Design of an assessment to probe teachers' content knowledge for teaching: An example from energy in high school physics

Eugenia Etkina,[1,*] Drew Gitomer,[1] Charles Iaconangelo,[1] Geoffrey Phelps,[2] Lane Seeley,[3] and Stamatis Vokos[4]

[1]*Department of Learning and Teaching, Graduate School of Education, Rutgers University, New Brunswick, New Jersey 08901, USA*
[2]*Educational Testing Service, Princeton, New Jersey 08541, USA*
[3]*Department of Physics, Seattle Pacific University, Seattle, Washington 98119, USA*
[4]*Physics Department, California Polytechnic State University, San Luis Obispo, California 93407, USA*

Research into teacher learning and practice over the last three decades shows that the teachers of a specific subject need to possess knowledge that is different from the knowledge of other content experts. Yet this specialized version of content knowledge that teachers need to plan instruction, respond to student ideas, and assess student understanding in real time is a theoretically elusive construct. It is crucial for the fields of precollege teacher preparation, teacher professional education, and postsecondary faculty professional development to (a) clarify the construct that underlies this specialized content knowledge, (b) operationalize it in some domain, (c) measure it in both static contexts and as it is enacted in the classroom, and (d) correlate its presence with "richness" of classroom instruction and its effect on student learning. This paper documents a piece of a multiyear, multi-institutional effort to investigate points (a)–(d) in the domain of energy in the first high school physics course. In particular, we describe the framework that we developed to clarify content knowledge for teaching in the context of high school energy learning. We then outline the process through which we developed, tested, and refined a "paper-and-pencil" assessment administered on a computer and discuss the substantive and psychometric features of several items based on a field test of the final form of the assessment. We choose to discuss these items for a dual purpose: to illustrate the application of our general framework and to present performance findings from a sample of 362 practicing high school teachers of physics.

I. INTRODUCTION

It is a truism that teachers "need" content knowledge to teach effectively. What is much less clear is the kind of content knowledge that is needed. To be sure, science teachers use content knowledge in ways that are different than a lab scientist or a practicing engineer would. In many teacher preparation programs, for instance, it is assumed that aspiring physics teachers develop subject matter knowledge in the introductory physics courses for scientists and engineers and, possibly, in more advanced courses that are intended for physics majors. Then the prospective teachers enroll in education courses in which they develop knowledge of pedagogy (including assessment techniques and ways to interact productively with diverse students) and

also hone their understanding of schools and communities. They typically take a course on science teaching methods, which gives them experience developing one or more lessons or units. Finally, they observe and subsequently student teach in a physics classroom. Through this process, it is hoped that content knowledge specifically tuned to teaching, which will be deployed daily in the classroom, develops with teaching experience.

Ms. Gonzales' physics class has been studying energy for a few days when the teacher asks a question to check on students' evolving understanding. The question is about a girl who has been hiking and reaches a very tall, level overhang that overlooks the valley below. Ms. Gonzales asks, "As the girl moves closer to the edge of the level overhang, does the gravitational potential energy of the girl and Earth increase, decrease, or stay the same?" About 15% of the class responds that it increases. What knowledge did Ms. Gonzales marshal in formulating the question? What knowledge fuels the interpretation that she will ascribe to each of the students' responses? What knowledge underlies the

*eugenia.etkina@gse.rutgers.edu

instructional action that she will devise as a result of the ways in which the class responded to her question?

Yet this specialized content knowledge is an elusive construct. It is crucial for the fields of precollege teacher preparation, teacher professional education, and postsecondary faculty professional development to (a) clarify the construct that underlies specialized content knowledge, (b) operationalize it in some domain, (c) measure it both in static contexts and in the classroom, and (d) correlate its presence with "richness" of classroom instruction and its effect on student learning.

This paper documents a piece of a multiyear, multi-institutional effort to investigate points (a)–(d) in the domain of energy in the first high school physics course. In particular, we describe the framework that we developed to clarify content knowledge for teaching (CKT) in the context of high school energy learning. We then outline the process through which we developed, tested, and refined a "paper-and-pencil" assessment administered on a computer and discuss the substantive and psychometric features of several items based on a field test of the final form of the assessment.

II. THEORETICAL FRAMEWORK

This study investigates the application of the construct of CKT to the domain of energy in physics (CKT-E). The following example serves to set the stage. Content experts know Kirchhoff's rules, but teachers need to know the additional following things: that students treat a battery as if it were a constant current source, consider current in a circuit as if it were used up as it passes through elements connected in series, and have powerful "flow" resources that can be built upon. Furthermore, teachers should expect that voltage and current are not well-differentiated ideas among students and pay special attention to the language that students use. In this sense, attention to language plays a different role in the life of a teacher whereas content experts can get away with using formally incorrect or sloppy language because their interlocutors know what they mean.

This example demonstrates how the CKT theoretical approach helps one to differentiate between the knowledge used by a teacher and the knowledge used by a content expert in the same discipline. CKT involves the intersection of teaching knowledge and skill (e.g., appreciating that students have incomplete or inaccurate conceptions) with specific content (e.g., energy or electrical circuits). Therefore, in studying CKT, a grain size of the whole subject domain (e.g., mathematics, physics, biology) may be too broad to operationalize with a single assessment. Content knowledge for teaching electric fields in physics is very different from content knowledge for teaching geometrical optics.

Thus, to assess CKT in a particular content area, we think about such knowledge as an integration of some general knowledge of learning and teaching processes that occur in any physics classroom and specific learning targets in a subject matter area that the students need to reach (e.g., what students should learn about forces, energy, momentum, magnetic field, etc.). The targets can be conceptualized in terms of disciplinary core ideas, crosscutting concepts, and science practices (three-dimensional learning of the *Next Generation Science Standards* [1]). Below we describe a generalizable, iterative process for operationalizing CKT in a particular area of physics, namely, energy taught in the context of mechanics in the high school physics course, culminating in the design, administration, and analysis of an assessment of this knowledge.

A. Content knowledge for teaching

The concept of CKT originated with the pedagogical content knowledge (PCK) work of Shulman [2] and was more fully developed by Ball and colleagues [3]. During the past three decades there has been significant progress in defining, categorizing, and assessing PCK domains [4,5].

There are several instruments that assess teachers' PCK in specific physics domains [6–9]. The Magnusson, Krajcik, and Borko model of science PCK [10] was the first attempt to detail this knowledge, and recently a new, revised model of PCK has emerged [11]. In this model PCK is just one component of teacher professional knowledge and practice. Similar to the PCK conceptualization of the specialized subject matter knowledge that teachers use, CKT is premised on the idea that teachers need to understand subject matter knowledge in ways that are specific to teaching, such as understanding the historical foundations of the concepts that students need to learn, structure of the curriculum that allows students to build coherent understanding [12–16], challenges that specific subject matter knowledge might present to students and how students may represent their understanding in nonstandard forms, knowing what knowledge representations are helpful, how to ask questions or provide explanations that can move understanding forward, etc. [17].

There have been several efforts to define and assess CKT, both in mathematics and in reading. In one of the first efforts to define and assess this construct, Hill, Schilling, and Ball [4] analyzed high-quality mathematics instruction, as well as student work and curricular materials, to develop a set of assessment items meant to measure a teacher's mathematics CKT. Focusing on elementary mathematics, they developed items that were meant to assess both a teacher's subject matter knowledge and their specialized knowledge for teaching in three domains: number concepts; operations; and patterns, functions, and algebra. From this study, they found evidence of multiple factors affecting teachers' scores on the assessment related to both content area and question type (subject matter knowledge and specialized content knowledge). Other studies have also worked to develop similar assessments in math. The *Measures of Effective Teaching* (MET) study developed a

set of CKT assessments in grades 4–5 math, grades 6–8 math, and grade 9 algebra [6]. The COACTIV study [10] led to the development of assessments of both teachers' subject matter knowledge and pedagogical content knowledge [11], and the international Teacher Education and Development Study developed assessment items to measure both teachers' subject matter knowledge and their CKT [18].

While most of the work on defining and assessing CKT has been done in mathematics, some work has also been done in reading. Phelps and Schilling [19] worked to develop a set of items to assess a teacher's elementary reading CKT and created a test that assessed both a teacher's subject matter knowledge and their PCK for teaching reading. Additionally, Kuncan, Hapgood, and Palinscar developed a constructed-response assessment meant to evaluate a teacher's specialized knowledge for teaching reading comprehension [20]. Other work has been done to assess reading CKT, including the *Language and Reading Concept Assessment*, which was designed to assess subject matter knowledge and the knowledge for teaching phonemic awareness, phonics, fluency, vocabulary, and reading comprehension [21]. Additionally, the MET project [5] developed assessments of CKT for grades 4–6 English language arts (ELA) and grades 7–8 ELA [6].

While CKT assessments have been created for math and reading, there have been no similar assessments developed for physics. The project on which this study is based is an effort to define, assess, and validate the construct of CKT in a narrow content area, specifically, the teaching of energy in the context of mechanics instruction in high school physics.

B. Content knowledge for teaching energy framework

The work described here is one of the products of a multiyear effort to develop and validate a set of substantively coherent measures that assess CKT in physics in the domain of energy, through both tests and evidence from instructional practice. These measures include results from a CKT assessment administered to teachers via computer, classroom observation and artifact protocols, and student performance data on energy tasks culled from other projects. Our project focused on one conceptual area so as to forge a tight theoretical and empirical link between CKT and practice. In this paper, we focus almost exclusively on the operationalization of the CKT construct and the corresponding assessment we developed.

The first step in this process was to develop a domain model of CKT-E. We developed the domain model through an extensive review of the literature. We further refined the model by using it to characterize the teaching of 32 expert physics teachers who volunteered to participate in an intensive study of physics teaching that included daily observation, artifact collection, and interviews during the teaching of an energy unit (2–4 weeks of instruction).

The domain model of CKT-E involves two components. The first component is the critical *Tasks of Teaching* (ToTs)

[22] that describe the key activities through which teachers and students enact practices that promote and support student learning. The model used in this study includes the following tasks of teaching: (I) anticipating student thinking around science ideas; (II) designing, selecting, and sequencing learning experiences and activities; (III) monitoring, interpreting, and acting on student thinking; (IV) scaffolding meaningful engagement in a science learning community; (V) explaining and using examples, models, representations, and arguments to support students' scientific understanding; and (VI) using experiments to construct, test, and apply concepts (see Appendix A for a complete list of the Tasks of Teaching in our context).

While it is not expected that teachers engage in all tasks of teaching in every lesson, we should be able to observe a teacher engaged in each of these tasks many times during the teaching of an energy unit. Further, while these tasks of teaching are not the only tasks in which teachers engage while teaching, the CKT theory assumes that for students to learn, teachers should engage in all of these tasks across each unit of instruction [6,10,18–22].

The second component of our domain model is the *Student Energy Targets* (SETs). It focuses on specific content targets of energy in mechanics for students and articulates features that are important in the domain (energy taught in the context of mechanics in a typical high school physics course, in our case) including concepts and skills, critical tasks in which those are manifested, and knowledge representations. The choice of targets is in line with existing research on the learning and teaching of energy [7,23–41] and combined 90-years physics teaching experience of three authors of this paper (E. E, L. S., and S. V.). Knowledge of prevalent student difficulties is implicitly represented in our student learning targets. Targets are separated into several broad categories: (A) connections of energy and everyday experiences [23–25]; (B) choice of system [26–29]; (C) identification of and differentiation between energy and other physics concepts [30,31]; (D) transfer of energy [32–39]; (E) use of mathematics; (F) use of representations [26,32–38]; and (G) use of other science practices [40].

From the point of view of the Next Generation Science Standards (NGSS),[1] these targets address disciplinary core ideas related to energy (targets A, C, D, and E), crosscutting concepts (targets B, C), and science practices (targets F and G). We also provided elaborations on each of these areas (see Appendix B). For example, for target B, choice of system, the elaborations are as follows:

---

[1]We must note here that when we were developing SETs, the science and engineering practices were not as widely adopted as they are now, therefore we did not create an assessment that purposefully assesses all three strands of 3D learning. While our student energy targets are not fully aligned with NGSS, they address most of the practices with some differences in nomenclature.

The student

(1) recognizes that the energy accounting for a phenomenon depends on the choice of system;

(2) explains the relative advantage of a given system choice (i.e., relative ease of analysis);

(3) recognizes that the choice of system determines whether springs or Earth do work on the system (i.e., if the spring or Earth are in the system they do not do any work on the system, but the system can possess elastic or gravitational potential energy); and

(4) identifies and differentiates between forms of energy and other physics concepts.

We conceptualize CKT-E as "residing" at the intersection of specific tasks of teaching with the student energy targets. In essence, we ask what knowledge a teacher would need to "have" to execute a particular task of teaching in the domain of energy to support a particular student energy target. This domain model allows us to design and interpret the findings of the CKT-E assessments, the classroom observation and artifact protocols, and assessments of student knowledge of energy. The next section elaborates on the interaction of student energy targets and tasks of teaching.

C. CKT-E residing at the intersection of ToTs with SETs

Our student energy targets represent a finite set of understandings that a teacher might hope that all of their students would construct through their energy learning experiences. Our tasks of teaching are the various tasks that a teacher would carry out in order to support student learning. CKT-E is the knowledge that a teacher is likely to draw on in their efforts to carry out ToTs in service of SETs. Operationalized in this way, CKT-E extends beyond the student learning targets and includes both disciplinary knowledge and pedagogical knowledge.

Consider the example of a teacher who is striving to help her students construct and refine their thinking around gravitational energy. She would like them to understand how the gravitational energy of an Earth-object system depends on the mass of the object and the height of that object above a reference height (SETs C1 and E1; see Appendix B). She would also like them to be able to apply and make sense of the mathematical expression for gravitational energy near Earth's surface, mgh. In order to help her students construct their own understanding of this relationship it is important that the teacher challenge students to reconcile this mathematical representation of gravitational energy with their own experience and intuition (ToT III. c; see Appendix A). A teacher who is striving to enact this particular ToT in support of these SETs might anticipate the following student question: *"I heard about the space probe that finally left the solar system forty years after it escaped Earth's gravity. If gravitational energy equals mgh then doesn't the gravitational energy just keep going up and up all the way to infinity?"* This is a wonderful question, and it is crucial to helping this particular student reconcile the mathematical representation for gravitational energy with his or her own ideas. In order to help the student engage productively with this question a teacher must bring specific disciplinary knowledge to bear. She must recognize that mgh is a mathematical approximation, which is only valid near Earth's surface. A more general mathematical approximation $-G(m_1 m_2/r_{12})$ is needed to describe the gravitational energy of objects that are far from Earth. The preceding disciplinary knowledge, which the teacher would call upon in this situation, extends beyond her energy learning targets for all students. Ball has categorized this disciplinary knowledge as horizon knowledge [3]. In addition to drawing on disciplinary knowledge, the teacher would call on specific pedagogical knowledge. In particular, she would need to recognize that many learners would expect a correct mathematical model for energy to be generalizable to all situations. She would also need to anticipate the inconsistencies that students may encounter when they attempt to apply this particular mathematical model for gravitational energy to situations that interest them. We would describe both the disciplinary and pedagogical knowledge to be part of this teacher's CKT-E.

The preceding example illustrates how any instructional situation that results from a particular ToT in support of specific SETs can reveal both disciplinary and pedagogical examples of CKT-E. The exact CKT-E that is engaged will depend on the SETs, the ToTs, and the instructional situation. Therefore, fully listing all examples of CKT-E is not productive. In our construction of a written CKT-E assessment, we have focused on a purposefully selected sample of ToT and SET combinations as instantiated through various instructional scenarios. Some of the items assess disciplinary knowledge of physics that may be relevant for a teaching situation but does not require detailed knowledge of student learning or of the school context to be answered correctly. We designate these items as content knowledge for teaching-disciplinary (CKT-D). One example would be the ability to identify an unarticulated system choice on the basis of energy conversions that a student described and assess on the spot whether the analysis is consistent with such choice. Solutions to the second set of items require disciplinary knowledge but also require an understanding of student learning and how to best teach students. These items are designated at content knowledge for teaching-pedagogical (CKT-P). About half of the items on the assessment are of the CKT-D type.

Each of the example items discussed includes two questions that are readily categorized as CKT-D and CKT-P. Not all of the items on this assessment can be sharply divided into these two categories. Rather, the assessment items are distributed along a continuum of

increasing pedagogical challenge. The items that we have categorized as CKT-P are those in which the principal cognitive challenge is directly related to pedagogy. We should also clarify that items that we have classified as CKT-P should not be considered to involve less sophisticated physics knowledge. On the contrary, many of these items require physics knowledge that is both sophisticated and subtle.

III. DEVELOPMENT OF THE CKT-E ASSESSMENT

In this section we outline the process that we followed.

A. Initial item development

A team of expert physics educators with support from a group of assessment experts led the design of the items. In designing CKT-E assessment items that probe the specialized content knowledge that teachers employ to support students' productive scientific engagement with the domain of energy in mechanics, we focused on the tasks of teaching and student energy targets of our domain framework. The entire assessment incorporates 15 different instructional scenarios (total of 50 questions as scenarios have multiple questions). The instructional scenarios are hypothetical but are all based on actual classroom experiences and/or video from high school physics classrooms. Each scenario provides the context for one or more individual items, including selected-response and constructed-response formats, to address authentic challenges to teaching energy. For each item we developed a rationale detailing the item design and justification for both the correct answer and the distractors (for selected-response items).

B. Expert teacher item review

The first of several iterations of item review and refinement was supported by a small group of expert high school physics teachers. These individuals attempted initial versions of an item and then came together to discuss different aspects of the item, guided by the following questions:
- Is the item clear and unambiguous?
- Is the target response clearly the best answer among the choices provided from a disciplinary perspective?
- Does each of the distractors present a compelling yet incorrect answer to the item?
- Does the item context represent a learning opportunity that is consistent with the goals of energy learning at the appropriate level?

In instances where the expert reviewers had concerns about items and/or answer options, items were revised, discarded, or replaced. Generally, the expert reviewers described items as challenging and appropriate for high school-level energy learning in physics.

C. Modeling of assessment using item response theory

Item response theory (IRT) models are a type of latent variable models where the observed responses are considered manifestations of a construct that cannot be observed directly. In our case the construct of interest is considered to be content knowledge for teaching energy (CKT-E). This construct assigns an overall CKT-E score to each teacher, which for the ith teacher is labeled Θ_i. A theta of zero can be interpreted as the average performance on our CKT-E assessment, and a theta of ± 1 can be interpreted as performance (ability) 1 standard deviation above or below the average.

There are several IRT models, such as two- or three-parameter logistic models (2-PL, 3-PL) that relate the examinee's estimated ability (Θ_i) to his or her probability of responding correctly to a given item (e.g., for an easier item an examinee of average ability might have a 0.60 probability of responding correctly whereas for a more difficult item the same average examinee might have a 0.40 probability of responding correctly). The IRT models are typically represented in a graphical form of a probability curve that allows us to estimate parameters of that curve. In the 3PL model the three parameters are as follows:
- slope (a parameter related to the maximum slope of the probability curve), which estimates how well the item discriminates respondents of different Θ_i.
- threshold (a parameter related to the difficulty of the item), which estimates the ability level Θ of a respondent who has a 50% chance of answering the item correctly (if guessing is taken into account); and
- pseudoguessing (a parameter related to the likelihood of guessing the correct answer), which estimates the probability that a respondent of very low ability answers correctly by chance.

Probability curves showing the probability of a correct response as a function of the three IRT parameters for the final versions of different items are shown later in the paper. (For an introduction to IRT, see Refs. [42,43]).

D. Pilot testing of a complete draft assessment

The next stage of our iterative development process involved a pilot test, which was administered via a proprietary assessment administration system to a group of 220 high school physics teachers across the country. Teachers were recruited through emails to local and national chapters of sections of the American Association of Physics Teachers, teacher mailing lists procured by Horizon Research, Inc., as well as advertisements in *The Physics Teacher*. Before administration of the pilot test, assessment experts on the team further vetted the items to ensure clarity, fairness, etc.

Results of the pilot test were used to further refine the assessment in the following ways:
- Test responses were analyzed using a 2PL-item response theory model. Each item was characterized

in terms of difficulty and ability to distinguish individuals in a reliable manner (discrimination index). Items that showed poor fit characteristics to the model and/or design expectations were further reviewed and either modified or removed from the assessment.

- Distractors were judged to be ineffective if they were selected by a very small number of teachers. Ineffective distractors were revised or replaced.
- Several constructed-response items were difficult to score or added little insight beyond what could be gained from a selected-response item. These items were also either revised to avoid scoring ambiguity or eliminated.

E. Field test of final CKT-E assessment

A revised final version of the assessment was administered via the same web portal to a group of 362 high school physics teachers. For all constructed-response items we developed scoring rubrics and iteratively found these rubrics to achieve interrater reliability of 90% or greater. An example of a scoring rubric for one of the constructed-response items is provided later in the paper (Sec. IV C). The revision process resulted in a psychometrically defensible assessment instrument. Using a 3-PL IRT[2] model [46], the assessment was able to reliably differentiate individuals' CKT ($r = 0.87$) across a broadly distributed range of scores (see Ref. [19] for full technical description). This allowed us to assign a single performance (ability) level theta to each of the teachers.

IV. DESCRIPTION OF REPRESENTATIVE FIELD TEST ITEMS

In this section we describe the design of three representative items that we have selected to illustrate specific purposes and characteristics of the complete assessment. Each of the following items includes two distinct questions based on the same instructional context.

Each of these items is aligned with one or more tasks of teaching and focuses on specific student energy targets, as shown in Table I. Furthermore, each item serves a specific rhetorical purpose, which we describe below.

- The *Two Blocks Item* showcases in detail how a given item context can be used to explore multiple tasks of teaching and how we interleave the assessment of

specific tasks of teaching with student energy targets. Specifically, the item focuses on ToT III: monitoring, interpreting, and acting on student thinking.

- The *Bouncing Basketball Item* illustrates how we assess the range in disciplinary and pedagogical knowledge that teachers use when interpreting and responding to learner ideas (CKT-D and CKT-P). Specifically, the item focuses on ToT II: designing, selecting, and sequencing learning experiences and activities.
- The *Puck Launcher Item* demonstrates how we assess teacher resources for interpreting student models and guiding students in the selection of experiments. Specifically, the item focuses on ToT VI: using experiments to construct, test, and apply concepts.

We also include a brief analysis of field test results for each of these items.

A. Two Blocks Item—Tasks of teaching and student energy targets

1. Overall design and structure of the item

Figure 1 shows a two-part item designed to assess both CKT-D and CKT-P.

The context for this item is typical of an instructional approach in which students use real-world data to construct generalizable scientific models. The teacher, Mr. Andreou, has challenged his students to compare the transfer of energy in two simple scenarios. In order for Mr. Andreou to help his students make sense of their experimental results he must apply the disciplinary knowledge needed to provide the most correct and complete account of the kinetic energies being equal. Specifically, he must realize that when equal forces are applied over equal distances the amount of energy transferred to the objects will be equal. This is the same disciplinary knowledge that Mr. Andreou is hoping to help his students apply. As a teacher, Mr. Andreou needs knowledge that goes beyond the knowledge his students need to select the correct answer (i.e., he needs to understand why the other answers are incomplete or incorrect and why students might be likely to choose them). In particular, Mr. Andreou needs to understand that compensatory reasoning alone does not provide a complete account for the equal kinetic energies of the blocks. The lighter block would have a higher speed, but this observation alone does not provide a complete account of the kinetic energies being equal. Based on the information provided in the item stem, only an analysis of work adequately explains the equal energies. If friction were not negligible, the frictional forces would be different and the net work done would be different; therefore, the kinetic energies would not be equal even though the lighter block would have a higher speed.

Recognizing the limitations of incomplete scientific arguments is a facet of content knowledge and one that is relevant for teaching (CKT-D). This is particularly salient

---

[2]The model selection was driven by model fit rather than heuristics about required number of examinees for each model. Although these rules were particularly important back when software options were limited, modern IRT software such as the MIRT package [44] in R statistical software [45] can easily apply priors to the guessing parameter in the 3-PL to ensure convergence. This resolves the well-known tendency for the guessing parameter to not converge, even in large samples.

TABLE I. Tasks of teaching and student energy targets for items featured in this paper.

Item	Task(s) of Teaching	Student Energy Target(s)
Two Blocks: Question 1	This question does not focus on a specific pedagogical task.	C3 Differentiates between energy and related concepts (e.g., force, power, stimulus, trigger, activation, speed, distance, temperature)
Two Blocks: Question 2	III. b) Interprets productive and problematic aspects of student thinking and mathematical reasoning III. c) Identifies specific cognitive and experiential needs or patterns of needs and builds upon them through instruction III. d) Uses interpretations of student thinking to support the teacher's instructional choices both in lesson design and during the course of classroom instruction	E4 Understands that conservation serves as a mathematical constraint on the outcomes of possible processes E5 Understands that the mathematical analysis of energy-related processes depends on the choice of initial and final state G5 Connects mathematical representations of energy to experiments
Bouncing Basketball: Question 1	This question does not focus on a specific pedagogical task.	E2 Understands the linear and non-linear mathematical relationships between forms of energy and the factors on which they depend
Bouncing Basketball: Question 2	II. a) Designs or selects and sequences learning experiences that focus on sense-making around important science concepts and practices, including productive representations, mathematical models, and experiments in science, that are connected to students' initial and developing ideas II. b) Includes key practices of science including experimentation, reasoning based on collected evidence, experimental testing of hypotheses, mathematical modeling, representational consistency, and argumentation	G3 Designs experiments to test competing hypotheses G4 Makes choices in data collection and analysis that allow for inferring the amounts and transfers of energy even when they cannot be measured directly G5 Connects mathematical representations of energy to experiments
Puck Launcher	VI. g) Encourages students to draw on experiments as evidence to support explanations and claims and to test explanations and claims by designing experiments to rule them out	G3 Designs experiments to test competing hypotheses G4 Makes choices in data collection and analysis that allow for inferring the amounts and transfers of energy even when they cannot be measured directly G5 Connects mathematical representations of energy to experiments

when a teacher attempts to help students assess and refine their own scientific arguments by suggesting new experiments for them to conduct, as is the case with question 2 of the Two Blocks Item. In order to correctly answer this question a teacher must differentiate one scenario in which the same exerted force will result in equal kinetic energy from three scenarios in which the result will be different kinetic energies. Responses A, B, and C would all result in different final kinetic energies for blocks of different masses. Only response D will result in the same kinetic energy for the two blocks. Both of these questions assess teacher resources for monitoring, interpreting, and acting on student thinking (ToT III). In order for a teacher to act on student thinking he should be able to identify both strengths and limitations in student reasoning by comparing that reasoning with correct and complete reasoning as in question 1. When student reasoning has limitations the

teacher should be able to identify cases in which the student's reasoning will lead to both correct and incorrect predictions, thereby, allowing the students' opportunities to refine their thinking as in question 2.

Question 1 assesses CKT-D, specifically differentiating between energy and related concepts (in this case, force [SET C3] and transfer of energy [SET D2]). Question 2 assesses CKT-P by addressing the following tasks of teaching: interpreting productive and problematic aspects of student thinking and mathematical reasoning (ToT III. b); and identifying specific cognitive and experiential needs or patterns of needs (ToT III. c); and then building upon these ToTs through instruction. Question 2 also addresses the related task of teaching of using interpretations of student thinking to support the teacher's instructional choices both in lesson design and during the course of classroom instruction (ToT III. d).

Mr. Andreou's class is in the middle of discussing possible factors that determine the change in kinetic energy of objects. Students have collected data in the following experimental setup:

Two blocks with different masses are free to slide on a very, very smooth table between two parallel lines. An air blower pushes each block horizontally, exerting the same constant force. Both blocks start from rest and cover the same distance on their track under the action of the air blower. The experimental data collected by the groups support the claim that the final kinetic energies of the two blocks are equal.

1. Of the following four student responses, which is the most correct and complete account of the kinetic energies of the two blocks being equal?

- The two blocks have equal final kinetic energies because the blower transfers to each of them equal amounts of energy per second. (11%)

- The two blocks have equal final kinetic energies because the higher final speed of the lighter block compensates for its smaller mass. (46%)

- The two blocks have equal final kinetic energies because the blower transfers to each of them equal amounts of energy per meter. (31%)

- The two blocks have equal final kinetic energies because when there is negligible friction, mechanical energy stays constant. (11%)

2. A student's written explanation states, "The two blocks have equal final kinetic energies when they cross the finish line because the blower pushed each block equally hard." If the student were to use similar reasoning to compare the final kinetic energies tor the two blocks in each of the variations of the experiment below, for which variation will the student's comparison of the final kinetic energies of the blocks be correct? Assume in each variation that the blocks start from rest.

- The experiment is repeated on the same very, very smooth table. The blower pushes the blocks with the same constant force for the same time interval. (17%)

- The experiment is repeated on a table that has small but not negligible friction. The blower exerts the same constant force on each block over the same distance. (13%)

- The very, very smooth table is slanted upward; the blower exerts the same constant force on each block uphill parallel to the track over the same distance. (8%)

- Instead of a blower, each block is pushed by the same compressed spring as the spring is released. The experiment takes place on the original very, very smooth table. (62%)

FIG. 1. Two Blocks Item: Percentages indicate proportion of teachers who chose a particular response.

2. Teacher performance on the item

Only 31% of teachers chose the correct response to question 1. Response B, which was selected by nearly one-half (46%) of the respondents, was the most popular response. Responses A or D were selected by about one-fifth of the respondents. We anticipated that this question would be challenging, guided by the results of an investigation on student understanding of the work-energy theorem conducted by the Physics Education Group at the University of Washington [47,48]. As part of that research, a similar physical situation (two blocks of different mass pushed by an air blower exerting the same constant force on a horizontal surface with negligible friction) was provided to students in a calculus-based university course. However, in that case, students were asked to predict the comparison of the final kinetic energies. After traditional instruction including standard laboratory, students had a very difficult time giving an accurate comparison with appropriate reasoning. In view of this research finding, we decided to assess teacher CKT-D with an easier question than the one used in the earlier research, namely, by providing respondents with the correct comparison and only asking for an explanation of this comparison.

Figure 2 shows item response characteristic curves for the two questions shown in Fig. 1. Question 009A had an estimated IRT 3-PL slope of 1.23, an estimated threshold of 1.08, and a pseudo-guessing parameter of 0.11 (the intercept of the curve with the probability axis). In comparison, question 009B had an estimated IRT 3-PL slope of 1.07, an estimated threshold of 1.01, and a pseudoguessing parameter of 0.51. Both slopes were within the range of reasonable values for test items. The difficulty indicated that the items were most informative for examinees approximately 1 standard deviation above the mean. Of particular note is the difference in pseudoguessing parameters, with responses to 009B being much noisier. Teachers were significantly more successful on the second question (CKT-P) than on the first, with 62% of teachers selecting the correct response. For the first question, teachers of average ability, $\Theta = 0$, had approximately a 25% probability of answering the question correctly. For the second question, even the teachers at the lowest ability levels had a 50% chance of answering the question correctly. For teachers 1 standard deviation above average
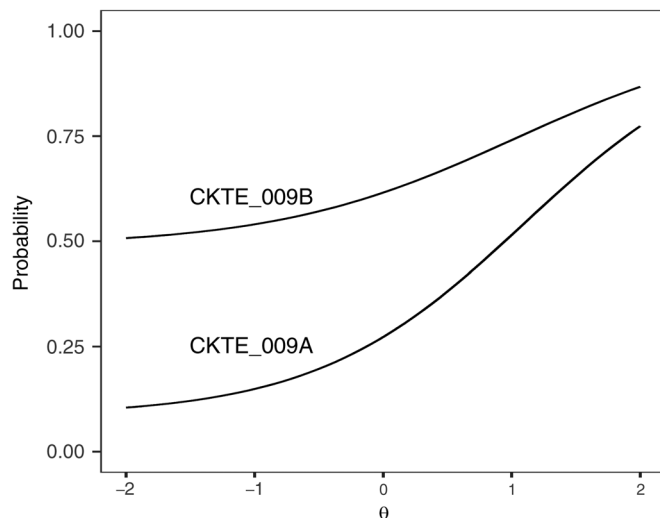
FIG. 2. Item response category characteristic curves for Two Blocks Item. The curves for each question show the probability that a teacher of a specified ability level, Θ, will answer the question correctly.

ability, $\Theta = 1$, the probability of answering the first and second questions correctly was approximately 45% vs 70%, respectively.

3. Implications of teacher performance on the item

The goal of this item was to evaluate teachers' ability to reason scientifically using the work-kinetic energy theorem and to measure teachers' ability to understand and build on student reasoning. We can explain the fact that most teachers chose distractor B in two ways: Either the teachers apply compensatory reasoning incorrectly (larger mass compensates for lower speed and, therefore, the kinetic energies are equal) or they confuse a description with an explanation. The inclusion of the word *because* in each answer option should have indicated to the teachers that the item was seeking an explanation of the data found by the students. Therefore, even if compensatory reasoning were correct for this item, it would not explain why the blocks have the same kinetic energy; it would just describe the difference in their speeds and masses while they have the same kinetic energy. Thus, the first implication for teacher preparation and professional development is to focus on the difference between an explanation and a description. The second question in the item assessed teachers' ability to build on student reasoning. We found that almost 40% of the teachers were unable to do so. Can the explanation be that they did not know how a student would determine the answer to the first question? There is very little literature on how to make activities that would help future teachers recognize the type of reasoning students utilized when answering certain questions. Our finding for the second question in the item suggests that such activities are absolutely necessary. (For a description of a disciplinary

course for preservice teachers that "uses metacognitive teaching strategies to promote the attainment of both disciplinary knowledge and pedagogical content knowledge," see Ref. [49].)

B. Bouncing basketball item: CKT-D vs CKT-P

1. Overall design and structure of the item

Figure 3 shows another two-part item designed to assess both CKT-D (Question 1) and CKT-P (question 2). Specifically, the first question asks whether the student has demonstrated an understanding of elastic energy and elastic force while the second asks to consider how a teacher can best help students understand these concepts.

The context of this item is also typical for a classroom environment. The teacher, Ms. Engel, hears the conversation between two students who are working on a problem. She needs to make a decision on how to help them resolve a typical difficulty. To be successful she first needs to recognize the physics nature of the difficulty and then choose (from the strategies suggested in the item) a productive instructional strategy.

Question 1 of the item assesses CKT-D—a teacher's disciplinary knowledge—specifically, the understanding of the difference in mathematical expressions for elastic force and elastic energy. In order to correctly answer the question a teacher needs to recognize that the ball and the floor exert forces on each other that are equal in magnitude and opposite in direction (Newton's third law). As the force magnitudes are the same and each is equal to the product of the object spring constant and the elastic deformation, the elastic deformation of the ball or floor is inversely proportional to the spring constant of the ball or floor. Thus, the higher the spring constant, the smaller the deformation (for a given force). It might look like the spring constant and the deformation of the object contribute equally to the forces that the objects exert. However, as the elastic energy is proportional to the square of the deformation, the object that has a greater deformation, namely the ball, will also have a larger amount of elastic energy (correct answer B). Understanding the difference between force and energy, as well as understanding the nature of mathematical relationships, is addressed by Student Energy Target E2: understanding the linear and non-linear mathematical relationships between forms of energy and the factors on which they depend. A teacher who meets this target will successfully answer this question. While meeting this target would be very useful to a teacher in this teaching context, she need not draw on any pedagogical insights in order to correctly answer this question.

Question 2 assesses CKT-P, or a teacher's ability to choose a productive strategy to help students figure out which of the objects does indeed have greater elastic energy when the ball compresses against the floor. Ms. Engel needs to acknowledge that the students, Marcos and Louisa, have known all of the concepts necessary to make

Two students in Ms. Engel's physics class are discussing the energetics of dribbling a basketball on a wooden floor. They agree that all of the kinetic energy gets converted into elastic energy for an instant when the basketball is compressed the most. They also agree that many objects can be modeled as springs, even basketballs and wooden floors. They are uncertain about whether there would be equal amounts of elastic energy in the ball and the floor. They call Ms. Engel over to share their ideas with her and get some help.

Marcos says, "We were thinking that when the ball compresses against the floor, the forces that the ball and the floor exert on each other would be equal and opposite, so maybe the amount of elastic energy in the floor is the same as the elastic energy in the ball."

Louisa responds, "I get that the forces are the same, but I am thinking that the ball compresses more than the floor, so shouldn't there be more energy stored in the ball?"

Marcos replies, "But the floor is more rigid and would have a higher spring constant. I think the larger k of the floor compensates for the smaller Δx in the $\frac{1}{2}k(\Delta x)^2$ equation, and the elastic energies are the same."

1. Is Marcos correct that the elastic energy of the ball and the floor would be the same?
    A.   Yes. (31%)
    B.   No. The elastic energy of the ball would be greater. (51%)
    C.   No. The elastic energy of the floor would be greater. (1%)
    D.   There is not enough information to compare these energies. (17%)

2. Which of the following activities would be most likely to provide Ms. Engel's students with additional insights about the relative amounts of elastic energy during the bounce of the basketball?
    A.   They could measure the spring constants and the displacements of both the floor and the ball and use those to compare the elastic energies. (10%)
    B.   They could compare the elastic energies of two non-identical springs when they are compressed with the same force. (70%)
    C.   They could do an experiment to see if a basketball bounces higher on a soft carpet surface or a hard concrete floor. (14%)
    D.   They could do an experiment to show that the same basketball will not bounce as high off the gym floor if it has first been put in a freezer. (6%)

3. Explain your selection and how the activity you selected might provide the students with additional insights about the relative elastic energy during the bounce of a basketball.

FIG. 3. Bouncing Basketball item. Percentages show teachers who chose a particular answer.

a quantitative comparison of the elastic energies. Together these students recognize that the interaction forces are equal, both objects may be modeled as springs with elastic energy given by ½$k\Delta x^2$, and the basketball will compress more than the floor. The teacher responding to this challenge needs to recognize that the students' ideas provide an ideal foundation for them to use a model of two springs with different spring constants compressed with the same force (answer B). These students could be encouraged to explore this model theoretically or experimentally to make a comparison of the elastic energies. The teacher could also recognize that the other possible answers either suggest experiments that are impossible to do or would not allow for a comparison of elastic energies.

To choose the correct answer the teacher needs to draw on the knowledge of the practice of physics (experimentation) and at the same time to apply that knowledge to anticipate the likely pedagogical results of various instructional activities. Picking the most productive activity for the basketball CKT-P question requires that a test subject think carefully about both the feasibility and the likely outcomes of the various candidate activities presented. However, a teacher might pick the right answer to this question without actually knowing the correct answer to question 1, as an appropriate experiment will help students come up with a correct decision based on the outcome of the experiment. The following ToTs address this ability: II. (a) designs or selects and sequences learning experiences that focus on

sense-making around important science concepts and practices, including productive representations, mathematical models, and experiments in science that are connected to students' initial and developing ideas; and II. (b) includes key practices of science including experimentation, reasoning based on collected evidence, experimental testing of hypotheses, mathematical modeling, representational consistency, and argumentation.

This combination of disciplinary knowledge and knowledge of teaching is what makes this item representative of the CKT-E assessment.

2. Teacher performance on the item

Figure 4 shows item response category characteristic curves for the Bouncing Basketball Item. The estimated IRT 3-PL parameters of Question 016A were a slope of 1.27, a threshold of 1.17, and a pseudo-guessing parameter of 0.33. Question 016B had an estimated slope of 1.68, a threshold of 0.92, and a pseudoguessing parameter of 0.29. These slopes were well within the range of acceptable values, and the difficulty parameters indicated that the items were appropriate for a large proportion of the examinees. The pseudoguessing parameters were high but not unreasonably so for an experimental assessment. For teachers with relatively low ability, $\Theta = -1$, the probability of answering both questions correctly was less than 50%. However, for teachers with average ability,
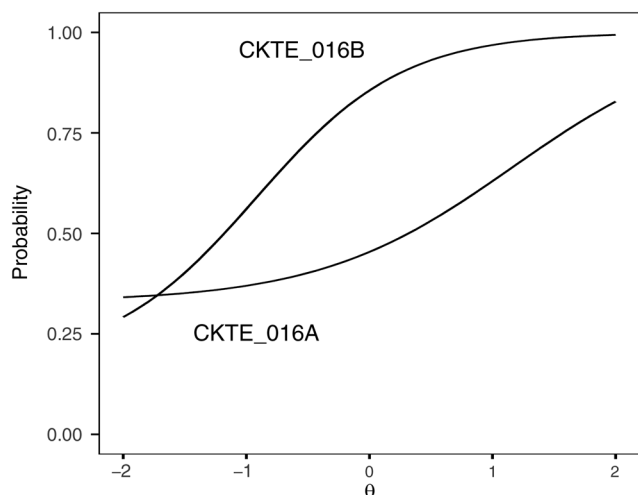
FIG. 4. Item response category characteristic curves for the Bouncing Basketball item.

$\Theta = 0$, there are substantial differences in the likelihood of answering each question correctly. While the likelihood of answering the first question correctly is still less than 50%, teachers at this ability level are very likely to answer the second question correctly. Even for teachers with high ability, $\Theta > 1$, the likelihood of responding correctly to the second question is substantially greater.

3. Implications of teacher performance on the item

Perhaps surprisingly, for this pair of questions we found many teachers who were able to correctly answer the second question even though they were not able to correctly answer the first one. In this case, it appears that selecting an appropriate pedagogical strategy is not contingent on the content knowledge necessary to predict the physics outcome of that pedagogical strategy. Many teachers would be able to guide the students toward a rigorous comparison of the elastic energies even though they made an incorrect energy comparison themselves. Possibly, if a teacher knows how to test different answers—quickly recognizing the correct answer is not necessary—the reasoning or experimentation will eventually lead there. In an upcoming paper we will further explore the complex relationship between supporting content knowledge and pedagogical reasoning.

C. Puck launcher item: Scaffolding scientific practices

1. Overall design and structure of the item

The *Next Generation Science Standards* (NGSS) present an ambitious vision of science classrooms where teachers guide students in authentic scientific practices. The development and testing of scientific models is the foundational thread in the practice of science. In order for teachers to guide their students in designing experiments to test competing hypotheses (Student Energy Target G3), teachers must be able to interpret student models and help

students design experiments that will critically test their models. The Puck Launcher Item is based on a classroom scenario in which students have been given an opportunity to connect energy ideas to real-life processes through experimental investigations (Student Energy Target G2). This item was developed to assess teacher resources for ToT VI. g: encouraging students to draw on experiments as evidence to support explanations and claims and to test explanations and claims by designing experiments to rule them out. This item includes two questions. Question 1 is a selected-response table, and question 2 is constructed response (Fig. 5).

2. Teacher performance on question 1: Interpreting student models

This item presents teachers with a situation in which there is a subtle difference between two student models. Both Jose and Sara agree that friction plays a significant role in slowing down the puck. They disagree about whether or not the air resistance will be significant. Jose thinks that friction is the dominant factor and that air resistance is negligible. Sara thinks that friction and air resistance are about equally important. Teacher responses to question 1 reveal that teachers were very successful in interpreting Jose's model. Ninety-six percent, 94%, and 92% of teachers were able to make correct predictions for the three experiments based on Jose's model, respectively. Teachers had more difficulty interpreting Sara's model. Only 67% of teachers recognized that Sara's model would also predict a significantly shorter distance on the rougher surface.

The item response curves for question 1 are shown in Fig. 6. Teacher performance on this question was categorized according to the number of correct predictions they identified: 1, 2, 3, 4, 5, or 6. Every teacher identified at least 1 correct prediction. This curve shows that teachers of above average overall ability were likely to get all six table predictions correct. In contrast, teachers of below average overall ability were likely to make at least one incorrect prediction. In the majority of these cases, these teachers had difficulty interpreting Sara's more nuanced model for the roles of friction and air resistance. We used these IRT curves to determine how teacher responses to this question would contribute to their overall assessed ability. The IRT curves for making one, two, three, or four correct predictions all showed similar variations with ability. Therefore, the first four correct table predictions did not influence the contribution of this item to our composite assessment of overall teacher performance. Only the fifth and sixth correct predictions contributed to the assessed teacher performance.

3. Teacher performance on question 2: Guiding experiment selection

Predicting outcomes of experiments based on student models is a necessary, but not sufficient, ability for guiding

The students in Ms. Rodriguez's class are conducting experiments in which they use a bent meter stick to launch a plastic puck across a smooth floor.

Ms. Rodriguez wants her students to:
  I.   Formulate a causal question about the scenario
  II.   Generate two or more different plausible models to answer their causal question
  III.   Design scientific investigations in order to rule out one or more of their models

Two students have decided that their causal question will be, "Why does the puck stop?" They are discussing plausible, energy-related reasons to explain why the puck stops.

> Jose: I think that almost all of the kinetic energy of the puck is converted into thermal energy in the puck and the floor as a result of friction.

> Sara: I agree about friction, but what about air resistance? I would think that about half of the kinetic energy is transferred to the air when the puck has to push the air out of the way.

> Jose: Yeah, I agree that a tiny bit would get transferred to the air. I just think that it is really small compared to the friction.

After a lot of discussion, the class converges on the following three experiments that they think might resolve the debate between Sara's model and Jose's model.

Rougher surface experiment – The students use a piece of cement board to perform the same experiment on a rougher surface.

Double height experiment – The students take a piece of paper twice the height of the puck, wrap it around the puck, and repeat the same experimental procedure.

Double pull experiment – The students pull the meter stick back twice as far and repeat the same experimental procedure.
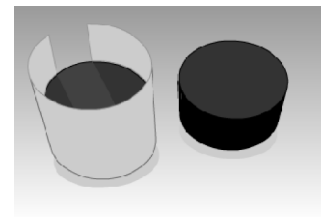
  1. **For the three experiments described, what is the expected outcome of the experiment based on each of Jose's and Sara's models. For each of the six cases, indicate whether the puck would be expected to travel: significantly farther, significantly less far, or about the same distance.**

	Predictions based on Jose's model			Predictions based on Sara's model		
	Significantly farther	Significantly less far	About the same	Significantly farther	Significantly less far	About the same
Rough surface experiment		(96%)			(67%)	
Double height experiment			(94%)		(91%)	
Double pull experiment	(92%)			(79%)		

  2. **Describe how you would use information such as that presented in the table to determine which experiments, if any, would be useful in resolving the debate between Jose and Sara.**

FIG. 5.   Puck launcher item. Percentages show teachers who chose a particular correct answer.

students in the selection of experiments. Teachers must also recognize that an experiment will only differentiate between models if the predicted results of the experiment are *different* based on the models. For example, based on the consistent model-based predictions shown in Fig. 5, only the double height experiment has different predicted outcomes based on Jose's and Sara's models. Therefore, only this experiment could help differentiate between the two student models. We assessed teacher understanding of this scientific practice based on their answers to the constructed-response question at the bottom of Fig. 5. It is important to note that even when teachers made incorrect predictions based on the student
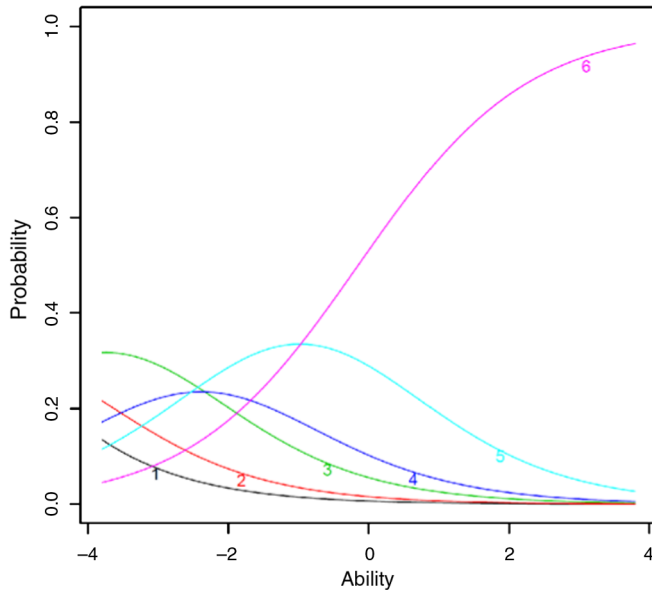
FIG. 6. Item response category characteristic curves for the puck launcher item. The number associated with each curve denotes the number of correct predictions (i.e., "6" means that the teacher made all six predictions correctly).

models they could still correctly apply this idea based on the predictions they selected. For example, if a teacher thought that in the rough surface experiment Sara's model would predict approximately the same distance and that Jose's

would not, they could correctly identify these experiments as being useful for resolving their debate.

We developed a 4-point scoring rubric to evaluate teacher performance on question 2. Figure 7 is a schematic showing the overall structure of this two-stage rubric. In the first stage, we evaluate whether teachers identified experiments for which they selected different predicted outcomes based on the two student models. In the second stage, we apply different criteria depending on whether the response satisfied the criterion in the first stage. For teachers whose response satisfied the initial criterion we determined whether their response explicitly or implicitly referenced the idea that an experiment must have different outcomes depending on which model is more correct. For teachers whose response did not satisfy the initial criterion we determined whether their response contained one or more features that were true and relevant to the question. Two example responses are provided in Fig. 7 to illustrate this rubric. On the left is the response of a teacher who only predicted different results for the double height experiment. This response, which scored a 4, satisfied the initial criterion and explicitly referenced the importance of different predicted outcomes. On the right is the response of a teacher who predicted different results for all three experiments. This written response did not satisfy the initial criterion. While the response is scientifically correct, it did not include ideas that are directly relevant to the proposed experiments or to the student models. This response scored a 1.
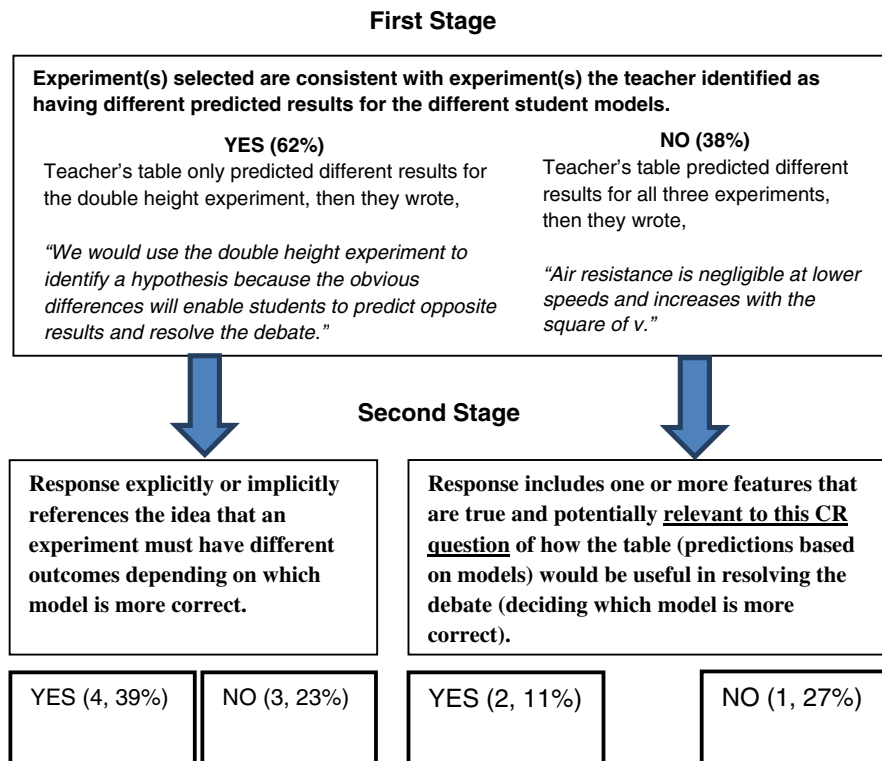

FIG. 7. Scoring rubric for the puck launcher constructed-response question.

Thirty-nine percent of teachers scored a 4 on this question by identifying a productive experiment(s) and articulating why that experiment(s) will discriminate between Jose's and Sara's models. This means that only 39% of all teachers were able to both fully apply and articulate this important scientific practice.

Experimental testability of models is at the heart of physics and is identified as a scientific practice within the NGSS. When teachers are able to interpret student models and help students select experiments to test those models they empower their students to be creators of scientific understanding. We found that most teachers were able to interpret student models and use them to make predictions. Teachers with above average overall assessed ability were able to interpret even subtle aspects of student models. However, only 39% of teachers were able to apply and articulate the idea that experiments must have different predicted outcomes for different models in order to discriminate between those models. This concept is foundational to the experimental enterprise. Teachers must be able to apply and articulate the concept in order to support their students when they strategically select their own experiments.

V. DISCUSSION

The goal of this paper is to describe a framework for measuring physics teachers' content knowledge for teaching energy (CKT-E) in the context of mechanics and to present an overview and examples of two kinds of questions, CKT-D and CKT-P, which help assess teachers' CKT-E. The framework is organized around Tasks of Teaching and Student Energy Targets. The tasks of teaching describe the activities in which teachers engage while teaching any content, and the student energy targets describe the disciplinary core ideas, science practices, and cross-cutting concepts that are important for student learning of energy in the context of mechanics. We described the process of creating and administering the assessment and presented examples of items with their analysis and teacher performance. These examples are indicative of the performance patterns shown on the assessment as a whole. Based on the analyses of items' IRT 3-PL curves we can say that the items reliably discriminate among teachers with strong CKT-E and weak CKT-E. Another specific finding is that teachers can sometimes successfully build on student reasoning and select an appropriate pedagogical strategy even in cases in which they themselves struggle with the disciplinary ideas. We also found that some teachers have difficulty recognizing the limitations of compensatory reasoning, applying mathematical models, and strategically selecting experiments to test conflicting hypotheses.

A limitation of our research is associated with the specific choices we made in the design of the items and the distractors. The items reflect what we consider to be important student learning targets and important teacher behaviors, informed by our reading of the literature. Therefore, they reflect our values and are informed by

our decades-long experience in the learning of physics and physics teacher education and professional development.

To find out whether the performance on the assessment correlates with other measures of CKT-E such as classroom teaching, teacher-designed assignments and assessments, as well as unit and lesson plans, we conducted an intensive study of 32 teachers who were among the 362 field test participants. From this group we collected a comprehensive data set including classroom videos of all lessons they taught in the energy unit, instructional artifacts, and assessments of student learning. Several articles that will describe relationships among these various data sets are in preparation. Preliminary analyses of the data indicate a positive correlation among CKT-E assessment performance, richness of instruction, and student learning.

One of the critical questions associated with CKT is whether it is a unique construct in and of itself or one that is simply a proxy for content understanding. Other research has found high correlations between measures of CKT and content, but this may be due to the fact that these studies have only considered teachers, for whom content knowledge may be a limiting factor on the level of one's CKT. In a companion study we ask whether we would find a smaller correlation with individuals who have similar content knowledge but no experience in teaching. When we tested a sample of undergraduate physics majors, individuals who have similar content preparation as the fraction of high school physics teachers who have an undergraduate major in physics or physics education, we found strong evidence that CKT is not reducible to content knowledge [50,51].

A caveat is in order. Although we have established empirically that CKT, as assessed in our instrument, is not reducible to pure content knowledge, we suspect that deep facility with the content (understood here to include the scientific practices of physics and the ways in which subdomains of physics knowledge are organized and interconnected), coupled with the ability to empathetically *imagine* ways in which learners might struggle with the material, could allow one to—in the moment—*construct* productive CKT from the prompts of the items. In practice, this is hard to do in time-sensitive, test-taking contexts, but it is definitely not impossible. Strictly speaking, then, CKT is strongly correlated with the work of teaching but need not be exclusively the purview of teachers, as very sophisticated, highly metacognitive individuals with deep grounding in the domain who are not teachers might conceivably also score highly on our assessment.

Nevertheless, thinking of the specialized knowledge of the discipline that teachers use in teaching as manifesting in the intersection of tasks of teaching and student energy targets is productive, valid, and generalizable. This approach can avoid the pitfalls of embedding a purely content question superficially in a classroom setting without changing its deep structure. Similarly, by indexing to a specific learning target, it avoids pedagogical questions that are agnostic about the intricacies of a specific area of the domain. Ultimately, it is a promising tool for those dimensions of teacher education and

professional learning programs whose goals are to help teachers improve their craft in disciplinarily rich ways.

APPENDIX A: TASKS OF TEACHING

This section provides a list of the tasks of teaching.

Task of teaching	Description	Specific tasks
I. Anticipating student thinking around science ideas	*While planning and implementing instruction teachers are able to anticipate particular patterns in student thinking. They understand and recognize challenges students are likely to confront in developing an understanding of key science concepts and mathematical models. Teachers are also familiar with student interests and background knowledge and enact instruction accordingly.*	*Teachers:* I. a) anticipate specific student challenges related to constructing scientific concepts, conceptual and quantitative reasoning, experimentation, and the application of science processes I. b) anticipate likely partial conceptions and alternate conceptions, including partial quantitative understanding about particular science content and processes I. c) recognize student interest and motivation around particular science content and practices I. d) understand how students' background knowledge both in physics and mathematics can interact with new science content
II. Designing, selecting, and sequencing learning experiences and activities	*Classroom learning experiences and activities are designed around learning goals and involve key science ideas, key experiments, and mathematical models relevant to the development of ideas and practices. Learning experiences reflect an awareness of student learning trajectories and support both individual and collective knowledge generation on the part of students.*	*Teachers:* II. a) design or select and sequence learning experiences that focus on sense-making around important science concepts and practices, including productive representations, mathematical models, and experiments in science that are connected to students' initial and developing ideas II. b) include key practices of science including experimentation, reasoning based on collected evidence, experimental testing of hypotheses, mathematical modeling, representational consistency, and argumentation II. c) address projected learning trajectories that include both long-term and short-term goals and are based on evidence of actual student learning trajectories II. d) address learners' actual learning trajectories by building on productive elements and addressing problematic ones II. e) provide students with evidence to support their understanding of short- and long-term learning goals II. f) integrate, synthesize, and use multiple strategies and involve students in making decisions II. g) prompt students to collectively generate and validate knowledge with others II. h) help students draw on multiple types of knowledge, including declarative, procedural, schematic, and strategic II. i) elicit student understanding and help them express their thinking via multiple modes of representation II. j) help students consider multiple alternative approaches or solutions, including those that could be considered to be incorrect

(Table continued)

APPENDIX A *(Continued)*

Task of teaching	Description	Specific tasks
III. Monitoring, interpreting, and acting on student thinking	*Teachers understand and recognize challenges and difficulties students experience in developing an understanding of key science concepts; understanding and applying mathematical models and manipulating equations; designing and conducting experiments, etc. This is evident in classroom work, talk, actions, and interactions throughout the course of instruction so that specific learning needs or patterns are revealed. Teachers also recognize productive developing ideas and problem solutions and know how to leverage these to advance learning.* *Teachers engage in an ongoing and multifaceted process of assessment, using a variety of tools and methods. Teachers draw on their understanding of learners and learning trajectories to accurately interpret and productively respond to their students' developing understanding.*	*Teachers:* III. a) employ multiple strategies and tools to make student thinking visible III. b) interpret productive and problematic aspects of student thinking and mathematical reasoning III. c) identify specific cognitive and experiential needs or patterns of needs and build upon them through instruction III. d) use interpretations of student thinking to support instructional choices both in lesson design and during the course of classroom instruction III. e) provide students with descriptive feedback III. f) engage students in metacognition and epistemic cognition III. g) devise assessment activities that match their goals of instruction
IV. Scaffolding meaningful engagement in a science learning community	*Productive classroom learning environments are community-centered. Teachers engage all students as full and active classroom participants. Knowledge is constructed both individually and collectively, with an emphasis on coming to know through the practices of science. The values of the classroom community include evidence-based reasoning, the pursuit of multiple or alternative approaches or solutions, and the respectful challenging of ideas.*	*Teachers:* IV. a) engage all students to express their thinking about key science ideas and encourage students to take responsibility for building their understanding, including knowing how they know IV. b) develop a climate of respect for scientific inquiry and encourage students' productive deep questions and rich student discourse IV. c) establish and maintain a "culture of physics learning" that scaffolds productive and supportive interactions between and among learners IV. d) encourage broad participation to ensure that no individual students or groups are marginalized in the classroom IV. e) promote negotiation of shared understanding of forms, concepts, mathematical models, experiments, etc., within the class IV. f) model and scaffold goal behaviors, values, and practices aligned with those of scientific communities IV. g) make explicit distinctions between science practices and those of everyday informal reasoning as well as between scientific expression and everyday language and terms IV. h) help students make connections between their collective thinking and that of scientists and science communities IV. i) scaffold learner flexibility and the development of independence IV. j) create opportunities for students to use science ideas and practices to engage real-world problems in their own contexts

(Table continued)

APPENDIX A *(Continued)*

Task of teaching	Description	Specific tasks
V. Explaining and using examples, models, representations, and arguments to support students' scientific understanding	*Teachers explain and use representations, examples, and models to help students develop their own scientific understanding. Teachers also support and scaffold students' ability to use models, examples, and representations to develop explanations and arguments. Mathematical models are included as a key aspect of physics understanding and are assumed whenever the term* model *is used.*	*Teachers:* V. a) explain concepts clearly, using accurate and appropriate technical language, consistent multiple representations, and mathematical representations when necessary V. b) use representations, examples, and models that are consistent with each other and with the theoretical approach to the concept that they want students to learn V. c) help students understand the purpose of a particular representation, example, or model and how to integrate new representations, examples, or models with those they already know V. d) encourage students to invent and develop examples, models, and representations that support relevant learning goals V. e) encourage students to explain features of representations and models (their own and others') and to identify/evaluate both strengths and limitations V. f) encourage students to create, critique, and shift between representations and models with the goal of seeking consistency between and among different representations and models V. g) model scientific approaches to explanation, argument, and mathematical derivation and explain how they know what they know. They choose models and analogs that accurately depict and do not distort the true meaning of the physical law and use language that does not confound technical and everyday terms (e.g., heat and energy). V. h) provide examples that allow students to analyze situations from different frameworks such as energy, forces, momentum, and fields
VI. Using experiments to construct, test, and apply concepts	*Teachers provide timely and meaningful opportunities throughout instruction for students to design and analyze experiments to help students develop, test, and apply particular concepts. Experiments are an integral part of student construction of physics concepts and are used as part of scientific inquiry in contrast with simple verification.*	*Teachers:* VI. a) provide opportunities for students to analyze quantitative and qualitative experimental data to identify patterns and construct concepts VI. b) provide opportunities for students to design and analyze experiments using particular frameworks such as energy, forces, momentum, field, etc. VI. c) provide opportunities for students to test experimentally or apply particular ideas in multiple contexts VI. d) provide opportunities for students to pose their own questions and investigate them experimentally VI. e) use questioning, discussion, and other methods to draw student attention during experiments to key aspects needed for subsequent learning, including the limitations of the models used to explain a particular experiment VI. f) help students draw connections between classroom experiments, their own ideas, and key science ideas VI. g) encourage students to draw on experiments as evidence to support explanations and claims and to test explanations and claims by designing experiments to rule them out

APPENDIX B: ENERGY-SPECIFIC STUDENT TARGETS (ENERGY-RELATED CONTENT AND PRACTICE IDEAS)

This section provides a list of energy targets for the students.

1. Connections of energy and everyday experiences

The student

1) uses energy ideas to interpret or explain everyday phenomena

2) recognizes the important role of internal energy in interpreting or explaining everyday phenomena

2. Choice of system

The student

1) recognizes that the energy accounting in a phenomenon depends on the choice of system

2) explains the relative advantage of a given system choice (i.e., relative ease of analysis)

3) recognizes that the choice of system determines whether springs or Earth do work (i.e., if the spring or Earth are in the system they do not do any work on the system, but the system can possess elastic or gravitational potential energy)

4) identifies and differentiates between forms of energy and other physics concepts

3. Identification of and differentiation between forms of energy and other physics concepts

The student

1) recognizes that energy cannot be observed directly and knows how different forms of energy correspond to different measurable physical quantities

2) recognizes and maintains a consistency of scale (microscopic or macroscopic) during energy analysis

3) differentiates between energy and related ideas (e.g., force, power, stimulus, trigger, activation, speed, distance, temperature)

4) distinguishes between forms of energy and energy transfers

4. Transfer of energy (environment → system; system → environment)

The student

1) recognizes that the energy of a system is always conserved but might not be constant

2) recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some cases

3) avoids double counting when analyzing processes involving work and energy

4) recognizes when to use compensatory models for tracking energy into and out of a system and when quantitative models are of limited use

5. Use of mathematics

The student

1) understands that when considering potential energy, it is important to think about the change. The zero level of potential energy is arbitrary, but the change is not. The energy of attraction is negative if the zero level is set at infinity.

2) can account for vector and scalar quantities in energy analysis

3) understands that work is a scalar quantity and the positive or negative sign of work does not indicate direction but addition or subtraction

4) connects forms of energy and the factors on which they depend through appropriate linear and non-linear mathematical relationships

5) applies conservation as a mathematical constraint on the outcomes of possible processes

6) recognizes that the mathematical analysis of energy-related processes depends on the choice of initial and final state and the choice of system

6. Use of representations

The student

1) selects/creates and uses appropriate verbal, mathematical, and graphical/pictorial representations (specific for energy, such as bar charts, energy diagrams, etc.) to describe, analyze, and/or communicate a physical situation or process

2) interprets different representations used to describe, analyze, and/or communicate a physical situation or process

3) understands the relationships between different representations of the same phenomenon and seeks consistency among different representations

4) understands standard technical representations and language used to communicate energy-related ideas

7. Use of science practices

The student

1) uses a range of representations to communicate ideas and illustrate or defend explanations

2) connects energy ideas to other learning and real-life processes and projects through experimental investigations, energy problem solutions, and engineering designs

3) designs experiments to test competing hypotheses

4) makes choices in data collection and analysis that allow for inferring the amounts and transfers of energy even when they cannot be measured directly

5) connects experiments and data to the mathematical representations of energy

6) evaluates and negotiates choices/options by considering the merits, limitations, and relative advantages of different engineering designs in terms of, for example, different choices of energy models for the same physical process

7) provides evidence-based arguments concerning energy processes and engineering designs

8) demonstrates consistency and coherence in model-based and evidence-based reasoning in making predictions and interpreting results

---

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

[2] L. S. Shulman, Those who understand: Knowledge growth in teaching, Educ. Res. **15**, 4 (1986).

[3] D. L. Ball, M. H. Thames, and G. Phelps, Content knowledge for teaching: What makes it special?, J. Teach. Educ. **59**, 389 (2008).

[4] H. C. Hill, S. G. Schilling, and D. L. Ball, Developing measures of teachers' mathematics knowledge for teaching, Elem. Sch. J. **105**, 11 (2004).

[5] Bill and Melinda Gates Foundation, Learning about teaching: Initial findings of the measures of effective teaching project, http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf, retrieved May 15, 2017.

[6] D. H. Gitomer and C. A. Bell, Evaluating teaching and teachers, in *APA Handbook of testing and assessment in Psychology Vol. 3 Test. Assess. Sch. Psychol. Educ.*, edited by K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, and M. C. Rodriguez (American Psychological Association, Washington, DC, 2013), pp. 415–444.

[7] R. Duit, Teaching, and learning the physics energy concept, in *Teaching and learning of energy in K-12 education*, edited by R. F. Chen, A. Eisenkraft, D. Fortus, J. Krajcik, K. Neumann, A. Nordine, and Jeffrey Scheff (Springer, Cham, 2014), pp. 67–85.

[8] N. D. Finkelstein, C. Turpen, S. Pollock, M. Dubson, S. Iona, C. Keller, and V. Otero, Evaluating a model of research-based practices for teacher preparation in a physics department: Colorado PhysTEC, in *AIP Conf. Proc.*, edited by P. R. L. Heron, L. McCullough, and J. Marx (AIP Press, Melville, NY, 2006), pp. 3–6.

[9] M. M. Keller, K. Neumann, and H. E. Fischer, The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest, J. Res. Sci. Teach. **54**, 586 (2017).

[10] J. Baumert, M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, S. Krauss, M. Neubrand, and Y.-M. Tsai, Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress, Am. Educ. Res. J. **47**, 133 (2010).

[11] S. Krauss, M. Brunner, M. Kunter, J. Baumert, W. Blum, M. Neubrand, and A. Jordan, Pedagogical content knowledge and content knowledge of secondary mathematics teachers, J. Educ. Psychol. **100**, 716 (2008).

[12] T. Mäntylä, Didactical reconstruction of processes in knowledge construction: Pre-service physics teachers learning the law of electromagnetic induction, Res. Sci. Educ. **42**, 791 (2012).

[13] T. Mäntylä, Promoting conceptual development in physics teacher education: Cognitive-historical reconstruction of electromagnetic induction law, Sci. Educ. **22**, 1361 (2013).

[14] T. Mäntylä and M. Nousiainen, Consolidating pre-service physics teachers' subject matter knowledge using didactical reconstructions, Sci. Educ. **23**, 1583 (2014).

[15] Y. Lehavi, E. Bagno, B.-S. Eylon, R. Mualem, G. Pospiech, U. Böhm, O. Krey, and R. Karam, Classroom evidence of teachers' PCK of the interplay of physics and mathematics, in *Key Competences in Physics Teaching Teaching and Learning, Springer Proceedings in Physics*, edited by T. Greczyło and E. Dębowska (Springer, Cham, 2017), Vol 190, pp. 95–104.

[16] E. Etkina, Pedagogical content knowledge and preparation of high school physics teachers, Phys. Rev. ST Phys. Educ. Res. **6**, 020110 (2010).

[17] D. L. Ball and H. Bass, Toward a practice-based theory of mathematical knowledge for teaching, in *Proceedings of the 2002 Annual Meeting of the Canadian Mathematical Educational Study Group*, edited by B. Davis and E. Simmt (CMESG/GCEDM, Edmonton, AB, 2003), pp. 3–14.

[18] S. Blömeke, R. Houang, and U. Suhl, TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models, IERI Monogr. Ser. Issues **4**, 109 (2011).

[19] G. Phelps and S. Schilling, Developing measures of content knowledge for teaching reading, Elem. Sch. J. **105**, 31 (2004).

[20] L. Kucan, S. Hapgood, and A. Sullivan Palincsar, Teachers' specialized knowledge for supporting student comprehension in text-based discussions, Elem. Sch. J. **112**, 61 (2011).

[21] J. F. Carlisle, R. Correnti, G. Phelps, and J. Zeng, Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading, Reading and writing **22**, 457 (2009).

[22] D. L. Ball, Bridging practices: Intertwining content and pedagogy in teaching and learning to teach, J. Teach. Educ. **51**, 241 (2000).

[23] National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (The National Academies Press, Washington, DC, 2012).

[24] National Research Council, *How People Learn: Bridging Research and Practice* (The National Academies Press, Washington, DC, 1999).

[25] National Research Council, *How People Learn: Brain, Mind, Experiences and School (Expanded Edition)* (The National Academies Press, Washington, DC, 2000).

[26] A. Van Heuvelen and X. Zou, Multiple representations of work–energy processes, Am. J. Phys. **69**, 184 (2001).

[27] B. A. Lindsey, Student reasoning about electrostatic and gravitational potential energy: An exploratory study with interdisciplinary consequences, Phys. Rev. ST Phys. Educ. Res. **10**, 013101 (2014).

[28] B. A. Lindsey, P. R. L. Heron, and P. S. Shaffer, Student ability to apply the concepts of work and energy to extended systems, Am. J. Phys. **77**, 999 (2009).

[29] B. A. Lindsey, P. R. L. Heron, and P. S. Shaffer, Student understanding of energy: Difficulties related to systems, Am. J. Phys. **80**, 154 (2012).

[30] G. Swackhamer, Cognitive resources for understanding energy, http://modeling.asu.edu/modeling/CognitiveResources-Energy.pdf, retrieved May 15, 2017.

[31] E. Brewe, Energy as a substancelike quantity that flows: Theoretical considerations and pedagogical consequences, Phys. Rev. ST Phys. Educ. Res. **7**, 020106 (2011).

[32] A. R. Daane, S. B. McKagan, S. Vokos, and R. E. Scherr, Energy conservation in dissipative processes: Teacher expectations and strategies associated with imperceptible thermal energy, Phys. Rev. ST Phys. Educ. Res. **11**, 010109 (2015).

[33] A. R. Daane, S. Vokos, and R. E. Scherr, Goals for teacher learning about energy degradation and usefulness, Phys. Rev. ST Phys. Educ. Res. **10**, 020111 (2014).

[34] R. E. Scherr, A. D. Robertson, L. Seeley, and S. Vokos, Content knowledge for teaching energy: An example from middle-school physical science, in *Proceedings of the Physics Education Research Conference 2013, Portland, OR*, edited by P. V. Englehardt, A. D. Churukian, and D. L. Jones (AIP, New York, 2014), pp. 321–324.

[35] A. R. Daane, S. Vokos, and R. E. Scherr, Learner understanding of energy degradation, in *Proceedings of the Physics Education Research Conference 2013, Portland, OR*, edited by P. V. Englehardt, A. D. Churukian, and D. L. Jones (AIP, New York, 2014), pp. 109–112.

[36] R. E. Scherr, H. G. Close, E. W. Close, V. J. Flood, S. B. McKagan, A. D. Robertson, L. Seeley, M. C. Wittmann, and S. Vokos, Negotiating energy dynamics through embodied action in a materially structured environment, Phys. Rev. ST Phys. Educ. Res. **9**, 020105 (2013).

[37] R. E. Scherr, H. G. Close, S. B. McKagan, and S. Vokos, Representing energy. I. Representing a substance ontology for energy, Phys. Rev. ST Phys. Educ. Res. **8**, 020114 (2012).

[38] R. E. Scherr, H. G. Close, S. B. McKagan, and S. Vokos, Representing energy. II. Energy tracking representations, Phys. Rev. ST Phys. Educ. Res. **8**, 020115 (2012).

[39] L. Seeley, S. Vokos, and J. Minstrell, Constructing a sustainable foundation for thinking and learning about energy in the twenty-first century, in *Teaching and Learning of Energy K–12 Education*, edited by R. F. Chen, A. Eisenkraft, D. Fortus, J. Krajcik, K. Neumann, J. Nordine, and A. Schef (Springer International Publishing, Switzerland, 2014), pp. 337–356.

[40] E. Etkina, Millikan award lecture: Students of physics—Listeners, observers, or collaborative participants in physics scientific practices?, Am. J. Phys. **83**, 669 (2015).

[41] Y. Lehavi and B.-S. Eylon, Integrating science education research and history and philosophy of science in developing an energy curriculum, in *History and Philosophy in Science Teaching* edited by M. R. Matthews (Springer, Cham, 2018), pp. 235–260.

[42] S. Embretson and S. Reise, *Item Response Theory*, Multivariate Applications Series (Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2013).

[43] R. K. Hambleton, H. Swaminathan, and J. Rogers, *Fundamentals of Item Response Theory*, Measurement Methods for the Social Science (Sage Publications, Inc., Newbury Park, CA, 1991).

[44] R. P. Chalmers, mirt: A multidimensional item response theory package for the R environment, J. Stat. Softw. **48**, 1 (2012).

[45] R Core Team, R: A language and environment for statistical computing, https://www.r-project.org/, retrieved May 15, 2017.

[46] D. Thissen and M. Orlando, Item response theory for items scored in two categories, in *Test Scoring*, edited by D. Thissen and H. Wainer (Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2001), pp. 73–140.

[47] R. A. Lawson and L. C. McDermott, Student understanding of the work-energy and impulse-momentum theorems, Am. J. Phys. **55**, 811 (1987).

[48] T. O'Brien Pride, S. Vokos, and L. C. McDermott, The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems, Am. J. Phys. **66**, 147 (1998).

[49] O. Eldar, B.-S. Eylon, and M. Ronen, A metacognitive teaching strategy for preservice teachers: Collaborative diagnosis of conceptual understanding in science, in *Metacognition in Science Education*, edited by A. Zohar and Y. J. Dori (Springer, Dordrecht, 2012) Vol. 40, pp. 225–250.

[50] G. Phelps, D. H. Gitomer, C. J. Iaconangelo, E. Etkina, L. Seeley, and S. Vokos, Assessing the Content Knowledge Used in Teaching Energy (to be published).

[51] C. J. Iaconangelo, G. Phelps, and D. H. Gitomer, Using Cognitive Diagnostic Modeling and Known Groups Analysis to Explore the Validity of the CKT Construct (to be published).