347.

Liu, X., & Lesniak, K. M. (2005). Students' progression of understanding the matter concept from elementary to high school: Matter concept. *Science Education*, *89*(3), 433–450. https://doi.org/10.1002/sce.20056

Lott, K., & Jensen, A. (2012). Changes matter! *Science and Children*, *50*(2), 54–61.

Lott, K., & Wallin, L. (2012). Modeling the states of matter in a first-grade classroom. *Science Activities: Classroom Projects and Curriculum Ideas*, *49*(4), 108–116. https://doi.org/10.1080/00368121.2012.706241

Merritt, J., Krajcik, J., & Shwartz, Y. (2008). Development of a Learning Progression for the Particle Model of Matter. In G. Kanselaar, V. Jonker, P. A. Kirschner, & F. J. Prins (Eds.), *International Perspectives in the Learning Sciences: Creating a learning world. Proceedings of the Eighth International Conference for the Learning Sciences – ICLS 2008, Volumes 2* (pp. 75–81). International Society of the Learning Sciences.

Mikeska, J. N., Anderson, C. W., & Schwarz, C. V. (2009). Principled reasoning about problems of practice. *Science Education*, *93*(4), 678–686. https://doi.org/10.1002/sce.20312

Mikeska, J. N., Brockway, D., Ciofalo, J., Jin, H., & Ritter, S. (2020). Examining variability in elementary science teachers' pedagogical content knowledge about phase change: Implications for teacher development and assessment. *Journal of Science Teacher Education*, *32*(4), 400–424. https://doi.org/10.1080/1046560X.2020.1831741

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

Mikeska, J. N., Phelps, G., & Croft, A. J. (2017). Practice-based measures of elementary science teachers' content knowledge for teaching: Initial item development and validity evidence. *ETS Research Report Series*, *2017*, 1–72. https://files.eric.ed.gov/fulltext/EJ1168726.pdf

Minner, D., Martinez, A., & Freeman, B. (2012). *Compendium of research instruments for STEM education*. Part 1: Teacher practices, PCK, and content knowledge.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered design: Layers, concepts, and terminology. In S. Downing, & T. M. ahway Haladyna (Eds.), *Handbook of test development*.

Nakhleh, M. B., & Samarapungavan, A. (1999). Elementary school children's beliefs about matter. *Journal of Research in Science Teaching*, *36*(7), 777–805. https://doi.org/10.1002/(SICI)1098-2736(199909)36:7<777::AID-TEA4>3.0.CO;2-Z

Nakhleh, M. B., Samarapungavan, A., & Saglam, Y. (2005). Middle school students' beliefs about matter. *Journal of Research in Science Teaching*, *42*(5), 581–612. https://doi.org/10.1002/tea.20065

National Academies of Sciences, Engineering, and Medicine. (2015). *Science teachers' learning: Enhancing opportunities, creating supportive contexts*. National Academies Press.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next Generation Science Standards: For States, by States*. The National Academies Press.

Özalp, D., & Kahveci, A. (2015). Diagnostic assessment of student misconceptions about the particulate nature of matter from ontological perspective. *Chemistry Education Research and Practice*, *16*(3), 619–639.

Palmeri, A., Cole, A., DeLisle, S., Erickson, S., & Janes, J. (2008). What's the matter with teaching children about matter? *Science and Children*, *46*(4), 20–23.

Park, S., & Suh, J. K. (2015). From portraying toward assessing PCK: Drivers, dilemmas, and directions for future research, In *Re-examining pedagogical content knowledge in science education* (pp. 104–119). Routledge.

Penuel, W. R., Harris, C. J., & DeBarger, A. H. (2015). Implementing the next generation science standards. *Phi Delta Kappan*, *96*(6), 45–49. https://doi.org/10.1177/0031721715575299

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. *ETS Research Report Series*, *2020*, 1–44. https://doi.org/10.1002/ets2.12295

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

del Pozo, R. M. (2001). Prospective teachers' ideas about the relationships between concepts describing the composition of matter. *International Journal of Science Education*, *23*(4), 353–371. https://doi.org/10.1080/095006901300069084

Pruitt, S. L. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education*, *25*(2), 145–156.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*(2), 117–148.

Royce, C. A. (2012). What's the matter? *Science and Children*, *49*(6), 22.
Richmond, G., & Manokore, V. (2011). Identifying elements critical for functional and sustainable professional learning communities. *Science Education*, *95*(3), 543–570.
Ryan, M., Gale, J., & Usselman, M. (2017). Integrating engineering into core science instruction: Translating NGSS principles into practice through iterative curriculum design. *International Journal of Engineering Education*, *33*(1), 321–331.
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.
Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, *81*(4), 530–565. https://doi.org/10.3102/0034654311423382
Sinapuelas, M. L. S., Lardy, C., Korb, M. A., & DiStefano, R. (2018). Toolkit to support preservice teacher dialogue for planning NGSS three-dimensional lessons. *Innovations in Science Teacher Education*, *3*(4). Retrieved from https://innovations.theaste.org/a-toolkit-to-support-preservice-teacher-dialogue-for-planning-ngss-three-dimensional-lessons/
Smith, P. S. (2010). *New Tools for Investigating the Relationship between Teacher Content Knowledge and Student Learning*. Presented at the 2010 NARST Annual Conference.
Smith, P. S., & Plumley, C. (2016). *A review of the research literature on teaching about the small particle model of matter to elementary students*. Horizon Research, Inc.
Smith, P. S., Plumley, C. L., & Hayes, M. L. (2017). Eliciting elementary teachers' PCK for the small particle model. Horizon Research, Inc.
Taber, K. S., & García-Franco, A. (2010). Learning processes in chemistry: Drawing upon cognitive resources to learn about the particulate structure of matter. *Journal of the Learning Sciences*, *19*(1), 99–142.
Talanquer, V. (2009). On cognitive constraints and learning progressions: The case of structure of matter. *International Journal of Science Education*, *31*(15), 2123–2136. https://doi.org/10.1080/09500690802578025
Tolbert, S., Stoddart, T., Lyon, E. G., & Solís, J. (2014). The next generation science standards, common core state standards, and English learners: Using the SSTELLA framework to prepare secondary science teachers. *Issues in Teacher Education*, *23*(1), 65–90.
Troncale, J. M. (2016). Sensing matter: Is it a liquid or solid? *Science and Children*, *54*(3), 58–63.
Tuttle, N., Kaderavek, J. N., Molitor, S., Czerniak, C. M., Johnson-Whitt, E., Bloomquist, D., Namatovu, W., & Wilson, G. (2016). Investigating the impact of NGSS-aligned professional development on PreK-3 teachers' science content knowledge and pedagogy. *Journal of Science Teacher Education*, *27*(7), 717–745.
Tyler, B., Estrella, D., Britton, T., Nguyen, K., Iveland, A., Nilsen, K., & Valcarcel, J. (2020). What education leaders can learn about NGSS implementation: Highlights from the early implementers initiative. Evaluation Report# 14. *WestEd*.
Valanides, N. (2000). Primary student teachers' understanding of the particulate nature of matter and its transformation during dissolving. *Chemistry Education: Research and Practice in Europe*, *1*(2), 249–262. Retrieved from http://www.chem.uoi.gr/cerp/2000_May/pdf/33-06valanides.pdf
Varelas, M., Pappas, C. C., & Rife, A. (2006). Exploring the role of intertextuality in concept construction: Urban second graders make sense of evaporation, boiling, and condensation. *Journal of Research in Science Teaching*, *43*(7), 637–666. https://doi.org/10.1002/tea.20100
Vowell, J., & Phillips, M. (2015). All aboard! The polar express is traveling to science—Understanding the states of matter while differentiating instruction for young learners. *Science Activities*, *52*(1), 1–8. https://doi.org/10.1080/00368121.2014.986039
Weishaar, S. (2011). Gummies solid or liquid? *Connect Magazine*, *24*(4), 6–9.
Wilson, S. M. (2016). Measuring the quantity and quality of the K-12 STEM teacher pipeline. *Education*, *1*, 859–2000.
Windschitl, M., Thompson, J., & Braaten, M. (2020). *Ambitious Science Teaching*. Harvard Education Press.
Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903. https://doi.org/10.1002/sce.21027
Yung, B. H. W., Zhu, Y., Wong, S. L., Cheng, M. W., & Lo, F. Y. (2013). Teachers' and students' conceptions of good science teaching. *International Journal of Science Education*, *35*(14), 2435–2461. https://doi.org/10.1080/09500693.2011.629375

How to cite this article: Mikeska, J. N., Cisterna, D., Lakhani, H., Bookbinder, A. K., Myers, D. L., & Vaval, L. (2022). Examining elementary science teachers' responses to assessments tasks designed to measure their content knowledge for teaching about matter and its interactions. *Science Education*, 1–37. https://doi.org/10.1002/sce.21779