Teaching assistants' performance at identifying common introductory student difficulties in mechanics revealed by the Force Concept Inventory

Alexandru Maries[1] and Chandralekha Singh[2]

[1]*Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA*
[2]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
(Received 22 October 2015; published 23 May 2016)

The Force Concept Inventory (FCI) has been widely used to assess student understanding of introductory mechanics concepts by a variety of educators and physics education researchers. One reason for this extensive use is that many of the items on the FCI have strong distractor choices which correspond to students' alternate conceptions in mechanics. Instruction is unlikely to be effective if instructors do not know the common alternate conceptions of introductory physics students and explicitly take into account students' initial knowledge states in their instructional design. Here, we discuss research involving the FCI to evaluate one aspect of the pedagogical content knowledge of teaching assistants (TAs): knowledge of introductory student alternate conceptions in mechanics as revealed by the FCI. For each item on the FCI, the TAs were asked to identify the most common incorrect answer choice of introductory physics students. This exercise was followed by a class discussion with the TAs related to this task, including the importance of knowing student difficulties in teaching and learning. Then, we used FCI pretest and post-test data from a large population ($\sim$900) of introductory physics students to assess the extent to which TAs were able to identify alternate conceptions of introductory students related to force and motion. In addition, we carried out think-aloud interviews with graduate students who had more than two semesters of teaching experience in recitations to examine how they reason about the task. We find that while the TAs, on average, performed better than random guessing at identifying introductory students' difficulties with FCI content, they did not identify many common difficulties that introductory physics students have after traditional instruction. We discuss specific alternate conceptions, the extent to which TAs are able to identify them, and results from the think-aloud interviews that provided valuable information about why TAs sometimes select certain alternate conceptions as the most common, which are instead very rare among introductory students. We also briefly compare the performance of the TAs with physics instructors in identifying common introductory student difficulties with FCI content using the same task and find that their performances are comparable. Furthermore, our data suggest that the ability to identify introductory students' difficulties was not correlated with whether the TAs obtained their undergraduate degree in the U.S. or elsewhere.

I. INTRODUCTION

Graduate students in physics across the United States have been playing an important role in educating the next generation of students for a long time. In particular, in the U.S., it is quite common for physics graduate teaching assistants (TAs) to teach introductory physics recitation or lab sections, which typically have lower enrollments than the "lecture" component of the course (20–40 compared to 100 or more in a lecture). In addition to the graduate TAs, in the last two decades, undergraduate TAs (sometimes referred to as learning assistants or LAs) have also been helpful in assisting faculty members in teaching large classes. Appropriate professional development of these TAs to help

them perform their duties effectively is an important task. Physics education researchers have been involved in research on identifying common beliefs and practices among physics TAs that have implication for effective teaching [1–10]. For example, research suggests that sometimes graduate TAs struggle to understand the value of thinking about the difficulty of a problem from an introductory students' perspective and believe that if they know the material and can explain it to their students in a clear manner, it will be sufficient to help their students learn [1,2]. Also, while graduate TAs are able to recognize useful solution features and articulate why they are important when looking at sample introductory physics student solutions provided to them, they do not necessarily include those features in their own solutions written for introductory physics courses [3–5]. Moreover, the TAs do not always engage in grading practices which are conducive to helping introductory physics students learn expertlike problem-solving strategies and develop a coherent understanding of physics [6,7].

It is also important to keep in mind that TAs may be given varying amounts of freedom regarding how to perform their teaching duties depending on the instructor. However, discussions with the TAs who participated in this study and others from the University of Pittsburgh (Pitt) suggest that except for broader guidelines such as whether to discuss homework problems followed by a quiz or whether to have group problem solving [11–16] followed by a quiz in the recitation, the TAs often have reasonable flexibility in how to perform their recitation duties. For example, many instructors meet with the TA only briefly at the beginning of the semester to outline general guidelines, e.g., answer student questions on the homework or solve problems on the board, and the TAs are left to their own devices for the rest of the semester except for some communication with the course instructor via email or during the grading of the exams. Thus, if TAs are knowledgeable about effective instructional approaches, they can make a significant contribution to introductory students' learning of physics in the recitation because they often have sufficient flexibility to lead the recitation in a manner that they believe is conducive to student learning.

To help TAs learn about effective pedagogy, many institutions offer professional development programs that are sometimes discipline specific [8–10]. For more information about professional development programs and research on recruiting and educating future teachers, see Ref. [10] and references therein. The effectiveness of these professional development programs can be enhanced if those leading them are knowledgeable about TAs' conceptions regarding introductory physics students' difficulties [17]. For example, TAs may be largely unaware of certain introductory student alternate conceptions, and if professional development instructors preparing TAs discuss introductory students' alternate conceptions and engage the TAs in discussion about how to help introductory physics students learn, the TAs may be better prepared to conduct their teaching duties. It is even possible that in order to convince the TAs, the professional development instructors may have to share quantitative data on introductory physics student performance, which show that those alternate conceptions are common. This type of activity in TA professional development programs has the potential to enhance TAs' teaching effectiveness as they design, adopt, and adapt activities to build on students' prior knowledge and help them develop a robust knowledge structure so that there is less room for those alternate conceptions. Similarly, if TA professional development instructors are aware that TAs know about certain student alternate conceptions, those can only be discussed briefly.

Thus, by focusing on what TAs know and do not know and gradually building their *pedagogical content knowledge*, or PCK for short [18–19] (more about PCK in the next section), they can be guided to learn and implement effective pedagogy. These considerations motivated us to carry out the research study discussed here involving the most well-known assessment in physics education [20], namely, the Force Concept Inventory (FCI), to evaluate TAs' knowledge of introductory student alternate conceptions in mechanics as revealed by the FCI. For each item on the FCI, the TAs were asked to identify the most common incorrect answer choice of introductory physics students. This exercise was followed by a class discussion with the TAs related to this task, including the importance of knowing student difficulties and addressing them effectively in order for learning to be meaningful. We have found that this type of activity in a TA professional development class engenders a rich discussion about introductory student difficulties and promotes the importance of thinking about their difficulties from their perspective in order to bridge the gap between teaching and learning.

Before we move on to discuss our research, we first provide a background on research involving the FCI. We note, however, that a comprehensive review of the research that has been conducted with the FCI is beyond the scope of this article.

A. Background on previous research involving the Force Concept Inventory

The FCI is a multiple choice survey developed in 1992 by Hestenes, Wells, and Swackhamer and later revised [21] after many early observations made by Halloun and Hestenes [22] and other physics education researchers [23–26] that many students enter and leave physics classes with conceptions that are not consistent with the scientifically accepted concepts in physics. The FCI was designed to assess student understanding of basic mechanics concepts related to force and motion and has been widely used for this purpose by many educators and physics education researchers. Similar assessments in mechanics have been designed for the same purpose by other physics education researchers [27–31]. Although the conclusion that the FCI consistently measures Newtonian thinking was subject to some early debate [32–35], the general consensus is that the FCI score is a good indicator of Newtonian thinking [36–38]. Some researchers have investigated the validity of the items on the FCI using Item Response Analysis [36,38]. Morris *et al.* [36] have argued that Item Response Analysis can be used to identify answer choices which do not discriminate between students in different performance groups based on FCI scores. Their analysis was also used to investigate student performance in more detail and gain further insights into student difficulties for some of the items on the FCI. Other researchers studied the FCI using the Rasch model [39] and concluded that the FCI "has succeeded in defining a sufficiently uni-dimensional construct for each population" (non-Newtonian and predominantly Newtonian). The analysis by Planinic *et al.* suggested that "the items in the test all work together and

there are no grossly misfitting items which would degrade measurement" [39].

The FCI has played a key role in convincing many educators that traditional teaching methods that are primarily lecture oriented and do not engage students in the learning process actively do not promote conceptual and functional understanding [40]. Several studies have demonstrated that many students enter and leave introductory physics courses with the same alternate conceptions that are inconsistent with the accepted scientific ways of reasoning. Indeed, the use of the FCI in traditionally taught classes (even those taught by popular instructors) gave an impetus to the field of physics education research (PER) as educators increasingly realized that traditional methods were not working as intended. Consequently, they began to develop and evaluate instructional strategies designed to promote functional understanding of physical phenomena [41–45]. The FCI has often been used to assess whether a particular instructional strategy is effective in promoting conceptual understanding. Hake [40] used the FCI for this purpose and found that courses that make significant use of interactive engagement methods result in higher normalized gains on the FCI than courses that employ traditional methods such as standard lectures. Hake defined interactive engagement methods as those "designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors." [40]. For more details on the types of interactive engagement methods that resulted in higher normalized gains reported by Hake, see Refs. [40] and [46]. The average normalized gain is defined as the ratio of the change in the average post-test score (after instruction of Newtonian concepts) with respect to the average pretest score (before instruction of Newtonian concepts) to the average maximum possible change from the average pretest score, i.e., average normalized gain $\langle g \rangle = (\langle \text{postpercent} \rangle - \langle \text{prepercent} \rangle)/(100\% - \langle \text{prepercent} \rangle)$. Hake's study was performed on more than 6000 students and included both college and high-school classes.

Other researchers have used the FCI to investigate correlations between FCI scores and various other indicators of student performance: normalized gain on the FCI [47], problem solving ability [48], scientific reasoning ability [49], mathematics preparation [50], SAT scores [51], representational consistency [52], etc. In almost all these instances, significant positive correlations were found.

The FCI has often been administered by physics education researchers and curriculum developers as a pretest to determine what initial knowledge students bring to the learning of physics. Knowing the initial knowledge state of students is important because instructional tools and pedagogies can be designed to take advantage of the knowledge resources students have and to effectively address the alternate conceptions that are not consistent with the accepted scientific way of reasoning about physical phenomena. In addition, the FCI has been administered as a post-test to determine what concepts are difficult for students even after instruction and how effective instruction was at addressing student difficulties.

B. Focus of our research: Pedagogical content knowledge related to FCI content

There are several theoretical frameworks that inspire our research. These theoretical frameworks focus on the importance of instructors familiarizing themselves with students' prior knowledge (including what students learn from traditional instruction) in order to scaffold their learning with appropriately designed curricula and pedagogies. In the context of this study, they point to the importance of being knowledgeable about student difficulties in order to help students learn better. For example, Piaget [53] emphasized "optimal mismatch" between what the student knows and where the instruction should be targeted in order for desired assimilation and accommodation of knowledge to occur. A related framework is the theory of conceptual change put forth by Posner *et al.* [54]. In this framework, conceptual changes or "accommodations" can occur when the existing conceptual understanding of students is not sufficient for or is inconsistent with new phenomena. They also suggest that these accommodations can be very difficult for students, particularly when students are firmly committed to their prior understanding unless instructional design explicitly accounts for these difficulties. This model suggests that it is important for instructors to be knowledgeable about student ideas, which are commonly applied to inappropriate contexts to make inferences. Within this framework, students can be motivated by an anomaly that provides a cognitive conflict and illustrates how their conceptions are inadequate in explaining a newly encountered physical situation, and they can become dissatisfied with their current understanding of concepts and improve their understanding. But instructors must be aware of what conceptions students have and what difficulties in learning physics these conceptions can lead to in order to design instruction that produces the desired cognitive conflict.

Being knowledgeable of what conceptions students have and the difficulties that these conceptions may lead to is one aspect of what Shulman defined as Pedagogical Content Knowledge (PCK) [18,19]. Shulman defines PCK as the subject matter knowledge *for teaching*, or, in other words, PCK is a form of practical knowledge used by experts to guide their pedagogical practices in highly contextualized settings. Shulman writes "Within the category of pedagogical content knowledge, I include […] the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and

formulating the subject that make it comprehensible to others." In addition, according to Shulman, PCK also includes "an understanding of what makes the learning of specific topics difficult: the conceptions and preconceptions that students bring with them to the learning of those most frequently taught topics and lessons." [18]. Shulman developed the concept of PCK in response to the growing trend of proliferating general educational research in teacher preparation programs. The development of PCK was in part due to Shulman's previous research on the reasoning processes of physicians [55], which he found to be domain specific and contrary to the general assumption that certain physicians possess a general trait of diagnostic acumen which makes them better diagnosticians than others. Shulman generalized this observation to conclude that good teachers not only possess domain specific knowledge, but also possess more practical knowledge about teaching that is domain specific (i.e., PCK). Shulman therefore encouraged research on teachers' PCK and the types of teacher preparation programs that are likely to improve and/or develop teachers' PCK. Since Shulman introduced the concept of PCK, much has been written about it [56–71]. For example, Grossman [58] includes PCK as one of the "four general areas of teacher knowledge [which are] the cornerstones of the emerging work on professional knowledge for teaching: general pedagogical knowledge, subject matter knowledge, pedagogical content knowledge, and knowledge of context", and argues that PCK (as opposed to their subject matter knowledge) generally has the greatest impact on teachers' classroom activities. Others have also stressed the importance of PCK in shaping instructional practice and discuss professional development programs which take PCK into account [65,66]. For example, Borko and Putman [66] describe the Cognitively Guided Instruction Project, a multiyear program of curriculum development, professional development, and research which has shown "powerful evidence that experienced teachers' pedagogical content knowledge and pedagogical content beliefs can be affected by professional development programmes."

Given the importance of PCK in shaping instructional practices, it is not surprising that researchers have attempted to document teachers' PCK [60,62,63] and others have attempted to document the development of teachers' PCK [64,67]. However, these tasks are challenging to carry out for multiple reasons such as the fact that much of the knowledge teachers have of their practice is tacit [68,69], or the fact that although there is a general consensus among researchers on PCK as a construct, its boundaries are not clearly delineated [70]. Also, extended observations are needed in order to recognize when teachers' PCK is instantiated in their practice [60]. To overcome some of these challenges, researchers have often used multi-method approaches to investigate teachers' PCK. For example, observational data are not sufficient because a teacher may use only a small portion of the representations he or she has at his or her disposal. In addition, observations do not provide insight into teachers' instructional decisions—we see what they are doing, but do not know why. Partly due to these issues, Loughran et al. [60] used both classroom observations and follow-up interviewing of teachers. The interviews encouraged teachers to articulate their knowledge and explored alternative representations that the teachers did not use during the teaching sessions. This investigative approach is quite time consuming both to carry out and analyze since both the observations and interviews provide lengthy qualitative data which require coding and analysis. Baxter and Lederman [71] provide a review of methods and techniques for studying PCK and the subject matter knowledge of teachers.

Partly due to all of the difficulties in carrying out an involved investigation of PCK, we developed a relatively straightforward method for delving into one particular aspect of PCK, namely, knowledge of student difficulties with particular topics. This method makes use of standardized multiple choice tests developed by physics education researchers and quantitative data from students taking these tests. Teachers are given a copy of a particular test (e.g., FCI), and for each item on the test they are asked to select what they expect would be the most common *incorrect* answer choice students would select after being instructed in the relevant topic. Then, quantitative student data are used to quantify the extent to which teachers are knowledgeable about common difficulties students have, which are revealed by the incorrect answer choices students commonly select. Previous research with K–12 teachers [17] has found that on items that have a strong distractor (i.e., student alternate conception), there is a large difference in learning gains between students taught by teachers who could identify the alternate conception and students taught by teachers who could not. It is therefore valuable to explore the extent to which teachers are knowledgeable about student alternate conceptions on items drawn from carefully designed standardized tests.

In this study, we used the FCI to investigate the extent to which graduate students are knowledgeable about introductory students' alternate conceptions related to force and motion using the method described above. Knowledge of the introductory student alternate conceptions that graduate students are and are not aware of can be especially useful in designing effective professional development programs and inform future research on identifying and documenting the pedagogical content knowledge of TAs.

II. METHODOLOGY

A. Participants

The participants in this study were twenty-five first year graduate students enrolled in a semester long mandatory

pedagogy oriented TA training class, which meets once a week for two hours. The graduate student population at Pitt is consistent with that of a typical research focused state university and the nationality of the graduate students varied: nine TAs were from the United States, nine were from China, and the other seven were from other countries (Asian and European). The TAs teach introductory recitations and labs. The recitations and labs are typically taught in a traditional manner. In the recitations, the TAs primarily answer student questions, solve problems on the board, and give students a quiz in the last 10–20 min. In the labs, the TAs start by demonstrating the procedures needed for that lab and the students closely follow the detailed procedures written in the lab manual. Thus, the TA training course (which is required of all first-year graduate students) is not focused on helping the TAs implement physics education research (PER) based curricula in the recitations or labs in a particular semester (e.g., University of Washington tutorials [72]), but is a general introduction to pedagogical issues in physics teaching and learning.

Since this is the first and last pedagogy-oriented semester long course most graduate students will ever take, it is designed to help graduate students become more effective teachers in general. During the course, graduate students get a general overview of cognitive research and PER during one two hour session, and discuss their instructional implications. The graduate students are also introduced to curricula and pedagogies based on PER that emphasize the importance of being knowledgeable about introductory students' difficulties in order to help them develop expertise in physics. Each week students complete various reflective exercises designed to help them perform their TA duties in a student-centered manner. For example, in one class, they discuss how to write effective problem solutions for introductory physics classes and what features should be included in solutions they hand out to students [3–5]. In another class, they are given sample student solutions and asked to grade them individually and in groups followed by a discussion about how to grade students to help them learn better [6,7]. In the second half of the semester each graduate student also leads an interactive discussion of the solution of a physics problem in the class in the manner in which they would lead a discussion if they were teaching introductory students and receive feedback from the other graduate students in the class (who are asked to pretend to be introductory students and ask questions) and the instructor.

This study focuses on issues related to the professional development of TAs who teach recitations and labs for introductory physics courses and typically have a closer association with introductory students than the course instructors and may even be in a better position to help introductory students learn if they are versed in effective pedagogy. At Pitt, the TAs generally hold regular office hours and interact with introductory students in the physics resource room where they help introductory students with any questions related to their introductory physics courses. In addition, recitation class sizes are usually much smaller than the sizes of lecture classes taught by instructors. Therefore, TAs who are knowledgeable about introductory student difficulties related to force and motion can play a significant role in improving introductory student understanding of mechanics and they can address introductory students' difficulties directly in their interactions with students. Of course, it is also important for physics instructors to be knowledgeable about student difficulties in order to design instruction to effectively address them. Therefore, we compare the performance of thirty physics instructors at identifying common student difficulties on the FCI with that of the TAs.

All but three of the graduate TAs in this study were teaching introductory physics recitations or labs concurrently for the first time (the ones who were not teaching were supported by fellowships but were still required to take this mandatory TA training class). Two of the three who were not teaching had physics teaching experience as undergraduates, either as a classroom teaching assistant or as a tutor for students in introductory physics courses. Only one student did not have teaching experience in physics, but this student tutored several mathematics students as an undergraduate. Also, the TA training course included some discussions of introductory student difficulties, but not in the specific context of the FCI (until after the TAs completed all tasks related to the FCI as described below).

B. Materials

The materials used in this study are the FCI (which was given to the TAs in the TA training course as explained below), the pre-post introductory student data collected over a few years (2008, 2009, and 2012) from approximately 900 algebra-based and more than 300 calculus-based introductory physics students at Pitt (see Tables VIII and IX included in the Appendix), the quantitative data obtained from the TAs in the TA training class, and the follow-up interview data. These data were used to determine introductory students' common alternate conceptions on each item on the FCI, to assess the knowledge physics TAs have of introductory student alternate conceptions and to understand the reasoning TAs use when selecting certain incorrect answers as the most common. We analyze TAs' performance at identifying introductory student alternate conceptions at Pitt by utilizing student FCI data from the same institution, which is a large, typical state university with about 18 000 undergraduate students.

Here, we focus on the performance of physics graduate TAs enrolled in a TA training course in identifying common introductory student alternate conceptions related to the FCI content. We will therefore not discuss or compare the performances of different introductory student populations (e.g., performance of students in algebra-based and

calculus-based courses). All classes from which these data were collected were taught in a traditional manner and the average unmatched normalized gain (all students who took the pretest and post-test were included regardless of whether they took both the pretest and post-test) was 0.26 for the algebra-based classes and 0.36 for the calculus-based classes; nearly identical gains are obtained for the matched data. These gains are comparable to gains for courses that do not employ interactive engagement methods as reported by Hake [40].

C. Methods

The study was carried out towards the end of the semester so that the majority of TAs had at least one semester worth of teaching experience as recitation and lab instructors. In order to assess how knowledgeable TAs (and instructors) are of introductory student alternate conceptions related to the FCI, the following task was designed (referred to as the "FCI-related PCK task"). The TAs were given a copy of the FCI and asked the following: For each question, identify which one of the four *incorrect* answer choices, in your view, would be most commonly selected by introductory physics students after instruction in relevant concepts if the students did not know the correct answer. In fact, they were asked to complete three different tasks related to the FCI in the following order:

  (i) while working individually, they were asked to identify the correct answers for each FCI question;
 (ii) while working individually, they were asked to complete the FCI-related PCK task described above; and
(iii) they repeated the FCI-related PCK task, while working in groups of two or three.

The TAs were allowed as much time as they needed for each task. All TAs finished the first task within the first 30 min and the second task within the first hour. The third task (group work) was completed by all groups within 40 min followed by a full class discussion about the FCI-related PCK task and why knowledge of student difficulties is critical for teaching and learning to be effective in general, which took the remaining class time. The TAs were not prompted to explain their reasoning for their choices, but in the class discussion certain items on the FCI were discussed in detail and the TAs mentioned their reasoning about why they expected certain incorrect answer choices to be most common among introductory students. In order to obtain an in-depth account of the TAs' reasoning (related to why they expect certain answer choices to be most common among introductory students), think-aloud interviews were conducted with some TAs. These interviews corroborated the findings from the in-class discussion.

As noted, we wanted to compare the performance of the TAs with physics instructors at Pitt. Therefore, thirty physics instructors were given a copy of the FCI and asked to complete the FCI-related PCK task at their convenience.

After the instructors had completed the task, we discussed the reasoning for their responses individually with some of them.

We note that the task given to TAs and instructors was framed such that they had to identify the most common incorrect option for each multiple choice question that introductory physics students would select *after* instruction if they did not know the correct answer (rather than *before* instruction), because individual discussions with some faculty members who had taught introductory physics before giving them the task indicated that they felt that they had no way of knowing the "preconceptions" of introductory physics students at the beginning of the course. Their reluctance to contemplate introductory physics students' difficulties about force and motion before instruction motivated us to ask them to identify the most common incorrect answer choice for each question if the student did not know the correct answer *after* instruction in relevant concepts. We note that it does not make a significant difference whether the question is phrased about introductory physics students' difficulties with each question in the post-test or pretest because the common alternate conceptions of introductory physics students rarely changed after traditional instruction. Instead, typically, fewer introductory students held the same common alternate conceptions; this was found to be true when we compared the pretest and post-test data of introductory students, but this analysis is not presented here. Therefore, the performance of physics experts at identifying these alternate conceptions also provides an indication of their knowledge of the initial knowledge state of introductory physics students related to FCI content. Although asking them to identify the most common alternate conception in a post-test made the task easier to complete, some TAs (and faculty members) who participated in the study were very concerned about their ability to identify introductory physics students' difficulties and explicitly noted that they have no way of knowing the most common difficulty of introductory physics students for the FCI questions.

In order to compare the FCI-related PCK performance of the TAs with that of the physics instructors (and also to compare the FCI-related PCK performance of subgroups of TAs with different nationalities), scores were assigned to each TA and each instructor. A TA or instructor who selected a particular incorrect answer choice as the most common incorrect choice in a particular question received a PCK score which was equal to the fraction of introductory physics students who selected that particular incorrect answer choice. If a TA or instructor selected the correct answer choice as the most common incorrect answer (a rare occurrence), he or she was assigned a score of zero because he or she was explicitly asked to indicate the *incorrect* answer choice which is most commonly selected by introductory students if they did not know the correct answer. For example, in question 2, the fractions of algebra-based

students who selected A, B, C, D, and E are 0.44, 0.25, 0.06, 0.21, and 0.04, respectively (see Table VIII included in the Appendix). Answer choice A is correct, thus, the score assigned to TAs or instructors for each answer choice if they selected it as the most common incorrect answer would be 0, 0.25, 0.06, 0.21, and 0.04 (A, B, C, D, and E). The total score a TA or instructor would obtain on the task for the entire FCI can be obtained by summing over all of the questions (this is referred to as "FCI-related PCK score"). A mathematical description of how this calculation was performed is included in the Appendix. These scores can be used to compare the performance of TAs with that of instructors and can also be used to determine if TAs and/or instructors performed better than if they were randomly guessing. More details on how this was done are provided in the Appendix.

We note that the approach used to determine the FCI-related PCK score weighs the responses of TAs and instructors by the fraction of introductory students who selected a particular incorrect response. This weighting scheme was chosen because the more prevalent an introductory student difficulty is, the more important it is for a TA or instructor to be aware of it and take it into account in their instruction. Nearly all the questions on the FCI have only one strong distractor (corresponding to an alternate conception) with only three questions having two distractors (questions 2, 18, 26) as shown in Table I. For questions with two comparable distractors, the weighting scheme weighs equal distractors equally. For example, for question 1, there are two distractors, options B and D selected by 25% and 21% of introductory students. If a TA selects B, he or she would receive the maximum score for this question, 0.25, but if he or she selects D, the score received would be 0.21, which is close to the maximum possible PCK score of 0.25 (thus, the weighting scheme is such that they are not penalized for selecting either distractor).

The researchers jointly determined a heuristic that if a particular alternate conception is held by more than 20% of introductory students in a particular context, it is important for TAs to be aware of this alternate conception. Therefore, the discussion in this article is primarily focused on the questions on the FCI in which the introductory student data collected at Pitt suggested that at least 20% of students selected an incorrect answer choice. There were 21 such questions as shown in Table I. We note that one limitation of the approach is that a question with multiple distractors would be weighted less than a question with only one distractor but there are only three questions with more than one distractor.

D. Research questions and approach used to answer them

The following four research questions (RQ1-RQ4) were developed for the purpose of investigating the FCI-related PCK of the TAs. For each research question, we provide details about the methods used to answer it.

RQ1: (i) Are there situations in which a significant fraction of TAs select answer choices that very few introductory students select? What are some common examples of reasoning that TAs use to select these answer choices?

(ii) For what concepts/areas are TAs' responses context dependent (i.e., the majority of TAs identify the most common student alternate conceptions in some contexts, but not in others)?

(iii) For what concepts/areas do TAs systematically underestimate the most prevalent alternate conceptions of introductory students?

(iv) What alternate conceptions can TAs identify across contexts?

As noted, knowledge of introductory student alternate conceptions can be helpful in determining what pedagogical approaches may be effective to help students learn better. Therefore, TAs' knowledge of student difficulties can play an important role in improving student learning. In addition, TAs should also have reasonable expectations regarding how many introductory students have certain alternate conceptions. If a TA significantly overestimates the prevalence of a certain type of alternate conception (e.g., he or she thinks that 50% of students have the alternate conception, whereas the percentage of introductory students with that difficulty is less than 10%), the TA may spend considerable time and effort attempting to help students with something the majority of them already know, and thus not use class time effectively in addressing student difficulties which are more common. Similarly, if a TA underestimates the prevalence of a certain type of alternate conception, he or she is unlikely to consider instructional strategies to address it. This prompted us to investigate instances in which a significant fraction of the TAs select answer choices on the FCI that they believe are most commonly selected by introductory students while those answer choices are actually selected by very few introductory students, and instances in which few TAs identify common student alternate conceptions. Knowledge of what TAs think are the common alternate conceptions of introductory students, but which are instead not common, and what alternate conceptions are common among introductory students, but the TAs are not aware of them can be valuable for developers of teacher preparation programs because it can provide them with an understanding of the TAs' prior knowledge regarding introductory students' common reasoning patterns. Furthermore, it is possible that the TAs are able to identify a particular student alternate conception in one context, but not in another. Knowing which alternate conceptions TAs are able to identify and in which contexts can help developers of TA professional development programs build on TAs' prior knowledge of student difficulties and guide them to identify the same alternate conceptions in other contexts. Finally, knowing what

alternate conceptions TAs can readily identify across contexts can also inform professional development programs (e.g., by providing a starting point for discussing student alternate conceptions).

Tables I and II provide detailed data on TAs' performance in identifying the alternate conceptions of introductory students which are discussed in detail in the results section. In particular we discuss the following:

- Alternate conceptions which many TAs expected to be very common among introductory students, which were instead not common among introductory students,
- Alternate conceptions which were common among introductory students which were not identified very well by TAs,
- Alternate conceptions for which TAs' performance at identifying them was context dependent (on some questions TAs were able to identify them but not on other questions),
- Alternate conceptions that were common among introductory students which the majority of TAs were able to identify, and
- Qualitative results from detailed think-aloud interviews with five graduate students which focused on what common reasoning TAs used to select certain answer choices (e.g., answer choices which were not common among introductory students). Because of the availability of the TAs for individual interviews, some of the interviewed TAs were included in the quantitative data (because they were in the TA training class in which the study was carried out) but others were not.

RQ2: How does the performance of physics TAs on the FCI-related PCK task compare with that of physics instructors?

While on the average physics instructors had significantly more teaching experience as a lecturer and more overall physics content knowledge than the TAs (the physics instructors all had PhDs whereas the TAs were in their first semester of graduate school), the TAs taught recitations and labs, graded homework, quizzes, and sometimes exams. In addition, the TAs, who were all first year graduate students, had taken introductory physics as undergraduates roughly four years prior to the study. Therefore, it is difficult to predict *a priori* how TAs would perform compared to the instructors (most of whom did minimal grading and had minimal direct contact with students in the large introductory classes) regardless of the fact that instructors had significantly more teaching experience as a lecturer. Furthermore, we note that a detailed analysis of the instructors' performance on the FCI-related PCK task will be discussed in a future article.

In order to answer this research question, the PCK scores of TAs enrolled in the TA training course (in the latter part of the course), who had never taught an independent introductory physics course as a lecturer, were compared with the PCK scores of instructors (all of whom had taught some introductory physics course in the near or distant past and several had taught them many times).

RQ3: Are physics TAs from the United States, who have been exposed to undergraduate teaching in the U.S., better at identifying introductory students' alternate conceptions than international physics TAs?

Graduate programs across the U.S. are populated by many international graduate students. According to American Institute of Physics (AIP) statistics, in 2013 almost half of the first year physics graduate students in U.S. universities were non-U.S. citizens [73] and about half of the awarded physics PhDs were to non-U.S. citizens [74]. By comparison, in 1973 only 1/3 of the first year physics graduate students were non-U.S. citizens. A majority of physics departments in the U.S. require that graduate students become TAs for undergraduate courses at least for one or two semesters. Since the influence of international graduate students in physics undergraduate education seems to be commensurate with that of graduate students from the U.S. (at least in terms of numbers of TAs), it is worthwhile comparing the knowledge that these two different groups of graduate students have regarding introductory students' common difficulties with physics. The educational backgrounds of these two groups of graduate students are very different, and it is unclear whether these backgrounds have a significant effect on developing an understanding of the difficulties of introductory students with physics, in particular, with concepts related to force and motion in the FCI.

Out of the 25 TAs who participated in this study, nine were from the U.S., nine were Chinese, and seven were from other countries (Asia and Europe). The FCI-related PCK scores of three groups of TAs were compared (US, Chinese, and other international students). The reason we divided the TAs into three groups is because the TAs from the U.S. were exposed to teaching in the United States as opposed to the international TAs, who were not exposed to U.S. teaching practices before graduate school and most were taught physics in their own native languages. The nine Chinese TAs were placed in a separate group because, although they fit the category of international TAs, it is possible that their backgrounds are different from the backgrounds of most of the other international TAs, and it was unclear whether these differences in backgrounds would translate to differences in performance on the FCI-related PCK task.

RQ4: To what extent do TAs identify introductory students' difficulties more often when working in groups than when working individually (i.e., do discussions improve TAs' understanding of introductory students' alternate conceptions related to force and motion as revealed by the FCI)?

Previous studies have found that student discussions improve performance on conceptual examinations

[13,14,43]. Mazur's Peer Instruction approach has been developed to take advantage of the fact that student discussions are more likely to converge to the correct answers rather than the incorrect answers. In particular, research suggests that if two students individually select different answers and one of them is correct, the student with the correct answer is more likely to convince the student with the incorrect answer through a discussion than vice versa. In addition, in the context of introductory calculus-based electricity and magnetism, Singh found that even if both students initially select an incorrect answer choice, in 29% of the cases, discussions among introductory students lead to the correct answer [13,14]. We investigated whether discussions among TAs that are centered on introductory students' alternate conceptions helped them identify the more common alternate conceptions.

The TAs completed the FCI-related PCK task (i.e., they identified the incorrect answers they believe would be most commonly selected by introductory students) both individually and in groups of two or three. We investigated whether TA discussions improved their knowledge of introductory student alternate conceptions. Two factors would indicate that discussions improve TAs' understanding of introductory students' alternate conceptions:

(i) better FCI-related PCK performance when working in groups compared to working individually, and

(ii) convergence to a more common introductory student alternate conception.

The second factor warrants further explanation: suppose that in the individual FCI-related PCK task, two TAs selected two different incorrect answer choices (that they thought would be most common among introductory students who did not know the correct answer). If at least one of the incorrect answer choices is connected to a common student alternate conception, we investigated how often the two TAs agreed on the incorrect answer choice which is selected by more introductory students. In order to answer this question, we identified all the instances in which two (or three) TAs who selected different incorrect choices in the individual FCI-related PCK task, while working in a group, agreed on one of the incorrect answers. Then, we determined how often the incorrect answer selected in the group FCI-related PCK task was more common (by 5% or more) among introductory students than the other answers selected by the TAs in the individual FCI-related PCK task.

III. RESULTS

Many TAs (and instructors) noted that the task of thinking from an introductory physics student's point of view was challenging; some even confessed that they did not feel confident about their performance in identifying the most common incorrect answers. Also, the task was posed as the identification of the most common incorrect answer of introductory physics students for each FCI

question *after* instruction if students did not know the correct answer. Thus, the primary data analysis in this section involves comparison of TAs' (or instructors') responses with introductory physics students' responses on each FCI question *after* instruction. However, as noted, our analysis revealed that introductory students' alternate conceptions are generally the same, except less pronounced after traditional instruction. Additionally, the analysis of the FCI-related PCK performance was conducted with the student data from both algebra-based and calculus-based classes yielding nearly identical results. Therefore, below, we present the analysis with the student data in the algebra-based courses.

A. Performance of TAs in identifying introductory physics students' alternate conceptions related to the FCI

There are 21 questions (out of 30) on the FCI that reveal alternate conceptions held by 20% or more introductory students: items 2, 4, 5, 9, 11, 13, 15, and 17–30. Analysis of the FCI-related PCK score of the TAs was conducted on each of these questions and the results are displayed in Tables I and II.

Table I shows the questions on the FCI in which 20% or more introductory students selected a particular incorrect answer choice, the percentages of introductory physics students who answered each question correctly, the percentages of introductory students who selected each incorrect answer choice ranked from most to least common (they were asked to select the correct answer for each question), the three incorrect answer choices most commonly selected by TAs (they were asked to select the most common incorrect answer for each question if introductory physics students did not know the correct answer), and the percentages of TAs who selected these answer choices. Correct answers are indicated by the green shading in Table I, and incorrect answer choices selected by 20% or more introductory students are indicated by the red font. In addition, the second column (titled ">RG") in Table I indicates whether TAs performed better than random guessing (RG) in identifying introductory physics students' most common incorrect answer for a particular question: "Yes" or blank field indicate that TAs performed or did not perform better than random guessing.

For each question on which more than 20% of introductory students had a particular alternate conception, Table II shows the normalized average FCI-related PCK scores of the TAs. Their scores were normalized on a scale from zero to 100 because for each question on the FCI, there is a minimum and a maximum possible score, which correspond to the smallest and largest fractions of introductory students who selected a particular incorrect answer choice among the four incorrect answer choices. The normalization was done in the following manner:

$$\text{normalized score} = 100 \times \frac{\text{average FCIrelatedPCK score- minimum possible score}}{\text{maximum possible score-minimum possible score}}.$$

The normalized FCI-related PCK score is zero if the TAs obtained the minimum possible score and 100 if they obtained the maximum possible score. This also provided a means to compare TAs' FCI-related PCK performance for different questions which have different minimum and maximum possible FCI-related PCK scores. Table II also shows the difficulty of each of these questions via the percentage of introductory algebra-based students who

TABLE I. Questions on the FCI on which at least 20% of introductory algebra-based students selected one incorrect answer choice in a post-test (questions 12 and 16 marked with an * had 19% of students selecting one answer choice, but they are included here because on these questions many TAs selected incorrect answer choices as most common among introductory students, but very few introductory students selected those answer choices), percentages of introductory algebra-based physics students who answered the questions correctly in a post-test, percentages of introductory students who selected each incorrect answer choice ranked from most to least common (they were asked to select the correct answer for each question), the three most common incorrect answer choices selected by TAs (they were asked to select the most common incorrect answer for each question if introductory physics students did not know the correct answer). The first column of the table lists the FCI question numbers and the second column titled ">RG" shows a "Yes" when the TAs on average performed better than random guessing (RG). Note that answer choices marked in red (e.g., 25% (B) for question 2) indicate that 20% or more introductory students selected that answer choice. The same answer choices are also colored in the case of the TAs to make them stand out (i.e., one should not expect that for the TAs, the red coloring is used for the questions selected by 20% or more of the TAs—the introductory student percentages determine the coloring).

FCI Question No.	>RG	Introductory student choices					TA choices		
		Correct	Rank of incorrect answer choices				First three most common incorrect answer choices		
			1st	2nd	3rd	4th	1st	2nd	3rd
2	Yes	44% (A)	25% (B)	21% (D)	6% (C)	4% (E)	68% (B)	24% (D)	4% (E)
4	Yes	60% (E)	39% (A)	1% (B)			84% (A)	12% (D)	4% (C)
5		24% (B)	44% (D)	17% (E)	12% (C)	3% (A)	40% (D)	32% (C)	28% (E)
9	Yes	53% (E)	20% (B)	19% (C)	5% (D)	4% (A)	76% (C)	12% (A)	8% (B)
11	Yes	35% (D)	48% (C)	8% (B)	5% (A)	4% (E)	60% (C)	20% (B)	20% (A)
12*	Yes	77% (B)	19% (C)	2% (D)	1% (A)	1% (E)	68% (C)	28% (A)	4% (E)
13		35% (D)	50% (C)	11% (B)	4% (A)		44% (B)	36% (A)	16% (C)
15		44% (A)	48% (C)	7% (B)	1% (D)		40% (C)	28% (D)	20% (B)
16*		73% (A)	19% (C)	4% (E)	2% (B)	2% (D)	36% (E)	28% (D)	16% (C)
17	Yes	24% (B)	62% (A)	10% (D)	3% (E)	1% (C)	72% (A)	16% (D)	12% (E)
18	Yes	28% (B)	42% (D)	25% (E)	4% (C)	1% (A)	52% (A)	28% (E)	16% (C)
19	Yes	51% (E)	29% (D)	12% (A)	5% (C)	3% (B)	76% (D)	12% (C)	8% (A)
20	Yes	49% (D)	27% (C)	16% (A)	4% (B)	4% (E)	56% (C)	40% (A)	4% (B)
21		33% (E)	38% (C)	13% (B)	9% (D)	7% (A)	44% (C)	36% (D)	20% (B)
22		45% (B)	33% (A)	16% (D)	3% (C)	2% (E)	40% (D)	28% (A)	4% (C)
23		39% (B)	23% (D)	18% (C)	15% (A)	5% (E)	28% (C)	24% (D)	24% (E)
24		70% (A)	22% (C)	5% (E)	2% (B)	2% (D)	68% (C)	16% (B)	12% (D)
25		23% (C)	53% (D)	12% (E)	9% (B)	3% (A)	44% (D)	24% (B)	24% (A)
26		14% (E)	41% (A)	32% (B)	9% (D)	3% (C)	52% (A)	32% (B)	12% (D)
27	Yes	58% (C)	26% (A)	13% (B)	2% (C)		68% (A)	16% (B)	8% (E)
28	Yes	59% (E)	32% (D)	6% (C)	2% (B)	1% (A)	52% (D)	36% (B)	8% (C)
29		71% (B)	23% (D)	3% (C)	2% (A)	1% (E)	44% (E)	28% (D)	20% (A)
30	Yes	26% (C)	59% (E)	10% (B)	3% (A)	2% (D)	88% (E)	4% (B)	4% (A)

TABLE II.   Questions on the FCI on which at least 20% of introductory algebra-based students selected one incorrect answer choice in a post-test (questions 12 and 16 marked with * had 19% of students selecting one answer choice, but they are included here because on these questions many TAs selected incorrect answer choices as most common among introductory students, but very few introductory students selected those answer choices), percentages of introductory algebra-based students who answered each question correctly (% intro. alg. correct), normalized gain (intro. alg. norm. gain), minimum possible FCI-related PCK score (min. pos. PCK score), maximum possible FCI-related PCK score (max. pos. PCK score), average FCI-related PCK score of TAs (TA avg. PCK score), and normalized average FCI-related PCK score of TAs (Norm TA avg. PCK score).

FCI item No.	% intro. alg. correct	Intro. alg. norm. gain	Min. pos. PCK score	Max. pos. PCK score	TAs	
					TA avg. PCK score	Norm TA avg. PCK score (%)
2	44	0.23	0.04	0.25	0.22	86
4	60	0.46	0	0.39	0.33	85
5	24	0.16	0.03	0.44	0.26	56
9	53	0.17	0.04	0.20	0.17	81
11	35	0.24	0.04	0.48	0.31	61
12*	77	0.43	0.01	0.19	0.13	67
13	35	0.27	0	0.50	0.14	28
15	44	0.25	0	0.48	0.21	44
16*	73	0.40	0.02	0.19	0.05	18
17	24	0.17	0.01	0.62	0.47	75
18	28	0.19	0.01	0.42	0.30	71
19	51	−0.06	0.03	0.29	0.24	81
20	49	0.25	0.04	0.27	0.22	78
21	33	−0.02	0.07	0.38	0.23	52
22	45	0.22	0.02	0.33	0.16	45
23	39	0.14	0.05	0.23	0.15	56
24	70	0.20	0.02	0.22	0.16	70
25	23	0.13	0.03	0.53	0.27	48
26	14	0.11	0.03	0.41	0.33	79
27	58	0.09	0	0.26	0.20	77
28	59	0.6	0.01	0.32	0.18	55
29	71	0.50	0.02	0.23	0.07	24
30	26	0.16	0.02	0.59	0.53	89
x	TAs' FCI-related PCK score is less than 50%.					
x	TAs' FCI-related PCK score is between 50% and 67%.					
x	TAs' FCI-related PCK score is more than 67%.					

answered each question correctly in a post-test and their normalized gain. The researchers jointly determined a heuristic that the performance of the TAs was "good" (and shaded green in Table II) if their normalized FCI-related PCK score was more than 2/3 of the maximum possible score, "average" (and shaded yellow) if their normalized score was between 1/2 and 2/3 of the maximum possible score, and "poor" (gray shading) if their

normalized score was less than 1/2 of the maximum possible score.

We note that there are two questions included in Tables I and II in which 19% of the introductory students had a particular alternate conception (questions 12 and 16, marked with an *), but these questions were considered insightful from the point of view of FCI-related PCK by the researchers. We discuss these questions because a significant fraction of TAs overestimated the prevalence of

alternate conceptions that were very rare among introductory students.

B. Results relevant to each research question

RQ1: (i) Are there situations in which a significant fraction of TAs select answer choices that very few introductory students select? What are some common examples of reasoning that TAs use to select these answer choices?

There were 21 questions on the FCI in which introductory students had strong alternate conceptions (i.e., more than 20% of introductory physics students selected one particular incorrect answer choice). As shown in Table II, in ten of those questions, TAs' normalized FCI-related PCK scores were not good (less than 2/3 of the maximum possible). Out of those ten questions, for half of them, TAs' FCI-related PCK scores were average (between 1/2 and 2/3 of the maximum possible normalized score) and for the other half they were poor (less than 1/2 the maximum possible normalized score). We now turn to discussing questions in which TAs' PCK performance was either poor or average and focus on the questions which revealed that many TAs expected introductory students to have certain alternate conceptions, while few introductory students actually had those alternate conceptions. As mentioned earlier, it is important for TAs to know that introductory (college level) students do not have certain alternate conceptions, because if they overestimate the prevalence of these alternate conceptions (which are not common amongst introductory students), they may spend a significant amount of time trying to address issues that are not critical for the students.

As noted, to identify common examples of reasoning TAs used to select certain answer choices, think-aloud interviews were conducted with five graduate students who all had more than two semesters' worth of teaching experience in introductory physics recitation classes. The interviews focused on questions on which the quantitative data suggested that TAs are likely to select answer choices which were not at all prevalent among introductory students (i.e., TAs expected the majority of introductory students to select a particular incorrect answer choice for a particular question, but very few introductory students selected it). These formal interviews corroborated the information we had obtained from the informal class discussion centered on different FCI questions and student alternate conceptions and were valuable in obtaining an in-depth account of how some TAs reason about the FCI-related PCK task in various contexts.

1. Newton's 2nd law: Impact of net force on velocity

A common student alternate conception is that the effect of a nonzero net force acting on an object is to make the object move at constant velocity. On Questions 22 and 25, which uncovered this alternate conception, the FCI-related PCK score of the TAs was very low (45% and 48%

respectively), in large part because many TAs expected that certain incorrect answer choices would be very common among introductory students, when in fact they were not. Question 22, for example, states that as a rocket moves from position "b" to position "c" in space, its engine produces a constant thrust (force on the rocket) and asks students about the speed of the rocket between "b" and "c" (neglecting all other forces). On this question, the most common alternate conception (held by 33% of students) was that the speed of the rocket is constant (answer choice A), and only 16% of introductory students thought that the speed of the rocket increases for a while and is constant thereafter (answer choice D). However, 40% of the TAs expected the latter alternate conception to be most common among introductory students and only 28% of the TAs identified the most common alternate conception.

In question 25, a woman is pulling a box across the horizontal floor. Here, the most common alternate conception was that the force the woman is exerting must be larger than the force which resists the motion of the box (53% of introductory students selected this in a post-test). On this question, 48% of TAs expected that introductory students would think that the force the woman is exerting should be compared to the weight of the box to determine if the box moves at a constant velocity even though those two forces act in orthogonal directions. These TAs selected answer choices A and B, which state that in order for the box to move at constant velocity, the force the woman is exerting must be larger than or the same as ($_{A \text{ or } B}$) the weight of the box. On the other hand, only 12% of introductory students selected one or the other of these answer choices in the post-test (see Table I). The percentage of introductory students who selected A and B on the pretest was virtually the same (11% as shown in Table VIII included in the Appendix), and, therefore, on this question nearly half the TAs were unaware of introductory students' preconceptions.

Nearly all interviewed graduate students expected that introductory students would be equally likely to select answer choices A and B in which comparisons are being made between the force exerted by the woman and the weight of the box as they are to select answer choice D, which incorrectly states that the woman must exert a larger force than the total force which resists the motion of the box. However, only 12% of the introductory students fell in the former category, while 53% fell in the latter. Interestingly, the percentage of TAs who selected answer choice A plus the percentage of TAs who selected answer choice B is equal to the percentage of TAs who selected answer choice D as the most common incorrect student answer choice. One interviewed graduate student stated

"Maybe D would be most popular because A and B would split the vote."

When asked to clarify what he meant he said that he expects that the percentage of introductory students who select option D would be roughly equal to the percentage of introductory students who select options A and B, and therefore, answer choice D would be most popular and stated

"I have a hard time picking between D and A and B combined."

Thus, interviewed graduate students often expected that many introductory students would incorrectly compare the horizontal force exerted by the woman with the weight of the box acting vertically. When asked why, one graduate student stated

"If you're pushing a box, I think you're very likely to think about how heavy the box is. They [students] will say 'oh, this is heavy'. They won't say 'oh, μ is large and it's heavy.' So they might think that the thing to compare it [the force exerted by the woman] to is the weight."

2. Newton's 3rd law

For Newton's third law, the TAs struggled to identify the common student alternate conception that larger and/or active objects exert larger forces on smaller and/or passive objects in contexts other than a small car colliding with a truck. In these less typical contexts of Newton's third law, namely, questions 15, 16, and 28, TAs' normalized FCI-related PCK score was either poor (44% and 18% for questions 15, and 16) or average (55% for question 28) as shown in Table II. In each of these questions, there was only one distractor prevalent amongst introductory students. On the other hand, many TAs identified incorrect answer choices that were rarely selected by introductory students. For example, in question 15, a car is pushing on a truck from behind causing it to accelerate. The most common alternate conception held by 48% of introductory students was that the car is exerting a larger force on the truck than vice versa. Almost no introductory students (3% in the pretest, 1% in the post-test) thought that the car exerts a force on the truck, but the truck does not exert a force on the car, and that the truck is being pushed forward because it is in the way of the car (answer choice D, see Table VIII). However, 28% of TAs selected this answer as the most common among introductory students (see Table I). Additionally, very few introductory students thought that the car exerts a smaller force on the truck than vice versa (answer choice B selected by 10% of introductory students in the pretest and 7% in the post-test as shown in Table VIII), but 20% of TAs believed that this type of thinking is the most common among introductory students (see Table I).

Interviews revealed that sometimes graduate students got distracted by the different answer choices which caused

them to select incorrect answer choices that were not common among introductory students. For example, on question 15, in answer choice D, "the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car," an explanation is provided for how forces arise (i.e., they stem from engines). Some graduate students thought that this detail would seem convincing to introductory students and that many of them would select this answer choice as the correct one. One interviewed graduate student motivated his choice of D as the most common incorrect answer among introductory students in question 15 by saying

"D [has an] explanation that the professor hasn't had a chance to give to students [the relation between engine and force] whereas C might be a statement that the professor had specifically driven in to students as not right. [So an introductory student might think:] 'Professor never talked about the engine, because the engine's giving the force right?' So they [introductory students] might fall for it that way."

On question 16, TAs had the lowest normalized FCI-related PCK score (18%) out of all the questions on the FCI because 80% of them selected answer choices which were only selected by a total of 8% of introductory students both in the pretest and in the post-test (answer choices C, D, and E as shown in Tables I and VIII). The most common answer choice of introductory students was the same as in question 15, namely, that the car exerts a larger force on the truck than the truck exerts on the car (answer choice C). In contrast, 36% of the TAs selected answer choice E as the most common incorrect answer choice among introductory students which states that neither the car nor the truck exert forces on each other and that the truck is pushed forward because it is in the way of the car, but only 4% of introductory students selected this answer choice in a post-test (1% in the pretest). Similar to question 15, 28% of TAs believed that introductory students think that the car exerts a force on the truck but the truck does not exert a force on the car (choice D selected by 2% of introductory students in a post-test) and 12% of TAs believed students think that the car exerts a smaller force on the truck than vice versa (choice B selected by 2% of introductory students).

On this question, the interviewed graduate students sometimes claimed that answer choice D (car exerts a force on truck but truck does not exert a force on car) would be even more attractive to introductory students than in question 15 for similar reasons (the explanation given about the engine creating a force). However, similar to question 15, only 2% of introductory students selected D as the correct answer. Some interviewed graduate students claimed that on question 16, answer choice E would be most common among introductory students (neither the car nor the truck exert any force on one another. The truck is

pushed forward simply because it's in the way of the car). For example, one interviewed graduate student reasoned that introductory students would select answer choice E in question 16 because

> "Now that it's at constant speed they [students] think that […] there's no acceleration so there's no forces on each other."

Even those interviewed graduate students who correctly identified C (car exerts larger force) as the most common introductory student incorrect answer choice in question 16 sometimes claimed that E (neither car nor truck exert forces on each other) would also be selected by a comparable (albeit somewhat smaller) percentage of introductory students. This uncertainty between options D and E as the most common incorrect answer choice exhibited by graduate students in interviews was suggested by the quantitative in-class data obtained from the graduate students in the TA training class: 28% selected option D and 36% selected option E as the most common incorrect answer choices. However, only 6% of introductory students selected these answer choices cumulatively.

The alternate conception that the "active" object exerts a force on the "passive" object, but the passive object does not exert a force on the active object was very rare among introductory students (2% in a post-test both on question 16—answer choice D and 2% on question 28—answer choice B). However, on question 28, 36% of TAs believed that this alternate conception is the most common among introductory students (answer choice B). Thus, when it comes to Newton's third law, these TAs are likely to spend considerable time attempting to address alternate conceptions that very few introductory students have.

3. Kinematics in two dimensions: Identifying trajectories of objects under various kinds of forces

There are three questions which ask students to identify trajectories of various objects which move under certain kinds of forces on which TAs' performance at identifying student alternate conceptions was average: Questions 12,
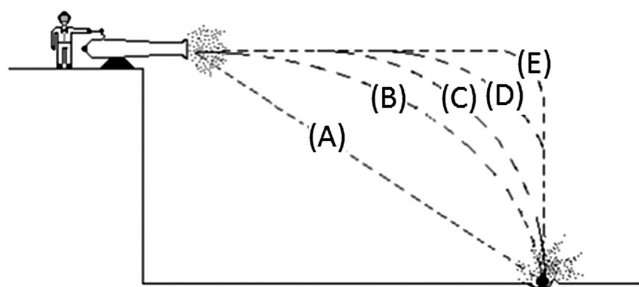
21, and 23 on which TAs' normalized FCI-related PCK scores were 67%, 52%, and 56%, respectively.

In question 12, a ball is shot from a cannon horizontally (see Fig. 1). The most common incorrect answer selected by introductory students is C (19%). Virtually no introductory students (only 1%) selected answer choice A in the pretest or in the post-test (see Table VIII), but 28% of TAs believed that this answer choice is the most common incorrect answer among introductory students as shown in Table I. We note that this path is (i) evidently contrary to the statement of problem (the path follows a straight line below the horizontal from the cannon, but the problem clearly states that the ball is shot horizontally) and (ii) contrary to everyday experience, which is what typically leads students to develop mental models that are inconsistent with physics principles.

Some interviewed graduate students selected answer choice A because they thought that introductory students would put together the vertical motion and the horizontal motion incorrectly into a straight line. It is important to note that once the graduate students latched on to this answer, they seemed to not consider the fact that introductory students have most likely seen projectile motion in real life, for example, a tennis player hitting a ball nearly horizontally, or a ball rolling off of a table. One interviewed graduate student stated

> "… they [introductory students] wouldn't tie in the acceleration to the curvature of the path so to me it seems like A [see Fig. 1] would be a more likely incorrect answer—so that path is just gonna follow a linear path to the ground."

In question 21, a rocket which is travelling horizontally (to the right), when reaching point b, starts its engine which produces a constant thrust (force) in the vertical direction until the rocket reaches point c (see Fig. 2 for the answer choices). The question asks students to identify the trajectory of the rocket between points b and c and the most common incorrect answer choice selected by 38% of the introductory students is C. Only 9% of introductory students selected answer choice D on the post-test (14% selected it on the pretest as shown in Table VIII), in which the rocket keeps traveling horizontally for a while after point b (see Fig. 2), but 36% of TAs selected this answer choice as the most common among introductory students as shown in Table I.

FIG. 1. Diagram for question 12 on the FCI, which provides the answer choices.
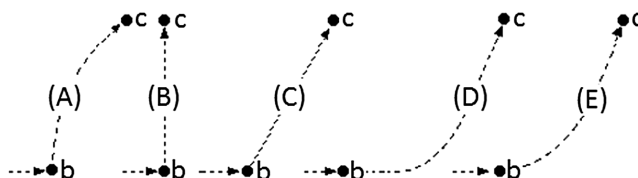

FIG. 2. Answer choices provided for question 21 on the FCI.

Some of the interviewed graduate students selected answer choice D because they expected that introductory students would think about the fact that trajectories in real life never change abruptly (i.e., snapping from one direction to another direction at an instant, such as the trajectories depicted in choices A, B, and C). Introductory students would therefore be tempted to select answer choice D, because the rocket continues to move horizontally for some time after the thrust is applied. For example, one interviewed graduate student said that answer choices A, B, and C would seem unnatural to students because in those trajectories, the rocket takes a sharp corner and selected answer choice D as the most common incorrect answer choice of introductory students. He said

> *"Well, it starts by something drifting sideways, and I think the closest analogous situation a student could successfully think about would be sliding sideways on ice. For instance if you are sliding sideways in a car and you hit the gas, you would not expect to immediately start rolling forward just because you hit the gas. So I would guess that a student would think that [answer choices] B and A would be very unlikely for that reason."*

He then reasoned that the trajectory depicted in answer choice D would look more "natural" to introductory students than the trajectory depicted in answer choice C, and concluded that answer choice D would be the most commonly selected incorrect answer choice of introductory students. However, as mentioned earlier, only 9% of introductory students selected answer choice D as the correct answer.

Question 23 also relates to the rocket and states that at point c the thrust immediately drops to zero. The question asks for the trajectory of the rocket after point c and provides the answer choices shown in Fig. 3 and the most common incorrect answer choice selected by 23% of introductory students is D. Choice E was selected by very few (7% in a pretest and 5% in a post-test, see Table VIII), but 24% of the TAs selected E as the most common incorrect answer choice of introductory students as shown in Table I.

The interviews indicated that it was very difficult for graduate students to predict the most common incorrect answer choice of introductory students on this question, partly because a student's answer to question 23 depends on his or her answer to question 21. Graduate students
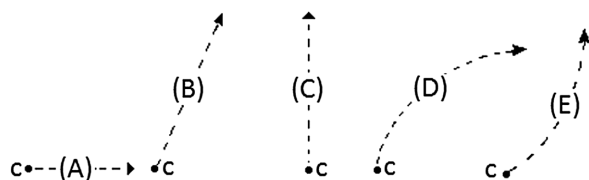
would sometimes say things like "if they [introductory] students think D in 21 (i.e., D is the most common incorrect answer choice), I don't see how they'd pick A here (question 23), but I'm not really sure about 21. Maybe they go for A (i.e., A is the most common incorrect choice), in which case A in 23 could make sense (as the most common incorrect choice)." Thus, they struggled to identify the most common incorrect answer choice of introductory students. One graduate student explicitly stated that he was guessing the most common incorrect answer of introductory students in this situation.

4. Identification of forces

In the context of an object at rest, many TAs thought that introductory students are unaware of any forces, including the force of gravity. Question 29 asks to identify all the forces acting on a chair at rest on a horizontal floor and nearly half of the TAs (44%) selected answer choice E (see Table I), which states that no forces act on the chair, while almost no introductory students selected this answer choice (1% in the post-test, 4% in the pretest as shown in Table VIII). This question has only one strong distractor choice (that the forces acting on the chair are the force of gravity, the normal force and a downward force exerted by the air) selected by 23% of introductory students but only 28% of the TAs were able to identify it correctly.

The think-aloud interviews indicated that graduate students are often distracted by the parenthetical statement included in answer choice E ("Since the chair is at rest there are no forces acting upon it") and they commented that this answer choice is a combination of student prior knowledge and what they had been taught in the course. They thought that students may incorrectly generalize that an object at rest has no forces acting on it (instead of the net force acting on it is zero) and conclude that there are no forces acting on the chair. For example, one graduate student said

> *"[Answer choice] E kind of incorporates some of what they learned […] 'since the chair is at rest, there are no forces acting upon it' is like 'since the chair is at rest there is no net force', um.. so that's what would make E exciting (for introductory physics students to select)."*

Another graduate student stated

> *"I think E is possible [as the most common incorrect answer choice of students] because it says that the chair is at rest and this reasoning here, 'oh since the chair is at rest there are no forces acting upon it'—that could convince students that the correct answer is E."*

5. Impetus view of motion

In question 13 which asks students to identify all of the forces that act on a ball after a boy throws a ball vertically upward, the most common incorrect answer choice selected

FIG. 3.　Answer choices provided for question 23 on the FCI.

by 50% of introductory students (and identified by only 16% of the TAs) is that on the way up the force of gravity and a force which decreases act on the ball and on the way down, it is only the force of gravity (answer choice C). On this question, one answer choice that was selected by very few introductory students (11%) is option B, which confuses velocity with force. It states that when the object is thrown, on the way up there is a steadily decreasing upward force and on the way down there is a steadily increasing downward force (gravity). This answer choice was selected by nearly half of the TAs (44%). In interviews, some of the graduate students claimed that it would be the most common because students may be confusing acceleration and velocity and using $F = ma$. For example, on the way up, students may think that the acceleration is decreasing (whereas it is the velocity that is decreasing) and use $F = ma$ to reason that this implies that the force is decreasing. Similarly, on the way down, they may conclude that the force is increasing. For example, one interviewed graduate student selected option B as the most common incorrect answer choice of introductory students and stated

"…I would say B here would be the most common incorrect. So decreasing upward force, so it's slowing down, so they would think there's a force pushing it up but slowing down, and on the way down a steadily increasing [force], so it's accelerating down, so I think they [students] would think that force of gravity as the object gets closer to the earth [would increase]."

RQ1 (ii) For what concepts or areas are TAs' responses context dependent (i.e., the majority of TAs identify the most common student alternate conceptions in some contexts, but not in others)?

Newton's 3rd law.—We found that the majority of TAs identified the alternate conception in the typical context (small car colliding with truck), but few TAs identified the alternate conception in less typical contexts.

Table III shows that the percentage of introductory students who hold the alternate conception related to Newton's third law did not vary significantly across different questions (varied from 32% to 48%), whereas the performance of TAs in identifying this alternate conception of introductory students varied significantly (from 40% to 84%). While the vast majority of TAs identified the alternate conception in a typical context (question 4—truck colliding with car—see Table III), they identified it much less often in the other two contexts (question 15—car pushing truck and speeding up, and question 28—student "a" pushing student "b"). We note that these are questions in which there is only one very strong distractor choice. However, the TAs were not able to identify that distractor choice equally well across different questions.

Newton's 2nd law.—While the TAs had some difficulty identifying introductory students' alternate conception that a constant nonzero net force implies a constant velocity in some contexts (as discussed earlier), they were aware of this alternate conception in most contexts. For example, in question 17 an elevator is being pulled up by a cable at constant velocity and 72% of introductory students claimed that the force exerted by the cable is larger than the forces exerted in the opposite direction (force of gravity—answer choice A, and force of gravity and the downward force exerted by the air—answer choice D). 88% of the TAs were aware that introductory students have this alternate conception in this context as shown in Table I. In question 24, a rocket is moving under the influence of a constant force which suddenly goes to zero and 22% of introductory students who selected answer choice C thought that the rocket will start to slow down (i.e., a constant force is needed to keep it moving at a constant velocity). This alternate conception was identified by 68% of the TAs. In question 26, a woman is exerting a constant horizontal force on a box which results in the box moving at a constant velocity. She then doubles the force she exerts on the

TABLE III. *Contrasting examples (related to Newton's 3rd law) of TAs' ability to identify introductory students' alternate conceptions.* Questions on Newton's 3rd law, percentage of introductory students who answered the questions incorrectly in the pretest and in the post-test, percentage of introductory students who hold the alternate conception in the pretest (Intro student alt. pre) and in the post-test (Intro student alt. post.), and percentage of TAs who identified the alternate conception. For convenience, brief descriptions of the FCI questions are provided.

Introductory student alternate conception	FCI item No.	% Overall incorrect intro pre	% Overall incorrect intro post	Intro student alt. pre.	Intro student alt. post	TA %
Newton's 3rd: while both objects exert forces on one another, if both	4	74	40	73%	39%	84
objects are active (i.e., collision), the larger object exerts the larger	15	75	56	61%	48%	40
force; if only one is active (i.e., car pushing truck), the active object	28	76	41	61%	32%	52
exerts a larger force on passive object than vice versa						

Questions
4. Truck colliding with car.
15. Car pushing truck and speeding up.
28. Student "a" puts his feet on student "b" and pushes against student "b."

box and the two most common alternate conceptions of introductory students are that the speed of the box also doubles (answer choice A selected by 41% of introductory students) and that the speed of the box increases, but is not necessarily double the original speed (answer choice B selected by 32% of introductory students). These alternate conceptions were identified by 52% and 32% of the TAs, respectively, as shown in Table I. In question 27, the woman suddenly stops applying the force and roughly one-quarter of the introductory students thought that the box will immediately come to a stop—an alternate conception identified by 68% of the TAs (see Table I). Thus, the TAs' ability in identifying this alternate conception was context dependent. We note, however, that it was only a few contexts in which they struggled to identify it, and in most situations they performed reasonably well at identifying this alternate conception.

RQ1 (iii) For what concept/area do TAs systematically underestimate the most prevalent alternate conceptions of introductory students?

For alternate conceptions that were uncovered in more than one question, TAs' performance in identifying them either was context dependent or was reasonably good across contexts. There was one alternate conception related to the impetus view of motion which the TAs struggled to identify. However, this alternate conception was only tested in one question on the FCI (question 13). The alternate conception is that after a boy throws a ball in the air vertically, when the ball is going up, the forces that act on the ball are the force of gravity and a steadily decreasing upward force. On the way down, only the force of gravity acts on the ball. This alternate conception was quite prominent among introductory students both in the post-test and in the pretest (50% and 64%, respectively), but it was identified by very few TAs (16%) as shown in Table IV. It is therefore unlikely that TAs would address this alternate conception directly in their instructional design.

One may wonder where the discrepancy between introductory students' and TAs' choices is coming from. It is possible that it is due to TAs not being aware of student alternate conceptions, but it is also possible that it is due to an issue with the test itself (e.g., there may be a difference between experts' and novices' perception of what the questions are asking). We note that the think-aloud interviews indicated that the TAs were interpreting the questions in the manner in which they were intended, and the research which resulted in the development of the FCI ensured that the novices were also interpreting the questions and answer choices as intended. This suggests that the discrepancy mentioned above is most likely due to TAs' lack of awareness of introductory students' alternate conceptions.

RQ1 (iv) What alternate conceptions can TAs identify across contexts?

1. Identification of distinct forces acting on an object

Regarding questions asking to identify all the distinct forces that act on an object, the vast majority of TAs were aware of the common alternate conceptions that objects moving in a particular direction must be acted upon by a distinct force acting in that direction (e.g., an upward "force of the hand" after an object is thrown vertically upwards). On the four questions that ask students to identify all of the forces that act on an object, namely, questions 5 (ball sliding in a horizontal, circular, frictionless channel), 11 (puck sliding on horizontal ice), 13 (boy throwing a ball in the air), and 18 (boy swinging on a rope) 80%, 63%, 65%, and 84% of introductory students, respectively, held this alternate conception after instruction. On these questions, 100% (question 5), 80% (question 11), 86% (question 13), and 90% (question 18) of the TAs identified this alternate conception.

2. Alternate conceptions related to interpreting strobe diagrams of motion

There are two questions (19 and 20) that ask students to interpret strobe diagrams of motion and TAs' performance in identifying student alternate conceptions in these questions was reasonably good. Question 19 provides positions

TABLE IV. *An example (related to impetus view of motion) in which TAs severely underestimate the percentage of introductory students who select a particular incorrect answer choice.* Introductory students' alternate conception related to an impetus view of motion, percentage of introductory students who answer the question incorrectly in the pretest and in the post-test, percentage of introductory students who hold this alternate conception in the pretest (Intro student alt. pre) and in the post-test (Intro student alt. post) and percentage of TAs (TA%) who identify it as the most common incorrect answer choice. A brief description of the question is also provided.

Introductory student alternate conceptions or difficulties	FCI item No.	% Overall incorrect intro pre	% Overall incorrect intro post	Intro student alt. pre	Intro student alt. post	TA %
Ball thrown vertically in the air: on the way up—steadily decreasing upward force and gravity, on way down, only gravity	13	88	65	64%	50%	16

Question
13. Ball thrown vertically in the air, no air resistance. Find the forces acting on the ball while in the air.
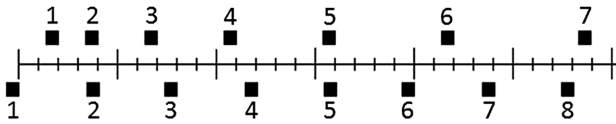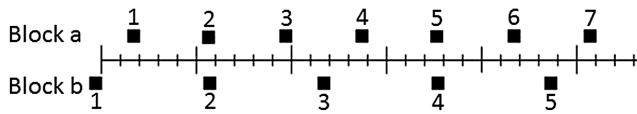
FIG. 4.   Diagram for question 19.

FIG. 5.   Diagram for question 20.

for two blocks at successive 0.2-sec intervals as shown in Fig. 4. The question asks if the blocks ever have the same speed and many introductory students (46%) confused position with velocity and said that the blocks have the same speed at position 2, 5, or both (answer choices A, C, and D). 88% of the TAs identified this alternate conception.

Question 20 also provides positions for two blocks, a and b, at successive 0.2-sec intervals as shown in Fig. 5 and asks students to compare the accelerations of the two blocks. 27% of introductory students who confused velocity with acceleration said that block b has a larger acceleration (answer choice C) and 56% of the TAs identified this as the most common alternate conception in this question.

3. Other alternate conceptions identified by the majority of TAs

There were two more alternate conceptions that the majority of the TAs identified, but these only occurred in one context each. In question 2, two balls roll off of a horizontal table and nearly half of the introductory students (46%) believed that the heavier ball will hit the floor either considerably closer than the lighter ball or halfway between where the lighter ball hits the floor and the base of the table (answer choices B and D). This alternate conception was identified by nearly all the TAs (96%).

In question 9, a puck is moving at some speed v_0 towards the right and is kicked vertically. Had the puck been at rest when kicked, the kick would impart a speed v_k to the puck. The question asks to find the speed of the puck after the kick. Here, there were two alternate conceptions: the speed of the puck is equal to the arithmetic sum of v_0 and v_k held by 19% of introductory students (answer choice C) and the speed of the puck after the kick is equal to v_k and independent of v_0 (i.e., the initial speed is irrelevant) held by 20% of introductory students (answer choice B). 88% of the TAs identified these alternate conceptions (76% identified the former and 12% identified the latter).

RQ2: How does the performance of physics TAs on the FCI-related PCK task compare with that of physics instructors?

TABLE V.   Numbers of instructors, TAs, and random guessers, averages (and percentage of maximum possible score), and standard deviations (Std. dev.) for the FCI-related PCK scores obtained (in determining introductory students' alternate conceptions on the FCI) out of a maximum score of 9.21.

	N	Average (% of max possible score)	Std. dev.
Instructors	30	6.25 (68)	0.90
TAs	25	6.01 (65)	0.78
Random guessing	24	3.71 (40)	0.93

As noted earlier, the maximum possible FCI-related PCK score of TAs or physics instructors on each question is equal to the maximum fraction of introductory students who selected an incorrect answer choice. The maximum possible FCI-related PCK score on the whole FCI survey is 9.21, which is the sum of these fractions for all the questions. Table V shows the average FCI-related PCK score of the TAs, instructors, and random guessers. ANOVA [75] revealed that not all three groups were comparable ($p < 0.001$), and we performed pairwise t tests to investigate the differences further. We found that both TAs and instructors performed better than random guessing (both p values for comparing TAs' and instructors' performance to random guessing are less than 0.001), but there was no significant difference between TAs and instructors in terms of their FCI-related PCK scores (p value 0.295).

RQ3: Are physics TAs from the United States, who have been exposed to undergraduate teaching in the U.S., better at identifying introductory students' alternate conceptions than international physics TAs?

Our analysis suggests that it was not the case that TAs from the U.S., on average, performed better than the others. In particular, the averages of these three groups of TAs (US, Chinese, and other international, who obtained 67%, 66%, and 62% of the maximum possible PCK score, respectively) were very similar as shown in Table VI and a nonparametric test, the Kruskal-Wallis test [75], indicated no significant differences between the groups ($p = 0.429$). In addition, the class discussion after the TAs completed all tasks related to the FCI suggested that

TABLE VI.   Numbers of TAs of different backgrounds, their averages (and percentage of maximum possible score), and standard deviations (Std. dev.) for the scores obtained in determining introductory students' alternate conceptions on the FCI out of a maximum of 9.21.

	N	Average (% of max possible score)	Std. dev.
US	9	6.20 (67)	0.70
Chinese	9	6.04 (66)	0.76
Other international	7	5.71 (62)	0.91

TABLE VII. FCI-related PCK performance of TAs in the individual and in the group PCK tasks: number of TAs or groups (*N*), averages (and percentage of maximum possible PCK score), and standard deviations (Std. dev.)

TAs' FCI-related PCK performance			
	N	Average	Std. dev.
Individual	25	6.01 (65%)	0.78
Group	12	6.59 (72%)	0.79

there were no differences between the TAs in this regard in terms of how they reasoned about student difficulties.

RQ4: To what extent do TAs identify introductory students' difficulties more often when working in groups than when working individually (i.e., do discussions improve TAs' understanding of introductory students' alternate conceptions related to force and motion revealed by the FCI)?

Table VII shows TAs' FCI-related PCK performance when they worked individually and in groups of two or three. ANOVA shows that the group performance was better than the individual performance ($p = 0.040$). In addition, the effect size as measured by Cohen's d (which is 0.74) shows a reasonably large effect of group discussions on TAs' FCI-related PCK performance (generally, effect sizes of 0.8 and higher are considered to correspond to large effects [75]).

Furthermore, there were 98 instances in which two or three TAs who did not all select the same incorrect answer choice in the individual PCK task, when working in groups, converged to one of their original answers pertaining to introductory students' common difficulties. In 73 of those instances (74%) the TAs converged to the "better" option (i.e., the more common incorrect answer choice of introductory students by 5% or more). It therefore appears that discussions among TAs tended to lead them to agree on a more common introductory student alternate conception.

IV. SUMMARY

Awareness of introductory physics students' common difficulties and being able to understand how students reason about physics are important aspects of pedagogical content knowledge because one can take advantage of introductory students' initial knowledge and design pedagogical approaches to account for these difficulties and help students learn better [76–78]. Our investigation used the FCI to evaluate this aspect of the pedagogical content knowledge for introductory mechanics of 25 TAs, who were all first-year physics graduate students enrolled in a TA training course. We also compared their FCI-related PCK performance with that of physics instructors with varying degrees of teaching experience. For each item on

the FCI, the TAs and instructors were asked to identify the most common incorrect answer choice of introductory physics students. We also had an in class discussion with the TAs about these issues after they completed this task individually and in small groups.

A. The majority of TAs identified the alternate conception related to Newton's third law in a typical context but fewer could identify it in less typical contexts

Introductory students' alternate conception that larger and/or active objects exert larger forces on smaller and/or passive objects was identified by nearly all of the TAs in the typical context (small car colliding with truck). However, in other contexts, fewer TAs (roughly half or less) were able to identify the alternate conception. In general, in order to help students learn, TAs should know how students' alternate conceptions instantiate in different contexts and what difficulties they lead to in order to make use of effective instructional approaches to help students transition from their naïve notions about force and motion to the Newtonian view. It appears that, at least in the context of Newton's third law, a sizable portion of the TAs (half or more) were unable to recognize the types of reasoning students would use in different contexts.

B. TAs were not able to identify introductory students' alternate conception related to an impetus view of motion

When someone throws an object in the air, many introductory students believe that there is a "force of the hand" that gradually decreases until the object reaches its highest point, and on the way down, only the force due to gravity acts on the object. Few TAs (16%) were able to identify this alternate conception although 64% and 50% of the introductory students before and after instruction displayed it.

C. Alternate conceptions held by very few students which TAs expected would be the most common

There were multiple instances in which TAs selected certain incorrect answer choices which they believed would be most common among introductory students, but those answer choices were very rarely selected by introductory students. For example, with regards to Newton's third law, many TAs, sometimes over 60% in certain contexts, expected that introductory students are unaware that both interacting objects exert forces on one another, but very few introductory students fit this category. For Newton's 2nd law, in the context of a woman pulling a box at constant speed across a horizontal floor, nearly half the TAs expected that students would compare the force exerted by the woman with the weight of the box, instead of comparing it with the total force that resists the motion of

the box, but very few introductory students did this. Other examples include selecting the trajectory of objects under various forces, for example, in the context of a ball under the influence of gravity shot horizontally—nearly all students were aware that the ball cannot take the trajectory that is not initially horizontal, but many TAs expected introductory students to select this trajectory. With regards to identification of forces, the TAs sometimes selected answer choices which state that no forces act on an object (e.g., when a chair is at rest on a horizontal floor) as the most common among introductory students, when virtually no introductory students (1%) selected these answer choices both before and after instruction.

D. Common reasoning TAs use when they incorrectly argue that certain answer choices must be popular among introductory students

Think-aloud interviews indicated that TAs can be distracted by certain details of particular answer choices. For example, with questions regarding Newton's third law, in the less typical contexts, interviewed graduate students incorrectly believed that the parenthetical statement that the car exerts a force on the truck because its engine is running, which provides an explanation for why forces between the two objects arise, would look convincing to students. They therefore concluded that this answer choice (car exerts a force on truck, but truck does not exert a force on the car) would be the most popular incorrect answer choice among introductory students for both questions 15 and 16. However, introductory students selected this answer choice very rarely. Similarly, regarding identification of forces, for question 29, interviewed graduate students sometimes believed that the parenthetical statement which stated that since the chair is at rest there are no forces acting on it would seem convincing to students. However, the quantitative data indicated that this is not the case.

Sometimes, graduate students expected that certain reasoning patterns would be common among introductory students when in fact they were not. For example, in question 12, some interviewed graduate students thought that students would not realize that the motion in the *vertical* direction is accelerated and select the straight trajectory. However, this type of reasoning is very rare among introductory students, partly because the context of the question (cannon shooting a ball) is familiar to students from everyday experience. Other questions show similar results.

E. Alternate conceptions that the TAs were able to identify

The TAs were aware that certain alternate conceptions are common among students. For example, TAs performed reasonably well in identifying the alternate conception that a constant nonzero net force acting on an object imparts a constant velocity to the object and in certain contexts (e.g.,

cable pulling up an elevator) the majority of them identified it (although in some contexts they did not). The TAs were also aware that introductory students believe that moving objects are acted upon by a distinct force in the direction of motion, that a heavier ball rolling off a table will strike the ground closer to the edge of the table than a lighter ball, and that when interpreting strobe diagrams of motion students confuse position with velocity. About half of the TAs were aware that when interpreting strobe diagrams of motion, students also confuse velocity with acceleration.

1. Other findings

We found that the ability to identify common introductory students' alternate conceptions on the FCI was not dependent on familiarity with U.S. teaching practices and that TAs and instructors exhibited comparable performance in identifying introductory students' alternate conceptions on the FCI. Finally, we found evidence to support that TA discussions improved their PCK performance in identifying common introductory students' alternate conceptions.

V. IMPLICATIONS FOR PROFESSIONAL DEVELOPMENT

Professional development leaders can benefit from knowing the prior knowledge of physics TAs with regards to common alternate conceptions of introductory students (i.e., which common alternate conceptions are TAs aware of, which ones are they not aware of, and which ones do they think are common but instead are not) when designing a professional development program. Additionally, tasks similar to the FCI-related PCK task described in this paper can be used to motivate a discussion about the importance of knowing student difficulties in order to help students learn. This can be followed by a discussion about how one should make use of these common difficulties as resources to design instruction to help students learn better.

Some TAs (and instructors) explicitly noted that this FCI-related PCK task was challenging and it was difficult for them to think about physics questions from an introductory physics student's perspective. In the think-aloud interviews, graduate students sometimes made comments which indicated that they found the task challenging (e.g., explicitly commenting "I don't know introductory students well enough…"). However, many TAs noted that the FCI-related PCK task was worthwhile and helped them think about the importance of putting themselves in their students' shoes in order for teaching and learning to be effective, especially after receiving introductory student data on how students actually performed, and discussing particular student alternate conceptions. Furthermore, we found that TA discussions of introductory students' difficulties led to a better understanding of students' initial knowledge state (and difficulties) and therefore, exercises

which encourage such discussions in the context of conceptual assessments such as the FCI could be beneficial and should be incorporated into professional development courses and workshops. Since similar studies on PCK related to other physics concepts would be valuable, we are currently carrying out such investigations focusing on other topics in mechanics and electricity and magnetism [79–82].

APPENDIX

1. Mathematical description of how the FCI-related PCK scores were calculated

We define indices i, j, and k that correspond to the following:
- i: index of TAs (25 TAs; it takes values from 1 to 25);
- j: FCI question number (30 questions; it takes values from 1 to 30);
- k: incorrect answer choice number for each question (4 incorrect answer choices; it takes values from 1 to 4).

Then, we let F_{jk} be the fraction of introductory physics students who selected incorrect answer choice k on item j (e.g., $F_{21} = 0.44$, $F_{22} = 0.06$, $F_{23} = 0.21$, $F_{24} = 0.04$). We let TA_{ijk} correspond to whether TA i chose incorrect answer choice k on item j (for a given i and j, $TA_{ijk} = 1$ only for the incorrect answer choice k, selected by TA i on item j, otherwise $TA_{ijk} = 0$). Then, the PCK score of the ith TA on item j (referred to TA_{ij}) is $TA_{ij} = \sum_{k=1}^{4}(TA_{ijk} \cdot F_{jk})$. Then, the total PCK score of the ith TA (TA_i) on the whole survey can be obtained by summing over all of the questions:

$$TA_i = \sum_{j=1}^{30} TA_{ij} = \sum_{j=1}^{30}\left[\sum_{k=1}^{4}(TA_{ijk} \cdot F_{jk})\right].$$

Also, the average PCK score of all of the TAs on item j

(referred to as $\overline{TA}_j$) can be obtained by taking an average over the TA scores on that particular question:

$$\overline{TA}_j = \frac{1}{25}\left(\sum_{i=1}^{25}TA_{ij}\right) = \frac{1}{25}\left(\sum_{i=1}^{25}\left[\sum_{k=1}^{4}(TA_{ijk} \cdot F_{jk})\right]\right).$$

A similar approach can also be adopted for the instructors ($I_{ij} = $ PCK score of the ith instructor on item j; $I_i = $ PCK score of the ith instructor on the whole survey; $\overline{I}_j = $ average PCK score of all instructors on item j) and for random guessers ($RG_{ij} = $ PCK score of ith random guesser on item j; $RG_i = $ PCK score of ith random guesser; $\overline{RG}_j = $ average PCK score of random guessers on item j). The PCK scores of each (i) TA-instructor-random guesser (GS_i, I_i, RG_i as described above) were used to obtain averages and standard deviations in order to perform t tests to compare the FCI-related PCK performance of TAs with that of the instructors and random guessers on the whole survey (and to compare different subgroups of TAs and instructors). In order to compare the PCK performance of these different groups on individual items, the averages and standard deviations of the PCK scores on that particular question (e.g., for question j on the FCI: TA_{ij}, I_{ij}, RG_{ij}) were used to perform t tests.

2. Comparison of TA performance with random guessing

The minimum FCI-related PCK score on this task is not zero. As long as a TA knows the correct answer for each question, he or she can randomly select one of the four incorrect answer choices. We therefore generated a population of 24 "random guessers" (24 was chosen in order to have a reasonable group size when performing ANOVA). Random guessing on this task would correspond to choosing one of the four incorrect answer choices for each question with equal probability (25%). Therefore, one-quarter of the random guessers always selected the first incorrect answer choice, one-quarter selected the second incorrect answer choice, etc. Therefore, each individual random guesser obtains a score for each question and these scores were used to perform comparison with the TA scores via ANOVA. We note that our choice of random guessers maximizes the standard deviation.

3. FCI pre- and post-test performance of the algebra-based and calculus-based courses

See details in Tables VIII and IX.

TABLE VIII. Percentages of algebra-based introductory physics students who selected each answer choice for each item on the FCI when it was given as a pretest and as a post-test and normalized gain (Norm. gain) on each item on the FCI. The percentages on the pretest are based on data from 601 students taught by two different instructors in two different semesters and the percentages on the post-test are based on data from 899 students taught by 4 different instructors over several years. The green shaded boxes indicate correct answers, and red font indicates incorrect answer choices selected by 20% or more students. All the courses were taught in a traditional manner which did not incorporate PER based teaching strategies.

FCI item No.	Pre-test algebra					Post-test algebra					Norm. gain
	A	B	C	D	E	A	B	C	D	E	
1	13	6	53	25	4	10	4	78	8	1	0.53
2	27	21	7	37	8	44	25	6	21	4	0.23
3	31	16	40	3	10	15	13	63	5	5	0.38
4	73	0	0	1	26	39	1	0	0	60	0.46
5	4	10	31	25	29	3	24	12	44	17	0.16
6	25	68	5	2	0	16	79	3	1	1	0.33
7	17	57	9	5	11	12	74	6	3	4	0.40
8	20	47	1	13	18	14	66	0	8	11	0.36
9	4	26	20	6	43	4	20	19	5	53	0.17
10	54	1	11	19	15	71	2	7	13	7	0.36
11	7	31	45	14	3	5	8	48	35	4	0.24
12	1	59	32	5	2	1	77	19	2	1	0.43
13	4	21	64	12	0	4	11	50	35	0	0.27
14	35	18	11	36	0	19	10	9	61	0	0.39
15	25	10	61	3	0	44	7	48	1	0	0.25
16	55	3	37	4	1	73	2	19	2	4	0.40
17	60	8	1	22	9	62	24	1	10	3	0.17
18	2	12	14	27	46	1	28	4	42	25	0.19
19	14	3	3	26	54	12	3	5	29	51	-0.06
20	19	6	36	32	8	16	4	27	49	4	0.25
21	7	21	23	14	35	7	13	38	9	33	-0.02
22	37	30	4	26	2	33	45	3	16	2	0.22
23	16	29	21	28	7	15	39	18	23	5	0.14
24	63	1	25	3	6	70	2	22	2	5	0.20
25	3	8	12	58	19	3	9	23	53	12	0.13
26	41	42	4	11	3	41	32	3	9	14	0.11
27	31	13	54	1	0	26	13	58	2	0	0.09
28	0	6	8	61	24	1	2	6	32	59	0.46
29	15	42	1	37	4	2	71	3	23	1	0.50
30	1	7	12	1	79	3	10	26	2	59	0.16
						Avg. normalized gain					0.26

TABLE IX   Percentages of calculus-based introductory physics students who selected each answer choice for each item on the FCI when it was given as a pretest and as a post-test. The percentages on the pretest are based on data from 364 students taught by three different instructors over several semesters and the percentages on the post-test are based on data from 296 students taught by two different instructors during two different semesters. The green shaded boxes indicate correct answers and red font indicates incorrect answer choices selected by 20% or more students. All the courses were taught in a traditional manner which did not incorporate PER based teaching strategies.

FCI item No.	Pre-test calculus					Post-test calculus					Norm. gain
	A	B	C	D	E	A	B	C	D	E	
1	8	3	82	6	1	7	1	86	6	1	0.24
2	39	24	4	26	7	55	27	4	15	1	0.25
3	17	8	66	4	5	15	4	76	3	2	0.30
4	59	1	0	0	39	23	0	1	0	76	0.60
5	5	23	20	27	25	3	47	13	29	9	0.31
6	17	79	3	1	1	12	86	2	1	0	0.32
7	9	76	4	3	7	7	85	3	1	5	0.39
8	11	62	0	12	17	8	74	0	8	10	0.33
9	2	20	20	6	53	1	9	14	3	73	0.44
10	70	3	6	12	8	91	1	3	3	3	0.69
11	8	16	42	30	4	6	2	28	61	3	0.44
12	0	84	14	2	0	0	90	10	0	0	0.41
13	4	12	51	33	0	3	5	31	61	0	0.42
14	21	12	5	63	0	14	9	5	72	0	0.25
15	29	6	63	2	0	42	5	52	0	1	0.19
16	72	1	23	2	3	86	2	7	1	4	0.50
17	63	21	1	12	4	46	44	2	6	2	0.29
18	3	27	9	34	29	0	55	2	33	11	0.38
19	11	3	3	20	62	8	2	3	12	75	0.35
20	11	4	22	54	6	9	2	16	71	3	0.36
21	6	11	26	11	46	4	6	23	11	57	0.20
22	27	51	2	19	1	20	67	2	11	1	0.34
23	9	48	16	24	3	5	64	13	14	3	0.32
24	74	1	18	2	4	87	1	9	1	3	0.48
25	2	9	29	51	10	3	6	51	38	2	0.31
26	35	29	1	14	20	20	28	3	7	42	0.28
27	22	7	66	4	1	14	6	75	3	1	0.28
28	2	4	6	45	44	1	3	3	20	72	0.51
29	6	61	2	27	3	1	83	1	13	2	0.56
30	3	7	30	1	59	3	7	44	1	45	0.20
						Avg. normalized gain					0.36

[1] C. Singh, Categorization of problems to assess and improve proficiency as teacher and learner, Am. J. Phys. **77,** 73 (2009); A. J. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, Phys. Rev. ST Phys. Educ. Res. **7,** 020110 (2011).

[2] C. Singh, Rethinking Tools for Training Teaching Assistants, *Proceedings of the 2009 Physics Education Research Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh, AIP Conf. Proc. No. 1179 (AIP, Melville, New York, 2009), p. 59.

[3] S. Y. Lin, C. Henderson, W. Mamudi, E. Yerushalmi, and C. Singh, Teaching assistants' beliefs regarding example solutions in introductory physics, Phys. Rev. ST Phys. Educ. Res. **9,** 010120 (2013).

[4] E. Yerushalmi, C. Henderson, W. Mamudi, C. Singh, and S. Y. Lin, The group administered interactive questionnaire: An alternative to individual interviews, *Proceedings of the 2011 Physics Education Research Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt, AIP Conf. Proc. No. 1413 (AIP, Melville, New York, 2012), p. 97.

[5] S. Y. Lin, C. Singh, W. Mamudi, C. Henderson, and E. Yerushalmi, TA-designed vs. research-oriented problem solutions, *Proceedings of the 2011 Physics Education Research Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt, AIP Conf. Proc. No. 1413 (AIP, Melville, New York, 2012), p. 255.

[6] E. Yerushalmi, E. Marshman, A. Maries, C. Henderson, and C. Singh, Grading practices and considerations of graduate students at the beginning of their teaching assignment, *Proceedings of the 2014 Physics Education Research Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (2015), p. 287 [arXiv:1601.02562].

[7] C. Henderson, E. Marshman, A. Maries, E. Yerushalmi, and C. Singh, Instructional goals and grading practices of graduate students after one semester of teaching experience, *Proceedings of the 2014 Physics Education Research Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (2015), p. 111 [arXiv:1601.02560].

[8] V. Otero, S. Pollock, and N. Finkelstein, A physics department's role in preparing teachers: The Colorado learning assistant model, Am. J. Phys. **78,** 1218 (2010).

[9] F. Lawrenz, P. Heller, and R. Keith, Training the teaching assistant: Matching TA strengths and capabilities to meet specific program goals, J. Coll. Sci. Teach. **22,** 106 (1992).

[10] *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices*, edited by C. Sandifer and E. Brewe (American Physical Society, PhysTEC, New York, 2015).

[11] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. 1. Group vs individual problem solving, Am. J. Phys. **60,** 627 (1992).

[12] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. 2. Designing problems and structuring groups, Am. J. Phys. **60,** 637 (1992).

[13] C. Singh, Impact of peer interaction on conceptual test performance, Am. J. Phys. **73,** 446 (2005).

[14] C. Singh, Effectiveness of group interaction on conceptual standardized test performance, *Proceedings of the 2002 Physics Education Research Conference, Boise, ID*, edited by S. Franklin, K. Cummings, and J. Marx (2002), p. 67 [arXiv:physics/0207106].

[15] A. J. Mason and C. Singh, Helping students learn effective problem solving strategies by reflecting with peers, Am. J. Phys. **78,** 748 (2010); Impact of guided reflection with peers on the development of effective problem solving strategies and physics learning, Phys. Teach. **54,** 295 (2016).

[16] A. J. Mason and C. Singh, Using reflection with peers to help students learn effective problem solving strategies, *Proceedings of the 2010 Physics Education Research Conference, Portland, OR*, edited by C. Singh, M. Sabella, and S. Rebello, AIP Conf. Proc. No. 1289 (AIP, Melville, New York, 2010), p. 41.

[17] P. M. Sadler, G. Sonnert, H. P. Coyle, N. Cook-Smith, and J. L. Miller, The influence of teachers' knowledge on student learning in middle school physical science classrooms, Am. Educ. Res. J. **50,** 1020 (2013).

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

[19] L. S. Shulman, Knowledge and teaching: Foundations of the new reform, Harv. Educ. Rev. **57,** 1 (1987).

[20] D. Hestenes, M. Wells, and G. Swackhammer, Force Concept Inventory, Phys. Teach. **30,** 141 (1992).

[21] I. Halloun, R. R. Hake, E. P. Mosca, and D. Hestenes, Force Concept Inventory (Revised, 1995); online (password protected) at http://modeling.la.asu.edu/R&E/Research.html and also printed in Mazur (1997).

[22] I. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53,** 1043 (1985).

[23] F. Reif, Educational challenges for the university, Science **184,** 537 (1974).

[24] J. Clement, Students' preconceptions in introductory mechanics, Am. J. Phys. **50,** 66 (1982).

[25] A. Arons, Student patterns of thinking and reasoning, Phys. Teach. **21,** 576 (1983).

[26] L. C. McDermott, Research on conceptual understanding in mechanics, Phys. Today **37,** 24 (1984).

[27] D. Hestenes and M. Wells, A mechanics baseline test, Phys. Teach. **30,** 159 (1992).

[28] R. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. **62,** 750 (1994).

[29] R. Thornton and D. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation, Am. J. Phys. **66,** 338 (1998).

[30] J. Mitchell, J. Martin, and T. Newell, Development of a concept inventory for fluid mechanics, *Proceedings, Frontiers in Education Conference, 2003* (IEEE, Boulder, CO, 2003), p. T3D-23.

[31] G. L. Gray, D. Evans, P. J. Cornwell, B. Self, and F. Constanzo, The Dynamics Concept Inventory Assessment Test: A Progress Report, *Proceedings of the 2005 American Society for Engineering Education Annual Conference* (American Society for Engineering Education, Portland, OR, 2005).

[32] D. Huffman and P. Heller, What does the force concept inventory actually measure?, Phys. Teach. **33,** 138 (1995).

[33] D. Hestenes and I. Halloun, Interpreting the Force concept inventory: A response to March 1995 critique by Huffman and Heller, Phys. Teach. **33,** 502 (1995).

[34] P. Heller and D. Huffman, Interpreting the force concept inventory: A reply to Hestenes and Halloun, Phys. Teach. **33,** 503 (1995).

[35] I. Halloun and D. Hestenes, The Search for Conceptual Coherence in FCI Data. Modeling Instruction Workshop website at Arizona State University [http://modeling.asu.edu/R&E/CoherFCI.pdf (14 December 2001)].

[36] G. A. Morris, L. Branum-Martin, N. Harshman, S. D. Baker, E. Mazur, T. Mzoughi, and V. McCauley, Testing the test: Item response curves and test quality, Am. J. Phys. **74,** 449 (2006).

[37] G. A. Morris, N. Harshman, L. Branum-Martin, E. Mazur, T. Mzoughi, and S. D. Baker, An item response curve analysis of the force concept inventory, Am. J. Phys. **80,** 825 (2012).

[38] J. Wang and L. Bao, Analyzing force concept inventory with item response theory, Am. J. Phys. **78,** 1064 (2010).

[39] M. Planinic, L. Ivanjek, and A. Susac, Rasch model based analysis of the force concept inventory, Phys. Rev. ST Phys. Educ. Res. **6,** 010103 (2010).

[40] R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66,** 64 (1998).

[41] N. Lasry, E. Mazur, and J. Watkins, Peer Instruction: From Harvard to the two-year college, Am. J. Phys. **76,** 1066 (2008).

[42] A. P. Fagen, C. H. Crouch, and E. Mazur, Peer Instruction: Results from a range of classrooms, Phys. Teach. **40,** 206 (2002).

[43] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Engelwood Cliffs, NJ, 1997).

[44] M. Wells, D. Hestenes, and G. Swackhamer, A modeling method for high school physics instruction, Am. J. Phys. **63,** 606 (1995).

[45] C. Singh, Hands-on homework for introductory science, J. Coll. Sci. Teach. **32,** 126 (2002); Exploration center for large introductory physics courses, Phys. Teach. **38,** 189 (2000).

[46] R. R. Hake, Interactive engagement methods in introductory mechanics courses, preprint, 5/1997, available on request to the author or at http://carini.physics.indiana.edu/SDI/.

[47] V. P. Coletta and J. A. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, Am. J. Phys. **73,** 1172 (2005).

[48] K. L. Malone, Correlations among knowledge structures, force concept inventory, and problem solving behaviors, Phys. Rev. ST Phys. Educ. Res. **4,** 020107 (2008).

[49] P. M. Pamela and J. M. Saul, Interpreting FCI normalized gain, pre-instruction scores, and scientific reasoning ability, AAPT Announcer **36,** 89 (2006).

[50] D. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable', Am. J. Phys. **70,** 1259 (2002).

[51] V. P. Coletta, J. A. Phillips, and J. A. Steinert, Interpreting force concept inventory scores: Normalized gain and SAT scores, Phys. Rev. ST Phys. Educ. Res. **3,** 010106 (2007).

[52] P. Nieminen, A. Savinainen, and J Viiri, Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning, Phys. Rev. ST Phys. Educ. Res. **8,** 010123 (2012).

[53] H. Ginsberg and S. Opper, *Piaget's Theory of Intellectual Development* (Prentice Hall, Englewood Cliffs, NJ, 1969).

[54] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accomodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66,** 211 (1982).

[55] A. S. Elstein, L. S. Shulman, and S. Sprafka, *Medical Problem Solving: The Analysis of Clinical Reasoning* (Harvard University Press, Cambridge, England, 1978).

[56] J. H. van Driel, N. Verloop, and W. de Vos, Developing science teachers' pedagogical content knowledge, J. Res. Sci. Teach. **35,** 673 (1998).

[57] P. L. Grossman, *The Making of a Teacher: Teacher Knowledge and Teacher Education* (Teachers College Press, New York 1990).

[58] P. L. Grossman, What are we talking about anyhow: Subject matter knowledge for secondary English teachers, in *Advances in Research on Teaching, Vol. 2: Subject Matter Knowledge*, edited by J. Brophy (JAI Press, Greenwich, CT, 1991), p. 245–264.

[59] J. Gess-Newsome and N. G. Lederman, *Examining Pedagogical Content Knowledge* (Kluwer Academic, Boston, 2001).

[60] J. Loughran, P. Mulhall, and A. Berry, In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice, J. Res. Sci. Teach. **41,** 370 (2004).

[61] H. Borko and R. T. Putnam, Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development, in *Professional Development in Education: New Paradigms and Practices*, T. R. Guskey and M. Huberman (Teachers College Press, New York, 1995).

[62] C. L. Ebert, An assessment of prospective secondary teachers' pedagogical content knowledge about functions and graphs, *Paper presented at the annual meeting of the American Educational Research Association*, (Atlanta, GA, 1993); http://eric.ed.gov/?id=ED366580.

[63] A. N. Geddis, B. Onslow, C. Beynon, and L. Oesch, Transforming content knowledge: Learning to teach about isotopes, Sci. Educ. **77,** 575 (1993).

[64] J. H. Van Driel and O. De Jong, Investigating the development of preservice teachers' pedagogical content knowledge, *Paper presented at the annual meeting of the National Association for Research in Science Teaching*, (St. Louis, MO, 2001); http://eric.ed.gov/?id=ED458146.

[65] N. G. Lederman, J. Gess-Newsome, and M. S. Latz, The nature and development of preservice science teachers' conceptions of subject matter and pedagogy, J. Res. Sci. Teach. **31,** 129 (1994).

[66] H. Borko and R. Putnam, Learning to teach, in *Handbook of Educational Psychology*, edited by D. Berliner and R. Calfee (Macmillan, New York, 1996), pp. 673–708.

[67] H. Akkoç and S. Yeşildere, Investigating development of pre-service elementary mathematics teachers' pedagogical content knowledge through a school practicum course, Procedia Soc. Behav. Sci. **2**, 1410 (2010).

[68] K. Carter, The place of story in the study of teaching and teacher education, Educ. Res. **22**, 5 (1993).

[69] D. M. Kagan, Ways of evaluating teacher cognition: Inferences concerning the Goldilocks Principle, Rev. Educ. Res. **60**, 419 (1990).

[70] J. J. Loughran, R. F. Gunstone, A. Berry, P. Milroy, and P. Mulhall, Science cases in action: Developing an understanding of science teachers' pedagogical content knowledge, *Proceedings of the Annual Meeting of the National Association for Research in Science Teaching* (New Orleans, LA, 2000); http://eric.ed.gov/?id=ED442630.

[71] J. A. Baxter and N. G. Lederman, Assessment and Measurement of Pedagogical Content Knowledge, in *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education*, edited by J. Gess-Newsome and N. G. Lederman (Kluwer Academic, Dordrecht, 1999).

[72] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Upper Saddle River, Prentice Hall, NJ, 1998).

[73] AIP Statistics: First-Year Graduate Physics Students, available online at: https://www.aip.org/statistics/physics-trends/first-year-graduate-physics-students.

[74] AIP Statistics: Physics PhDs Awarded in the U.S., available online at: https://www.aip.org/statistics/physics-trends/physics-phds-awarded-us.

[75] G. V. Glass and K. D. Hopkins, *Statistical Methods in Education & Psychology* (Allyn & Bacon, Boston, MA, 1996).

[76] A. Maries and C. Singh, Exploring one aspect of pedagogical content knowledge of teaching assistants using the test of understanding graphs in kinematics, Phys. Rev. ST Phys. Educ. Res. **9**, 020120 (2013).

[77] A. Maries and C. Singh, Performance of graduate students at identifying introductory students' difficulties with kinematics graphs, *Proceedings of the 2014 Physics Education Research Conference, Minneapolis, MN*, edited by A. Churukian, P. Engelhardt, and D. Jones (2015), p. 171 [arXiv:1601.04354].

[78] J. R. Thompson, W. M. Christensen, and M. C. Wittmann, Preparing future teachers to anticipate student difficulties in physics in a graduate-level course in physics, pedagogy, and education research, Phys. Rev. ST Phys. Educ. Res. **7**, 010108 (2011).

[79] D. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, s12 (2001).

[80] C. Singh and D. Rosengrant, Multiple-choice test of energy and momentum concepts, Am. J. Phys. **71**, 607 (2003).

[81] L. Rimoldini and C. Singh, Student understanding of rotational and rolling motion concepts, Phys. Rev. ST Phys. Educ. Res. **1**, 010102 (2005).

[82] C. Singh, Student understanding of symmetry and Gauss's law of electricity, Am. J. Phys. **74**, 923 (2006).