

Investigating the effect of question-driven pedagogy on the development of physics teacher candidates' pedagogical content knowledge

Marina Milner-Bolotin, Davor Egersdorfer, and Murugan Vinayagam

*Department of Curriculum and Pedagogy, Faculty of Education, The University of British Columbia,
2329 West Mall, Vancouver, British Columbia V6T 1Z4, Canada*

(Received 29 April 2016; published 7 September 2016)

This paper describes the second year of a multi-year study on the implementation of Peer Instruction and PeerWise-inspired pedagogies in a physics methods course in a teacher education program at a large research university in Western Canada. In the first year of this study, Peer Instruction was implemented consistently in the physics methods course and teacher candidates were asked to submit five conceptual multiple-choice questions as a final assignment. In the second year of the study we incorporated PeerWise online tool to facilitate teacher candidates' design of conceptual questions by allowing them to provide and receive feedback from their peers, and consequently improve their questions. We have found that as a result of this collaboration teacher candidates improved their pedagogical content knowledge as measured by the rubric developed for the study.

DOI: [10.1103/PhysRevPhysEducRes.12.020128](https://doi.org/10.1103/PhysRevPhysEducRes.12.020128)

I. INTRODUCTION: ADDRESSING THE CHALLENGES OF PHYSICS TEACHER EDUCATION

This study aims to investigate how an online collaborative educational technology implemented in a physics methods course can support physics teacher candidates (TCs) in acquiring relevant pedagogical content knowledge (PCK) [1]. The PCK concept was introduced by the former AERA President, Lee Shulman, who in his 1985 presidential address suggested to view teachers' knowledge as a combination of content and pedagogical knowledge (CK and PK, respectively). Shulman referred to PCK as knowledge *for teaching specific content*, which among other dimensions includes knowledge of content, multiple ways of representing it, and supporting students in overcoming potential difficulties in learning it. In the last three decades a number of studies have been conducted to expand and clarify different PCK components [2]. For this study, the PCK clarification by Magnusson *et al.* is especially relevant, as in addition to the original PCK construct suggested by Shulman, they explicitly focus on teacher's knowledge of teaching purposes and relevant assessment strategies [3]. The importance of teachers' grasp of various ways of assessing student learning, often referred to as evaluation *for*, *as*, and *of*, learning has been emphasized by both researchers and policy makers, which can be seen in the documents produced by the National Education Association and other bodies [4,5]. Since evaluation is

often question driven, it is not surprising that a key element of PCK is teacher's ability to ask questions that elicit student conceptual difficulties and promote meaningful understanding [6].

The concept of PCK has been widely used in mathematics and science (M&S) teacher education [2,7–11]. Yet few North American teacher graduates have acquired PCK sufficient for effective teaching [12,13]. In Canada, we find a number of reasons for this persisting problem. First, most secondary teacher education programs, as post-baccalaureate professional certifications, are comparatively short, lasting between 8 and 24 months [14]. Second, teacher education programs often assume that most TCs have already mastered the necessary CK, and only lack the general PK. However, substantial research indicates that the M&S knowledge of many North American B.Sc. graduates is rather limited [15,16]. Even though a significant number of TCs are lacking the CK necessary for teaching M&S, subject-specific methods courses represent less than 10% of the total time devoted to teacher education [14]. Third, there is an increased emphasis on general education courses that are often divorced from the subject-specific context and are intended to equip TCs with generic pedagogical strategies. Yet ample research has demonstrated that applying general teaching techniques to a subject-specific context is a difficult task for novice teachers to master [10].

This lack of attention to the development of PCK of future M&S teachers creates significant challenges. Numerous international student M&S assessments, such as PISA, have shown that limited teachers' PCK has direct negative consequences for student learning [17–20]. It also threatens the success of M&S educational reforms [5,21]. If we are to consider teaching as a serious profession, we need to reconsider the role of PCK in teacher education [22].

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

In addition to possessing the essential PCK components, teachers must be willing to promote active student engagement in their classrooms; this is what Magnusson *et al.* referred to as orientation to teaching science [3]. In this paper, we use the broad term *interactive engagement pedagogies* to describe student-centered pedagogies that encourage learners to take ownership of their learning [23,24]. To understand how teacher education programs can help M&S TCs develop their PCK and be open to incorporating these interactive engagement pedagogies into their teaching practice, we refer to situated learning theory [25], which will serve as the theoretical framework of the current study.

II. THEORETICAL FRAMEWORK FOR INVESTIGATING INTERACTIVE ENGAGEMENT PEDAGOGIES IN TEACHER EDUCATION

A. Situated learning theory

Situated learning theory posits that meaningful learning happens when students are engaged with authentic activities, in authentic context and culture [25]. It stems from Vygotsky's social constructivist theory of learning [26], which emphasizes both the learners' social environment and the subject-matter context. Situated learning theory states that knowledge cannot be transferred from authority, such as teachers or textbooks, to the learners. Instead, it asserts that knowledge is constructed by the learners who become active members of authentic learning environments in which they co-create the knowledge and experience its applicability [27]. According to this theory, students' social interactions with peers and with teachers are crucial for learning.

This theory is particularly relevant to teacher education. In order for TCs to be ready and willing to adopt novel learning environments in their own teaching, they have to experience them as *learners* and to reflect on their pedagogical effectiveness as *teachers* [6]. Subject-specific methods courses are natural and safe environments for TCs to do so. One of these pedagogies is called Peer Instruction [28]. It is powered by student discussions of conceptual multiple-choice questions that utilize common misconceptions as plausible but incorrect alternatives (distractors) [29]. Designing these conceptual questions and providing feedback to the conceptual questions contributed by their peers, while being supported by more experienced educators plays a vital role in helping TCs construct their PCK. There are a number of reasons for that. First, coming up with high-quality conceptual multiple-choice questions is not easy [30,31]. The process of formulating these questions helps TCs improve their knowledge of the subject matter and of the relevant curriculum (content knowledge). In the process of coming up with questions and responding to their peers' questions, TCs become aware of their own misunderstandings and gain a new appreciation of the potential student difficulties (content-specific pedagogical knowledge). As future teachers, they also become aware of

the importance of conceptual questioning in science, as opposed to asking questions about factual knowledge. As TCs become more familiar with the process, they gradually become comfortable with viewing assessment as a learning opportunity as opposed to the means of summative evaluation [6]. We will discuss the role of questioning in M&S education in the next section.

B. The role of questioning in mathematics and science teaching and teacher education

Since Socrates, educators have relied on questions to gauge student knowledge, promote comprehension, and encourage critical thinking. Well-crafted questions arouse curiosity, generate meaningful peer interactions, and help students reach new insights [32,33].

Unfortunately, it seems that teachers often ask lower-level questions that promote factual information recall, rather than higher-order thinking. While lower-level questions have their value, effective questions should span across all cognitive domains, depending on the desired learning outcomes. By using a variety of questions, teachers can develop pedagogical strategies suitable for different classroom scenarios and diverse student populations.

Multiple-choice questions are frequently used by teachers for both practical and pedagogical reasons [Fig. 1]. Despite a popular mistrust of this form of assessment, well-crafted multiple-choice questions can promote meaningful learning [31,34].

It was found that when designing these conceptual multiple-choice questions, it is crucial that they include plausible distractors targeting common student misconceptions [32]. These types of questions require teachers to understand the relevant science content and be able to anticipate student potential difficulties with learning it.

The difficult and creative work of questioning is a skill that teachers should begin acquiring in teacher education and continue improving during their teaching [35]. However, to the best of our knowledge, few teacher education programs focus on helping TCs develop necessary skills for asking conceptual M&S questions [36]. In order to see how teacher education programs can help TCs acquire these skills, we turn to a family of instructional methods called *interactive engagement pedagogies*. In the next section we discuss how these pedagogies can be used to model effective questioning in physics teacher education.

C. Interactive engagement pedagogies in physics methods courses

Interactive engagement pedagogies comprise several teaching methods that encourage students to ask and answer questions, test new ideas, actively interact with peers and instructors, and take ownership of their learning [27]. These pedagogies incorporate extensive formative assessment that is often enabled by modern technologies [16,37].

There has been much research over the past decades on the implementation and pedagogical effectiveness of

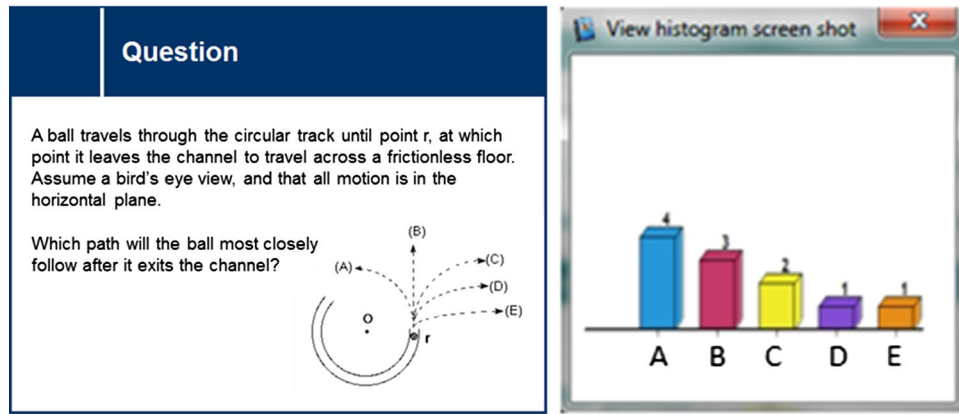


FIG. 1. An example of a conceptual mechanics multiple-choice question from Force Concept Inventory and the distribution of TCs' responses [34].

interactive engagement pedagogies within post-secondary M&S courses. These pedagogies have been shown to be more effective than traditional teacher-centered methods in enhancing student conceptual understanding, positive attitudes about M&S, and problem solving skills [38]. An interactive engagement pedagogy, often found in postsecondary physics classrooms, that utilizes multiple-choice questions is called Peer Instruction (PI) [28].

D. Peer Instruction

PI incorporates classroom response systems (clickers), various personal electronic devices, or flashcards to engage students in answering conceptual multiple-choice questions that target student difficulties by using common misconceptions as distractors (Fig. 1) [28,39]. After displaying initial responses to the question, the students are invited to discuss the question with their peers. This is followed by a repeated individual voting and an all-class discussion. PI in M&S classrooms results in increased student engagement, frequent and continuous feedback to both students and the instructor regarding the level of student understanding, and improved student learning [24]. The increased availability of emergent technologies has contributed towards promotion of interactive engagement pedagogies, and specifically PI, in K-12 classrooms [37,39,40]. However, there is extensive research evidence demonstrating that PI success lies not in the technology itself, but in the instructor's PCK [33,41]. Thus, in order to prepare future M&S teachers for successful implementation of PI and other interactive engagement pedagogies in their classrooms, these technology-enhanced interactive engagement pedagogies should be introduced in teacher education [6]. One of these new technologies, PeerWise [42], seems ideally suited for putting PI into action in teacher education.

E. PeerWise online collaborative system

PeerWise is an online platform for hosting student-authored multiple-choice questions and promoting student

engagement through online collaboration [42]. PeerWise allows students to answer, rate, and comment on multiple-choice questions created by their peers (Table I) [14]. PeerWise has chiefly been implemented in large undergraduate science courses, with results indicating that student engagement through question-creation produces positive learning outcomes [30,43,44]. The goal of these large scale (hundreds of students) studies was for students to learn the science content and enhance their academic performance, and not on learning how to ask pedagogically effective conceptual questions. While these PeerWise studies have focused on peer feedback and discussion of student-designed questions, none have discussed the use of PeerWise in teacher education, where questions were revised as a result of peer and instructor feedback. Thus our study is unique in that it focuses on investigating how physics TCs can improve their PCK, while asking, answering, critiquing, and improving conceptual multiple-choice physics questions.

As an educational technology, PeerWise is user friendly, requires minimal software learning time, and provides ample opportunities for student collaboration. Through PeerWise, the TCs are able to support each other in refining and improving their multiple-choice questions. At the end of the course, TCs also benefit from having a repository of collaboratively designed and peer-reviewed multiple-choice conceptual questions that they can use in their practicum teaching. In the following section we outline the goals and the methodology of the current study.

III. METHODOLOGY

A. Study goals

The study aimed to investigate how physics TCs' engagement with designing, answering, and commenting on conceptual multiple-choice physics questions via a PeerWise platform influenced the development of their physics PCK and openness towards using interactive engagement pedagogies in their own teaching. More

TABLE I. An example of a PeerWise question posted by a TC in the fall of 2014 and of the discussion that followed. Our interpretation of this development is posted in the right column.

	Original PeerWise content: Posted during the first week of classes in 2014	Analysis of TCs' posts
Question designed by TC G	Forces—Part 1/2: Bob and Yoshiko are having a tug-of-war on the wooden floor in the gymnasium. Bob has a mass of 78.5 kg and Yoshiko weighs 980 N. If Bob pulls with a force of 75 N and is winning the tug-of-war (pulling Yoshiko towards him), what force is Yoshiko pulling with on the rope F_y ? (Assume $g = 9.81 \text{ m/s}^2$)	It is clear that TC who posted this question intended to test student understanding of Newton's II and III laws. TC also uses the words "weight" and "mass" to make sure students understand the difference between the terms. Lastly, judging by the question's title it is the first question in a sequence of questions on forces, thus opening doors for further discussion.
A	70 N	While it is hard to judge where the distractors come from, it is clear that they are purposefully chosen to be less and more than the correct value. It is unclear why the TC chose these specific distractors.
B (correct)	75 N	
C	80 N	
D	55 N	
Explanation	No matter who is winning the tug-of-war the forces applied on either end of the rope <i>must</i> be equal in magnitude, but opposite in direction. Therefore the force applied by Yoshiko on the rope is $F_y = 75 \text{ N}$.	The author's explanation supports the correct answer. However, it is unclear if the TC can anticipate students' potential problems. They just state that the forces <i>must</i> be equal, but they don't explain why one student wins the tug-of-war, while the other loses.
Comment 1	Instructor: As I mentioned in class, it is a fantastic question. However, in the explanation, you should mention the reasons why one person is going to win. I guess you will be doing it in the second part of the question. A picture could have made it more fun as well.	The instructor notices this point and offers to expand on this concept in the consequent question. The problem was also discussed in class.
Comment 2	Can you explain more for me? I don't get why forces applied on either end of the rope must be equal when Bob is winning...(TC F)	One of the TCs actually raises this question. She asks this specific question, indicating what she does not understand.
Comment 3	Sorry, but I need a more detailed explanation. If the forces on the rope are equal which force is it that gives Bob the ability to pull Yoshiko towards him? (TC C)	Another TC also asks for a more detailed explanation. Both TCs C and F forget that the students pulling on the rope are experiencing frictional forces by the ground which eventually creates imbalance allowing one student to win.
Comment 4	Great conceptual question and a good improvement on the previous version. I would recommend putting in a picture to make things clearer (especially in the explanation). Also, perhaps elaborate a bit more in your explanation, and maybe mention why you put in the different distractors? The question is a great way to test students' knowledge of Newton's III Law. (Teaching Assistant)	The teaching assistant models constructive feedback: he indicates positive aspects of the question and suggests some ways for improvement. He also clearly indicates the physics concept behind the question (Newton's III law).

specifically, our objective was to determine if physics TCs' PCK improved over the course of an academic term as a result of a PeerWise intervention. Therefore, the research hypothesis of the study was formulated as follows:

Research hypothesis: The pedagogical content knowledge of physics TCs, as measured by the multidimensional rubric shown in Table II, has improved significantly as a result of TCs' participation in designing and developing conceptual multiple-choice questions using the PeerWise tool during a 13-week long physics methods course.

To test this research hypothesis we designed the following study.

B. Study context

This action research study [45] took place in a secondary physics methods course in a teacher education program at a large research university in Western Canada. Eight TCs were enrolled in this required physics methods course in the fall of 2014. The three-credit course took place over a period of 13 weeks, with a break during the ninth and tenth weeks for a short school practicum. Each TC had at least a bachelor's degree in physics, chemistry, or engineering. In Canada, physics is taught in grades 8–10 as part of general science studies, and in grades 11 and 12 as a separate subject. The secondary physics methods course focused on the grade 11 and 12 physics curriculum, and was designed to help TCs acquire physics-related PCK. The course was taught by the first author of this study, while the second author was the course teaching assistant. The course was designed on the pass or fail basis, such as the risk of failing the course was extremely low if not nonexistent. Historically, it is very unusual for a TC to fail a method course, as compared to failing a three-month-long school practicum. PI pedagogy was modeled in most classes to facilitate discussions about physics concepts, the different approaches students might take and where they might experience difficulties, and how a teacher could help them overcome these challenges.

Questions in class were presented through PI, and utilized clickers to allow TCs to answer them in a risk-free anonymous manner [28,45]. The instructor presented the TCs with a question that was carefully chosen by her to engage common student difficulties with fundamental physics concepts. Each TC then considered the problem on their own, and submitted an answer using a clicker. The answers were then displayed via a histogram, such that the fraction of the class giving each answer could be seen [Fig. 1].

Conceptual questions, which used common misconceptions as distractors, produced voting results that revealed TCs' knowledge gaps and misunderstandings. However, after discussing their answers with their peers and voting again on the question, TCs were able to figure out the correct response, experience potential student difficulties, and generate multiple ways of explaining relevant concepts.

Finally, any issues that were raised by TCs during this process were resolved with a class discussion facilitated by the instructor. These discussions also focused on how the TCs would utilize these PI pedagogies in their own classrooms. This allowed TCs to experience PI pedagogy both as learners and as teachers [6,45].

As part of the course, TCs were required to complete an assignment where they designed (or modified existing) conceptual multiple-choice physics questions and uploaded them on PeerWise. These questions had to incorporate meaningful distractors, an explanation of the correct answer, and the logic behind each distractor. TCs were also asked to answer questions posted by their peers and provide constructive comments on these questions. After receiving comments on their own questions, the TCs had to address these comments by modifying and improving their own questions [Table III].

The physics methods course required TCs to participate in an online discussion through commenting on conceptual questions posted by other TCs on PeerWise, as well as responding to these comments. Studies show that online discussions can stimulate critical thinking and promote cognitive engagement [46,47]. The asynchronous nature of an online as compared to a face-to-face discussion provides students with more reflection time, allowing them to carefully construct and craft their responses, which is especially relevant to future teachers [48]. The process of translating their ideas, claims, and opinions into well-articulated verbal, graphical, and symbolic expressions helps students to sharpen their arguments and improve communication skills. On the other hand, face-to-face discussions are especially beneficial for promoting critical analysis, reasoning and argumentation. In fact, it has been shown that online and face-to-face discussions comprise differing yet complementary natures of discourse, which suggests that a combination of both would be the most beneficial for TCs [47]. That is why in our course we monitored TCs' comments on PeerWise and followed up with class discussions when we saw that they had trouble grasping concepts. This meant that the content covered in our semiweekly classes was heavily informed by what the TCs were doing online on PeerWise.

C. Operationalizing PCK: Research instruments

In order to evaluate the quality of TCs' PeerWise questions, we made use of a quantitative assessment rubric modified from our previous study [Table II] [6]. The rubric was explicitly developed for rating the quality of conceptual multiple-choice questions. The ratings were based on the following seven PCK-inspired dimensions: question's cognitive level, explanation's cognitive level, question targets students' difficulties, science accuracy, distractors' quality, tcs' justification of the answer, and clarity of the question (Table II). Each measure was evaluated on a five point Likert scale. The cognitive level

TABLE II. A rubric for evaluating pedagogical and content effectiveness of conceptual questions. For the purpose of this study, the content knowledge part of the rubric comprises both content knowledge and pedagogical content knowledge aspects. Note: CA—Correct answer and IA—Incorrect answer.

Level	Content knowledge					Pedagogical knowledge	
	Question's cognitive level	Explanation's cognitive level	Question targets students' difficulties	Science accuracy	Distractors' quality	TCS' justification of the answer	Clarity of the question
1	Knowledge	Knowledge	Doesn't target any concept difficulty	Major mistakes in the question and in solutions	All irrelevant distractors	No answer justification	Unclear, misleading and inaccurate
2	Comprehension	Comprehension	Targets a minor concept ineffectively	An accurate question but an inaccurate and unclear solution		Incomplete CA justification; no IA justification	
3	Application	Application	Targets a minor concept effectively	The question is clear; the solution is accurate but unclear	Half of distractors are meaningful	Incomplete CA and IA justification	Minor problems in the question or in solution
4	Analysis	Analysis	Targets a few conceptual difficulties	Both question and solution are somewhat clear and accurate		Complete and accurate CA justification only	
5	Synthesis and Evaluation	Synthesis and Evaluation	Clearly targets major conceptual difficulties	Both question and solution are clear and accurate	All distractors are meaningful	Complete and accurate CA and IA justification	Both question and solution are very clear and accurate

TABLE III. PeerWise pedagogical cycle sequence.

Cycle	Week	Expected tasks
1	1	Author at least six questions and upload them to PeerWise
	2	Answer at least 10 questions designed by your peers Provide comments on at least five of the questions you answered
2	3	Review all relevant comments on your questions and respond when applicable Modify at least three of your questions to address the comments Author at least three additional questions
	4	Answer at least 10 questions designed by your peers Provide comments on at least five questions you answered

Cycle 2 repeats every two weeks for the duration of the course.

of the questions and explanations were assessed using Bloom’s taxonomy [49,50]. There is a precedent of using Bloom’s taxonomy to rate the quality of multiple-choice questions, specifically questions that were created on PeerWise [43]. For each question, we identified if answering the question required retrieval of factual knowledge (i.e., what unit is used for measuring forces?), comprehension of the concept through connecting multiple representations, application of the concepts outside of the context they were initially learned, or synthesis and evaluation of multiple concepts and ideas. Following up our previous study, we grouped the dimensions of the rubric into two categories—content knowledge and pedagogical knowledge [6]. However, it is clear that these dimensions are inherently overlapping, thus creating the PCK construct. For example, to design a question that targets students’ conceptual difficulties and possible misconceptions, a teacher needs to acquire both content and pedagogical knowledge (Table II). Thus, this dimension could have belonged to both content and pedagogical knowledge categories. For the purpose of the analysis we decided to place this category under the content knowledge construct. This had no limiting consequences for our analysis as we investigated PCK in its entirety.

There were three researchers in this study who served as raters of the questions: the course instructor, the teaching assistant, and a research assistant. The instructor was the leader of the study. She is a physics educator with more than 20 years of secondary and post-secondary physics teaching and educational research experience, 10 of those years being spent in teacher education. The teaching assistant holds a B.Sc. degree in molecular and cell biology and a M.Ed. in science education, and had also taken introductory university-level physics courses. The research assistant holds a M.Sc. degree in mathematics and a M.Ed. in mathematics education. The researchers independently rated each of the questions using the rubric. To ensure consistent rating, initially only the first ten questions were independently evaluated by the researchers. The ratings were then compared, and in each case where discrepancies existed, the ratings were discussed and resolved. The aim was to ensure that the rating process was clear, as little

biased as possible, and all the consequent ratings by the team were consistent. The diverse backgrounds of the researchers meant that each brought different experiences and insights to the rating process. In times when the agreement on a rating could not be reached, the lead researcher’s argument was given more weight due to her extensive experience. Once a baseline for the rating had been set, the rest of the questions were rated independently without further discussion. This process was implemented in order to improve the consistency of the independent ratings by the three researchers. The resulting estimate of the interrater reliability demonstrated an overall degree of agreement among the raters, which is described in the following section.

D. Analysis of the validity and reliability of the instrument

While using statistical instruments, researchers must make sure that the instruments are reliable and valid for the study’s population. Below we provide information on how we measured the interrater reliability and internal reliability of the instrument (Tables VI and VII in the Supplemental Material [51]), as well as how we calculated the statistical significance of the research findings.

1. Statistical analyses

To determine the statistical significance of the differences between the pre- and postintervention scores of the PCK of physics TCs (as based on their multiple-choice conceptual questions), we conducted Wilcoxon signed ranks tests (two related samples) across the seven dimensions (Tables IV and V). We provide the discussion of

TABLE IV. Question data collection times.

Time period	Corresponding week	Number of questions collected
Time 1	Weeks 2–3	51
Time 2	Weeks 11–13	52

TABLE V. Wilcoxon signed ranks tests (two related samples).

		Ranks			Test Statistics	
		<i>N</i>	Mean Rank	Sum of Ranks	<i>Z</i>	Asymp. Sig. (2-tailed)
QuestionCL_Post- QuestionCL_Pre	Negative ranks	13 ^a	18.23	237.00	Based on negative ranks	0.004
	Positive ranks	30 ^b	23.63	709.00		
	Ties	8 ^c				
	Total	51				
QuestionClarity_Post- QuestionClarity_Pre	Negative ranks	9 ^d	19.67	177.00	Based on negative ranks	0.000
	Positive ranks	34 ^e	22.62	769.00		
	Ties	8 ^f				
	Total	51				
ExplanationCL_Post- ExplanationCL_Pre	Negative ranks	9 ^g	20.11	181.00	Based on negative ranks	0.000
	Positive ranks	36 ^h	23.72	854.00		
	Ties	6 ⁱ				
	Total	51				
QuestionTargetDiff_Post- QuestionTargetDiff_Pre	Negative ranks	9 ^j	18.72	168.50	Based on negative ranks	0.002
	Positive ranks	30 ^k	20.38	611.50		
	Ties	12 ^l				
	Total	51				
ScienceAccuracy_Post- ScienceAccuracy_Pre	Negative ranks	9 ^m	20.11	181.00	Based on negative ranks	0.000
	Positive ranks	37 ⁿ	24.32	900.00		
	Ties	5 ^o				
	Total	51				
DistractorsQual_Post- DistractorsQual_Pre	Negative ranks	6 ^p	4.33	26.00	Based on negative ranks	0.000
	Positive ranks	37 ^q	24.86	920.00		
	Ties	8 ^r				
	Total	51				
TCJustificationAns_Post- TCJustificationAns_Pre	Negative ranks	12 ^s	14.29	171.50	Based on negative ranks	0.000
	Positive ranks	32 ^t	25.58	818.50		
	Ties	7 ^u				
	Total	51				

^aQuestionCLPost < QuestionCLPre |^bQuestionCLPost > QuestionCLPre |^cQuestionCLPost = QuestionCLPre |^dQuestionClarityPost < QuestionClarityPre |^eQuestionClarityPost > QuestionClarityPre |^fQuestionClarityPost = QuestionClarityPre |^gExplanationCLPost < ExplanationCLPre |^hExplanationCLPost > ExplanationCLPre |ⁱExplanationCLPost = ExplanationCLPre |^jQuestionTargetDiff Post < QuestionTargetDiff Pre |^kQuestionTargetDiff Post > QuestionTargetDiff Pre |^lQuestionTargetDiff Post = QuestionTargetDiff Pre |^mScienceAccuracyPost < ScienceAccuracyPre |ⁿScienceAccuracyPost > ScienceAccuracyPre |^oScienceAccuracyPost = ScienceAccuracyPre |^pDistractorsQualPost < DistractorsQualPre |^qDistractorsQualPost > DistractorsQualPre |^rDistractorsQualPost = DistractorsQualPre |^sTCJustificationAnsPost < TCJustificationAnsPre |^tTCJustificationAnsPost > TCJustificationAnsPre |^uTCJustificationAnsPost = TCJustificationAnsPre |

TABLE VI. Item analysis from SPSS output.

	<i>N</i>		Mean	Variance	SD	
	7		26.605	19.307	4.394	
Statistics for scale	Mean	Minimum	Maximum	Range	Max/Min	Variance
Item means	3.801	2.790	4.392	1.602	1.574	0.334
Item variances	0.770	0.446	1.240	0.793	2.777	0.068
Inter-item covariances	0.331	0.114	0.604	0.491	5.311	0.017
Inter-item correlations	0.435	0.187	0.780	0.593	4.165	0.020

Total item statistics	Scale mean if item deleted	Scale variance if item deleted	Corrected item total correlation	Squared multiple correlation	Cronbach's Alpha if item deleted
QuestionCL	23.3916	16.539	0.427	0.446	0.841
QuestionClarity	22.2136	14.984	0.641	0.647	0.814
ExplanationCL	23.8155	15.065	0.585	0.484	0.821
QuestionTargetDiff	22.4693	14.363	0.596	0.479	0.819
ScienceAccuracy	22.4660	14.046	0.688	0.692	0.805
DistractorsQual	22.6926	13.263	0.592	0.387	0.825
TCJustificationAns	22.5825	13.662	0.672	0.558	0.807
Reliability coefficient for 7 Items	Cronbach alpha 0.841			Standardized Cronbach alpha 0.843	

the statistical results of these tests in the results section below.

2. Interrater reliability

In order to demonstrate that the ratings of the three independent raters were consistent and did not have a negative effect on the quality of the ratings, the interrater reliability (IRR) statistic was computed. The IRR statistic is widely used in studies where trained raters are required for

analyzing collected data. Since the raters were not randomly sampled from a larger population of raters, IRR was assessed by using two-way mixed, consistency, average-measures intraclass correlation coefficients (ICC). Hence, this ICC was used to assess the consistency of the ratings on the seven dimensions across all the multiple-choice questions, as shown in Table VI in the Supplemental Material [51]. According to Cicchetti [52], the resulting ICC average measures were in the excellent range for all

TABLE VII. Intraclass correlation coefficient.

Dimension		Intraclass correlation ^b	95% confidence interval		<i>F</i> test with true value 0			
			Lower Bound	Upper Bound	Value	df1	df2	Sig.
QuestionCL	Single measures	0.628 ^a	0.529	0.717	6.067	102	204	0.000
	Average measures	0.835 ^c	0.771	0.884	6.067	102	204	0.000
QuestionClarity	Single measures	0.782 ^a	0.713	0.839	11.739	102	204	0.000
	Average measures	0.915 ^c	0.882	0.940	11.739	102	204	0.000
ExplanationCL	Single measures	0.802 ^a	0.738	0.854	13.138	102	204	0.000
	Average measures	0.924 ^c	0.894	0.946	13.138	102	204	0.000
QuestionTargetDiff	Single measures	0.918 ^a	0.888	0.941	34.448	102	204	0.000
	Average measures	0.971 ^c	0.960	0.979	34.448	102	204	0.000
ScienceAccuracy	Single measures	0.780 ^a	0.711	0.837	11.630	102	204	0.000
	Average measures	0.914 ^c	0.881	0.939	11.630	102	204	0.000
DistractorsQual	Single measures	0.886 ^a	0.847	0.918	24.423	102	204	0.000
	Average measures	0.959 ^c	0.943	0.971	24.423	102	204	0.000
TCJustificationAns	Single measures	0.817 ^a	0.757	0.866	14.376	102	204	0.000
	Average measures	0.930 ^c	0.903	0.951	14.376	102	204	0.000

Two-way mixed effects model where question effects are random and measures effects are fixed.

^aThe estimator is the same, whether the interaction effect is present or not.

^bType C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

^cThis estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

dimensions, (poor: less than 0.40; fair: 0.40–0.59, good: 0.60–0.74; excellent: 0.75–1.00), indicating that the three independent raters had a high degree of agreement and also suggesting that all the dimensions were rated consistently. Measures of ICC in the excellent range suggest that a minimal amount of rating error was introduced by the three independent raters. Therefore, the statistical power for subsequent analyses is not substantially reduced. Consequently, the ratings for all the dimensions were deemed to be suitable for use in the hypothesis testing for this study.

3. Internal reliability

A preexisting instrument designed and developed by the team [6] was used for rating all the multiple-choice questions [Table II]. In order to confirm the robustness of the instrument's ability to evaluate the TCs' PCK based on their conceptual multiple-choice questions, internal consistency of the seven dimensions was calculated by using Cronbach's alpha reliability estimate. It is commonly used as a statistical index of internal consistency for gauging whether items (the seven dimensions) in an instrument measure the same construct (TCs' PCK) of interest when designing, developing, or refining an instrument. SPSS 21 was used to estimate the Cronbach alpha statistic across 103 multiple choice questions for the seven dimensions, as shown in Table VI in the Supplemental Material [51]. The overall Cronbach alpha reliability estimate (based on standardized items) for the instrument was evaluated to be 0.843, which indicates a good internal consistency of the dimensions in the instrument.

E. Data collection

Data were collected using the PeerWise online collaborative tool described above [42]. The TCs submitted 298 questions, 844 answers with explanations, and 329 comments during the term, and these were stored in a central database that the supervisor and teaching assistant had access to. In order to prevent bias and protect the privacy of all the TCs, all personal details were anonymized by assigning codes. The questions were then exported into a pdf document, which was then used for further analysis. Since our goal was to evaluate the change in TCs' PCK, we chose two time periods that corresponded to the beginning and the end of the course: time 1 included questions submitted in weeks 2 to 3 of the course; time 2 included questions submitted in weeks 11 to 13 of the course (Table IV). We did not include questions from the first week of the course as the TCs were still familiarizing themselves with PeerWise and what was required of them in the question creation process. Likewise, we did not include the last week of the course as the TCs had many assignments due at this time, and we feel that as a result of this pressure many of their final questions were rushed and

of poorer quality. Thus, we evaluated 103 conceptual multiple-choice questions out of 298 questions (35%) submitted by TCs.

IV. RESULTS AND DISCUSSION

One of our objectives was to determine whether TCs had improved their PCK as measured by the rubric as a result of developing conceptual physics multiple-choice questions over the course of a 13-week long academic term. Statistical analyses were based on 103 multiple choice questions (51 preintervention and 52 postintervention questions) (Table IV).

The ratings, based on TCs' PCK, were analyzed for statistically significant differences between pre- and post-intervention scores. We used both graphical (histograms, normal Q-Q plots, and normality curves) and numerical (Shapiro-Wilk, Kolmogorov-Smirnov, skewness, and kurtosis) methods to assess the center, shape, and spread of the data distributions across all the dimensions for both samples. Data distributions for all the seven dimensions reflected the likelihood of deviating from normality (Tables VIII and IX in the Supplemental Material [51]). Because of the non-normal data distributions of the pre- and postintervention scores, nonparametric tests were considered for further statistical analysis. The Wilcoxon Signed Ranks tests were used to analyze the differences between the two related samples. Table V shows the Wilcoxon Signed Ranks test results for the seven dimensions underpinning the TCs' PCK. These test results indicate that postintervention scores showed statistically significant increases in all the dimensions that measure the TCs' PCK. For each dimension, about 60% of the post-intervention questions showed an improvement, and all seven dimensions had a significantly higher number of positive rankings (indicating significant improvement) (Table V).

In summary, these results strongly suggest that for all seven dimensions of the rubric (Table II), the ratings of the multiple-choice questions produced by the TCs improved significantly during the methods course. All of the ratings were significantly higher at the post-test as compared to the pretest ($p < 0.005$). Therefore, we can justifiably conclude that TCs were able to improve their skills in asking meaningful conceptual multiple-choice questions, in designing scientifically meaningful distracters, and providing scientifically plausible and pedagogically sound explanations to their questions. This means that over the course of 13 weeks, TCs were able to acquire skills that enabled them to produce more sophisticated (higher level on Bloom's taxonomy) multiple-choice conceptual questions, which is one of the manifestations of their improved PCK.

From our experience with TCs and observations of the dynamics of the physics methods course, we think that a number of factors might have contributed to these results. First, pedagogies such as Peer Instruction draw TCs'

TABLE VIII. Descriptive statistics and normality tests for preintervention.

	N		Mean		Standard deviation		Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
QuestionCL	51		3.0458	0.0895496	0.63951		-0.592	0.333	1.183	0.656
QuestionClarity	51		4.0654	0.1108571	0.79168		-0.476	0.333	-0.422	0.656
ExplanationCL	51		2.4641	0.1072851	0.76617		-0.051	0.333	-0.395	0.656
QuestionTargetDiff	51		3.8497	0.1267367	0.90508		-0.742	0.333	0.852	0.656
ScienceAccuracy	51		3.7647	0.1170792	0.83611		-0.285	0.333	-0.809	0.656
DistractorsQual	51		3.2941	0.1624022	1.15978		-0.304	0.333	-1.185	0.656
TCJustificationAns	51		3.6863	0.1407765	1.00535		-0.304	0.333	-1.001	0.656
Valid N (listwise)	51									

Tests of normality for preintervention

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
QuestionCL	0.195	51	0.000	0.908	51	0.001
QuestionClarity	0.156	51	0.003	0.906	51	0.001
ExplanationCL	0.189	51	0.000	0.924	51	0.003
QuestionTargetDiff	0.233	51	0.000	0.885	51	0.000
ScienceAccuracy	0.134	51	0.024	0.945	51	0.019
DistractorsQual	0.219	51	0.000	0.897	51	0.000
TCJustificationAns	0.113	51	0.125	0.914	51	0.001

^aLilliefors significance correction.

TABLE IX. Descriptive statistics and normality tests for postintervention.

Descriptive statistics for postintervention										
	N		Mean		Standard deviation		Kurtosis			
	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error		
QuestionCL	52		3.3782	0.0915524	0.66019		-1.139	0.330	1.845	0.650
QuestionClarity	52		4.7115	0.0782266	0.56410		-2.113	0.330	3.865	0.650
ExplanationCL	52		3.1090	0.0958853	0.69144		-0.142	0.330	-0.287	0.650
QuestionTargetDiff	52		4.4167	0.1153811	0.83203		-1.445	0.330	1.320	0.650
ScienceAccuracy	52		4.5064	0.1037629	0.74825		-1.573	0.330	1.388	0.650
DistractorsQual	52		4.5192	0.0879984	0.63457		-1.015	0.330	-0.073	0.650
TCJustificationAns	52		4.3526	0.1076952	0.77660		-1.388	0.330	1.659	0.650
Valid N (listwise)	52									

Tests of normality for preintervention

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
QuestionCL	0.173	52	0.000	0.879	52	0.000
QuestionClarity	0.407	52	0.000	0.585	52	0.000
ExplanationCL	0.140	52	0.013	0.961	52	0.086
QuestionTargetDiff	0.297	52	0.000	0.734	52	0.000
ScienceAccuracy	0.284	52	0.000	0.705	52	0.000
DistractorsQual	0.353	52	0.000	0.744	52	0.000
TCJustificationAns	0.202	52	0.000	0.800	52	0.000

^aLilliefors significance correction.

attention to their own gaps in science content knowledge. As TCs become aware of their own knowledge deficiencies they start paying closer attention to their own knowledge, as well as to the knowledge of their peers.

Second, PeerWise provided a safe online forum for TCs to explore conceptual questions, relevant explanations, as well as potential student difficulties. Many of the comments initiated on PeerWise sparked conversations within the classroom and were incorporated into in-class discussions during the methods course, which allowed for more in-depth conversations. In addition, as TCs were asked to revise their questions, they paid close attention to the feedback provided by their peers, by the instructor and by the teaching assistant, as well as to the face-to-face in-class discussions.

Third, as each of the TCs was required to provide feedback to questions submitted by their peers, TCs learned how to both provide and receive feedback in a positive and constructive manner. We often observed discussions on PeerWise where TCs responded to the feedback by justifying their reasoning and clarifying their initial ideas. This was also timely as TCs had to be prepared to receive ample feedback from their supervising teachers during the practicum.

The results of both the quantitative and the qualitative (written course evaluations, classroom discussions, and the write-ups of the focus group meetings) data analyses complement each other to provide robust results on the growth of TCs' PCK over the course of the term.

TCs' thoughts on PeerWise:

TC A: "...I felt my multiple choice question writing was greatly improved and I received commendations on it from my School Advisors (school teachers supervising TCs during their practicum)..."

TC B: "I think for myself,...getting away from the number-crunching questions to conceptual questions, and understanding that even though number-crunching can feel harder, it doesn't really show how well you know something compared to conceptually, and for me PeerWise really helped me realize that. High school kids, they just want to crunch numbers and go, they don't want to think. It made me focus more on conceptual stuff."

TC C: "I think it really improved my question rating and also my assessment of questions. Like looking online and finding questions, it drastically improved my ability to go...do I want to include that in my test? What's that actually testing? And it did help. [TC B agrees]"

TCs' thoughts on the use of clickers and multiple-choice questions:

TC D: "I really liked them [clickers]. Now when people talk about clickers...I'm all over it, I know all about it. I

really like using it. I like the anonymity of it; I like the kinda just gauging where you are compared to others. It's a very useful tool."

TCs' thoughts on conceptual questions in physics:

TC D: "It gets you away from thinking physics is a math class. It gets you to really understand the science."

TC A: "The math is just applying the concepts...the conceptual part is the physics, the rest of it is just applying math to the physics concepts to get an answer."

TCs' thoughts on how the quality of their own questions has changed over the course:

TC D: "I thought they [student's multiple-choice physics questions] got way better. It was harder and harder for me to find points to nit-pick out for revisions and things when I was giving feedback."

TC B: "It got easier to write...it got faster for myself. I find it almost easier to make conceptual questions than number crunching ones."

TC D: "PeerWise, that exercise gave me the confidence to be able to go and say: ok, this question is ok, but if I reword it like this, or if I change the options from these to these, that would actually make it a better question."

TC A: "After writing all those questions and putting in a lot of work into it, especially in the beginning...but towards the end the questions were coming easier and faster. And then on my practicum, when I did go look at other resources...when I was looking at other people's questions I was like: 'that's junk, that's junk, that one could work if I could tweak it this way...ooh I'll take that one, that's garbage'...and I was able to...quickly zip through it and made my homework out of that."

V. CONCLUSIONS AND IMPLICATIONS

This paper describes the second year of a multiyear study on the implementation of PI and PeerWise pedagogies in physics methods courses in a teacher education program. In the first year of this study, PI was implemented in a physics methods course and TCs were asked to submit five conceptual multiple-choice questions as a final assignment [6]. Based on the results of that study, we concluded that TCs benefited from PI and from the opportunity to design their own conceptual multiple-choice questions as part of the course assignment [45]. However, since the question design was the final course assignment, it was difficult for TCs to share and reflect on their questions, as they were submitted in an offline format at the very end of the course. Only the instructor and teaching assistant read and analyzed all of these questions. This lack of collaboration among TCs and a chance to revise their questions was a missed opportunity. The implementation of PeerWise coupled with PI during the second year, described in this paper,

supported a deeper engagement with the questions than the first year of the study. PeerWise online engagement increased the TCs' opportunities to develop questioning skills, learn from each other and improve their PCK.

The purpose of this study was to investigate how physics TCs' engagement with designing, answering, and commenting on conceptual multiple-choice physics questions during a physics methods course that utilized PI and a PeerWise platform influenced the development of their physics PCK and attitudes towards using interactive engagement pedagogies in their own teaching. In our study TCs experienced PI on a regular basis during the course and were asked to collaborate on authoring their own conceptual multiple-choice questions using PeerWise outside of class. The cyclical nature of the course structure allowed TCs to create, evaluate, and discuss conceptual multiple-choice questions on PeerWise and explore the nuances of distractor quality, language, and interpretation during class hours using PI. This study has shown that PI enhanced by PeerWise can be successfully implemented in a physics methods course, as it helped improve TCs' PCK, engaged them in meaningful discussions inside and outside of the classroom, and promoted TCs' collaboration on designing physics teaching resources. We have also shown that during the 13-week course, TCs were able to significantly improve their physics knowledge, as well as the knowledge of physics pedagogy (PCK), through the development of more effective questioning skills as expressed in the multiple-choice questions they contributed on PeerWise.

Teacher education programs often lag in the implementation of educational technologies due to the time required for instructors to learn how these technologies can be implemented, and the lack of opportunities to use these technologies in K-12 classrooms during the practicum and consequent teaching [45,53]. It is worth mentioning that there are a number of low-tech alternatives to using clickers, such as flashcards or voting, which can also produce powerful learning [39]. The PeerWise system is free, and only requires that students have access to a computer and the Internet outside of class. Thus, the lack of access to computers during class time should not preclude M&S educators from using PI and PeerWise. PI and PeerWise-enhanced learning environments support student learning outside of the classroom, creating an opportunity for teachers to flip the classroom. These teachers could then use class time for discussing student conceptual difficulties, rather than using this precious time for information transfer [54]. This reasoning also applies to teacher education. In our study, PI and PeerWise provided an excellent opportunity to help TCs develop their questioning skills, enhance their PCK, as well as begin designing pedagogically effective teaching resources. Most importantly, these pedagogies helped to create a learning community where acquiring PCK was valued, where making a mistake was considered to be a learning opportunity, and where TCs

were encouraged to discuss physics teaching ideas with each other and the instructor.

VI. CHALLENGES AND LIMITATIONS

Because of the nature of the study, it has a number of limitations. First, it was situated in the context of physics teacher preparation. While the teacher education program at the university in question has more than 600 students, only eight of them were future physics teachers. As a result, the study had a small number of participants, which limits its generalizability. Second, the small size of the study allowed for the creation of personal connections with the participants, making it easier for the instructor to promote interactive engagement pedagogies and conceptual physics understanding during the methods course. This might prove a greater challenge for educators teaching larger methods courses. Third, the course was only 13 weeks long. Thus, we had relatively little time to support TCs in developing positive attitudes about interactive engagement and conceptual learning. It took us several weeks to help TCs understand what a good effective conceptual question is. Fourth, while we used clickers for our PI implementation, not every methods course might have them available. However, other technologies and low-tech tools such as flashcards can be used to implement PI, and current research indicates that the technology used for PI implementation does not affect its effectiveness [39]. Conversely, TCs have to realize that the availability of a particular technology is not a necessary condition for the successful implementation of PI. Fifth, the biggest challenge for our study was the two-week practicum in the middle of the physics methods course. During this time TCs interacted with various physics teachers, some of whom did not believe in interactive engagement and chose to teach traditional plug-and-chug physics courses. This created a cognitive conflict for some of the TCs as they had to negotiate the values promoted in the methods course with the pedagogical philosophies of their host teachers. Moreover, although the sophistication of the questions designed by TCs increased during the course, we did not have sufficient data to evaluate TCs' ability to articulate their pedagogical considerations in their comments. This is likely due to the nature of the course structure, where an online discussion that began on PeerWise often continued during class, thus not being formally resolved or documented on PeerWise. While we feel that this was one of the strengths of the PeerWise-enhanced PI approach, we were unable to quantitatively measure the quality of in-class conversations, and therefore cannot report on the quality of these discussions.

Lastly, we realize that it is a quasiexperimental study, as with a low number of participants (we rarely have more than 12 physics TCs in a methods course) it was impossible for us to have a control and an experimental groups. This obviously limited the generalizability of the results. There

is a possibility to conduct a study with mathematics and chemistry TCs in the future. The challenge will be to make sure that the instructors in these courses are open to using this pedagogy and the teaching methods and learning environments in different sections will be comparable.

VII. FUTURE DIRECTIONS

As we complete the second year of this study there are a number of important reflections that are guiding us as we move forward.

One of the major problems we had in implementing PeerWise-supported PI was that we could not track discussions started in PeerWise that were eventually resolved in face-to-face class meetings. Going further, we would like to capture how these discussions sparked by PeerWise questions are developed during consequent lessons. It is also important to understand if and how these discussions were incorporated in TCs' consequent teaching during the practicum.

In their written and informal feedback, TCs mentioned multiple times that they would like to have opportunities during class where they can teach mini-lessons using the PI pedagogies they learned in the course. Therefore, going into our third year we plan to make mini-lessons a part of the course structure. Furthermore, we plan to record these lessons so that TCs can reflect on their teaching styles and improve them. This will also help expose TCs to many different teaching styles, as they will be expected to watch their peers teach, and provide feedback. This will also provide opportunities to collect richer data and investigate how PI and PeerWise-enhanced pedagogy affects TCs' teaching. It would also be interesting to follow up with TCs during their 13-week long practicum and explore if and how they incorporate PI pedagogy with their students.

At the end of this physics methods course, we ended up with a large database of conceptual multiple-choice questions created by the TCs. While not all of these were high quality questions, there were many that were improved over time to create powerful conceptual physics questions. This would be an indispensable resource for beginning physics teachers. That is why in future courses we plan to create a compilation of the best questions in the course that TCs can use in their future physics teaching.

Lastly, in the future, we will focus on developing methods for evaluating the influence of the reflective practices implemented in the course on TCs' acquisition of PCK. As research on teachers' growth through reflective practice has gained a wider recognition in the last decades, it is important to investigate how the reflective practices implemented in our methods courses coupled with PeerWise and Peer Instruction influence TCs' pedagogical transformation [55,56].

ACKNOWLEDGMENTS

We would like to thank Mr. Paul Denny (the father of PeerWise), from the University of Auckland who has been kindly supporting us with PeerWise implementation during all these years. We would like to thank former graduate students Heather Fisher and Alexandra MacDonald who have been a part of the research team in the earlier stages of the project. We are especially grateful to all the physics TCs who participated in the study and provided constructive feedback for continuous course improvements. We also would like to acknowledge the financial support of the University of British Columbia Teaching and Learning Enhancement Fund. In addition, we would like to acknowledge two anonymous reviewers who provided very useful feedback on the earlier draft of the paper.

-
- [1] L. S. Shulman, Those who understand: Knowledge growth in teaching, *Educ. Res.* **15**, 4 (1986).
 - [2] H. Jing-Jing, A critical review of Pedagogical Content Knowledge' components: nature, principle and trend, *Int. J. Educ. Res.* **2**, 411 (2014).
 - [3] S. Magnusson, J. Krajcik, and H. Borko, Nature, sources, and development of pedagogical content knowledge for science teaching, in *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education*, edited by J. Gess-Newsome and N. G. Lederman (Springer Netherlands, Dordrecht, 1999), p. 95–132.
 - [4] T. K. Angelo and K. P. Cross, *Classroom Assessment Techniques: A Handbook for College Teachers*, 2nd ed., The Jossey-Bass Higher and Adult Education Series (Jossey-Bass Publishers, San Francisco, 1993).
 - [5] The National Education Association, *Teacher Assessment and Evaluation: The National Education Association's Framework for Transforming Education Systems to Support Effective Teaching and Improve Student Learning* (National Education Association, Washington, DC, 2010).
 - [6] M. Milner-Bolotin, H. Fisher, and A. MacDonald, Modeling active engagement pedagogy through classroom response systems in a physics teacher education course, *LUMAT: Research and Practice in Math, Science and Technology Education* **1**, 523 (2013).
 - [7] S. Blömeke and S. Delaney, Assessment of teacher knowledge across countries: a review of the state of research, *ZDM Math. Educ.* **44**, 223 (2012).
 - [8] M. J. Koehler and P. Mishra, Technological pedagogical content knowledge, in *The SAGE Encyclopedia of*

- Educational Technology*, edited by M. J. Spector (SAGE Publications, Los Angeles, 2015), p. 782–785.
- [9] J. H. van Driel, A. Berry, and J. Meirink, Research on science teacher knowledge, in *Handbook of Research on Science Education*, edited by N. G. Lederman and S. K. Abel (Routledge, New York, 2014), Vol. 2, p. 848–870.
- [10] J. A. Luft and P. W. Hewson, Research on teacher professional development knowledge in science, in *Handbook of Research on Science Education*, edited by N. G. Lederman and S. K. Abel (Routledge, New York, 2014), Vol. 2, p. 889–909.
- [11] M. G. Jones and M. Leagon, Science teacher attitudes and beliefs: Reforming practice, in *Handbook of Research on Science Education*, edited by N. G. Lederman and S. K. Abel (Routledge, New York, 2014), Vol. 2, p. 830–847.
- [12] D. E. Meltzer, M. Plisch, and S. Vokos, A report by the Task Force on Teacher Education in Physics, Report No. (T-TEP), 2012.
- [13] E. Etkina, Pedagogical content knowledge and preparation of high school physics teachers, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020110 (2010).
- [14] M. Milner-Bolotin, Using PeerWise to promote student collaboration on design of conceptual multiple-choice questions, *Phys. Canada* **70**, 149 (2014).
- [15] C. E. Wieman, *Issues Sci. Tech.* **1** (2012).
- [16] N. Lasry, J. Guillemette, and E. Mazur, Two steps forward, one step back, *Nat. Phys.* **10**, 402 (2014).
- [17] W. H. Schmidt *et al.*, *Teacher Education Matters: A Study of Middle School Mathematics Teacher Preparation in Six Countries* (Teachers College Press, New York, 2011).
- [18] OECD, PISA 2012 Results: What students know and can do—student performance in mathematics reading and science, 2014.
- [19] P. Brochu, M.-A. Deussing, K. Houme, and M. Chuy, Measuring up: Canadian results of the OECD PISA study: The performance of Canada’s youth in mathematics, http://cmec.ca/Publications/Lists/Publications/Attachments/318/PISA2012_CanadianReport_EN_Web.pdf.
- [20] C. M. Schroeder, T. P. Scott, H. Tolson, T. Huang, and Y. Lee, A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States, *J. Res. Sci. Teach.* **44**, 1436 (2007).
- [21] T. Feder, Europe reflects on a decade of higher education reforms, *Phys. Today* **63**, 24 (2010).
- [22] S. Tobias and A. Baffert, *Science Teaching as a Profession: Why It Isn’t How It Could Be* (Research Corporation for Science Advancement, Tucson, AZ, 2009).
- [23] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordta, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410 (2014).
- [24] R. 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).
- [25] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*, Learning in Doing: Social, Cognitive and Computational Perspectives (Cambridge University Press, Cambridge, England, 1991).
- [26] L. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).
- [27] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (The National Academies Press, Washington, DC, 2002).
- [28] E. Mazur, *Peer Instruction: User’s Manual*, Prentice Hall Series in Educational Innovation (Prentice Hall, Upper Saddle River, NJ, 1997).
- [29] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [30] J. Hardy, S. P. Bates, M. M. Casey, K. W. Galloway, R. K. Galloway, A. E. Kay, P. Kirsop, and H. A. McQueen, Student-Generated Content: Enhancing learning through sharing multiple-choice questions, *Int. J. Sci. Educ.* **36**, 2180 (2014).
- [31] J. L. Little, E. L. Bjork, R. A. Bjork, and G. Angello, Multiple-choice tests exonerated, at least of some charges fostering test-induced learning and avoiding test-induced forgetting, *Psychol. Sci.* **23**, 1337 (2012).
- [32] I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, Designing effective questions for classroom response system teaching, *Am. J. Phys.* **74**, 31 (2006).
- [33] I. D. Beatty, A. Feldman, W. J. Leonard, W. J. Gerace, K. S. Cyr, H. Lee, and R. Harris, Teacher learning of technology-enhanced formative assessment, in *Paper presented at the National Association for Research in Science Teaching, Baltimore, MD* (2008), p. 29.
- [34] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [35] M. H. Bowker, Thought & Action: The NEA Higher Educ. *J.* **26**, 127 (2010).
- [36] T. Tofade, J. Elsnor, and S. T. Haines, Best practice strategies for effective use of questions as a teaching tool, *Am. J. Pharm. Educ.* **77**, 155 (2013).
- [37] M. Milner-Bolotin, A. Kotlicki, and G. Rieger, Can students learn from lecture demonstrations: The role and place of interactive lecture experiments in large introductory science courses, *J. Coll. Sci. Teach.* **36**, 45 (2007).
- [38] J. M. Fraser, A. L. Timan, K. Miller, J. E. Dowd, L. Tucker, and E. Mazur, Teaching and physics education research: bridging the gap, *Rep. Prog. Phys.* **77**, 032401 (2014).
- [39] N. Lasry, Clickers or Flashcards: Is there really a difference?, *Phys. Teach.* **46**, 242 (2008).
- [40] C. E. Wieman, W. K. Adams, P. Loeblein, and K. K. Perkins, Teaching Physics Using PhET Simulations, *Phys. Teach.* **48**, 225 (2010).
- [41] M. K. Smith, W. B. Wood, K. Krauter, and J. K. Knight, Combining peer discussion with instructor explanation increases student learning from in-class concept questions, *CBE Life Sci. Educ.* **10**, 55 (2011).
- [42] PeerWise, (The University of Auckland, NZ) <http://peerwise.cs.auckland.ac.nz/> (Accessed April 22 2016).
- [43] S. P. Bates, R. K. Galloway, J. Riise, and D. Homer, Assessing the quality of a student-generated question repository, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020105 (2014).
- [44] H. A. McQueen, C. Shields, D. J. Finnegan, J. Higham, and M. W. Simmen, Peerwise provides significant academic benefits to biological science students across diverse

- learning tasks, but with minimal instructor intervention, *Biochem. Mol. Biol. Educ.* **42**, 371 (2014).
- [45] M. Milner-Bolotin, Learning physics teaching through collaborative design of conceptual multiple-choice questions, *Can. J. Action Res.* **16**, 22 (2015).
- [46] J. Guiller, A. Durndell, and A. Ross, Peer interaction and critical thinking: Face-to-face or online discussion?, *Learn. Instr.* **18**, 187 (2008).
- [47] L. M. PytlikZillig, C. A. Horn, R. Bruning, S. Bell, X. Liu, K. O. Siwatu, M. C. Bodvarsson, D. Kim, and D. Carlson, Face-to-face versus computer-mediated discussion of teaching cases: Impacts on preservice teachers' engagement, critical analyses, and self-efficacy, *Contemp. Educ. Psychol.* **36**, 302 (2011).
- [48] R. Joiner, D. Jones, and J. Doherty, Two studies examining argumentation in asynchronous computer mediated communication, *Int. J. Res. Meth. Educ.* **31**, 243 (2008).
- [49] B. S. Bloom, *Taxonomy of Educational Objectives: Cognitive Domain* (Longman, New York, 1956), Vol. 1.
- [50] T. Lord and S. Baviskar, Moving students from information recitation to information understanding: Exploiting Bloom's Taxonomy in creating science questions, *J. Coll. Sci. Teach.* **37**, 40 (2007).
- [51] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.12.020128> for statistical analysis of the data.
- [52] D. V. Cicchetti, Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology, *Psychol. Assess.* **6**, 284 (1994).
- [53] M. Milner-Bolotin, Rethinking Technology-Enhanced Physics Teacher Education: From Theory to Practice, *Can. J. Sci. Math. Tech. Educ.* **16**, 284 (2016).
- [54] E. Mazur, Farewell, Lecture?, *Science* **323**, 50 (2009).
- [55] S. Kemmis, in *Rethinking Educational Practice Through Reflexive Inquiry*, edited by N. Mockler and J. Sachs (Springer, New York, 2011), p. 11.
- [56] P. Hunt and R. McGovern, Towards the reflective practitioner, in *Using Group-Based Learning in Higher Education*, edited by L. Thorly and R. Gregory (Kogan Page Limited, London, 1994), Vol. 1.