Investigating learning assistants' use of questioning in online courses about introductory physics

Jianlan Wang,[1,*] Yuanhua Wang,[2] Kyle Wipfli,[3] Beth Thacker,[3] and Stephanie Hart[4]

[1]College of Education, Texas Tech University, Lubbock, Texas 79409, USA
[2]School of education, College of applied human sciences, West Virginia University, Morgantown,, West Virginia 26506, USA
[3]Physics & Astronomy department,, Lubbock, Texas 79409, USA
[4]Terry B. Rogers College of Education and Social Sciences, West Texas A&M University, Canyon, Texas 79016, USA

(Received 2 November 2022; accepted 23 January 2023; published 23 February 2023)

Learning assistants (LAs) play an important role in inquiry-oriented physics instruction in large classrooms. LAs increase the teacher-student ratio and provide timely support to student exploration. Questioning is believed to be an advantageous strategy for LAs to scaffold student learning and maintain students as the agent of knowledge construction. Unfortunately, there are few instruments for the measurement and preparation of LAs' competency of questioning. It remains obscure how LAs use questioning in their interaction with students and how their questioning contributes to students' conceptual understanding. In this study, we developed two instruments that included a coding scheme to assess LAs' practice of questioning from class videos and written questions to assess LAs' knowledge of analyzing various situations that they may encounter while teaching introductory physics. We used the two instruments to measure performed and narrated pedagogical content knowledge of questioning (PCK-Q) of four LAs in two inquiry-based physics courses taught online. We examined the validity and reliability of both instruments, gauged LA-student interaction in online settings, delineated the LAs' PCK-Q, and suggested how the LAs' questioning contributed to students' physics learning. We also discussed the use of the two instruments for the purposes of both LA assessment and preparation.

DOI: 10.1103/PhysRevPhysEducRes.19.010113

I. INTRODUCTION

Introductory physics like classical mechanics and electromagnetism is a gateway to science, technology, engineering, and mathematics (STEM) majors [1]. There have been increasing efforts of transforming introductory physics courses from lecturing and factual recitation to student-centered knowledge construction known as inquiry teaching [2,3], such as ISLE [4] and Modeling Instruction [5]. Inquiry teaching requires that teachers facilitate student-paced exploration with support that meets students' needs in different contexts [6,7], which is challenging given the large enrollment in introductory physics courses. Undergraduate learning assistants (LAs) are pivotal to course reformation since they increase the teacher-student ratio so timely support becomes more practical. Meanwhile, LAs enable near-peer teaching where students seek help from more-knowledgeable peers [8]. Research yields empirical evidence showing that LA programs promote students' conceptual understanding [9–11] and positive attitudinal shift toward physics [12,13]. However, the mechanism remains unclear how the presence of LAs or LA-student interaction entails positive learning outcomes.

It is argued that LAs improve student learning [6,14], but there have been few studies that directly examine how LAs interact with students in inquiry-oriented settings or how LA-student interaction contributes to student conceptual learning. Research regarding LAs mainly follows the pre-post design to compare students' learning outcomes with and without LAs [15,16]. The details of LA-student interaction remain a black box. For example, do LAs scaffold student learning with strategies like questioning as expected, or do they address students' difficulties by directly imparting what students need? What pedagogical competencies do LAs need to provide effective guidance? What should LAs do when they are unable to address students' needs with inquiry teaching strategies? How to efficiently prepare LAs with inquiry-teaching practices given their limited access to educational theories or field practicum prior to a LA program?

Another deficiency of knowledge lies in the preparation of LAs' pedagogical competencies. LA assessment has

[*]jianlan.wang@ttu.edu

been limited to LAs' subject matter knowledge [17], students' evaluation of LA performance [18], and LAs' self-perception of their pedagogical competencies [19]. There has been little direct assessment of LAs' pedagogical competencies in physics instruction. Despite interventions like pedagogical preparation, empirical evidence is lacking about whether and how well LAs transfer pedagogical instruction that they have received to their teaching practices. It is questionable whether LAs support student learning through inquiry teaching considering that successful implementation of inquiry teaching is rare even in classes taught by experienced teachers [2]. LA programs may face the same challenge of theory-practice disconnection like other science teacher preparation programs. Thus, more evidence is needed to ascribe the improvement in students' achievement to inquiry-oriented strategies conducted by LAs.

In this study, we took an initial step to design two instruments for the assessment and preparation of LAs' pedagogy of questioning, a coding scheme (Instrument 1) and written questions (Instrument 2). We focused on questioning because it is a promising approach to diagnose students' ideas, prompt productive discourses, and promote student thinking [20,21]. Questioning is more advantageous than guiding statements because it maintains students' sense of agency for problem solving [22]. We concentrated on one pedagogy also because the time of pedagogical instruction is limited for LAs to engage in depth with multiple pedagogies. We used the framework of pedagogical content knowledge to describe LAs' competencies of questioning [23,24], which we refer to as PCK-Q in this study. Our research questions are as follows:

Q1. Are the coding scheme and written questions reliable to measure LAs' PCK-Q?

Q2. Are the coding scheme and written questions valid to measure LAs' PCK-Q?

Q3. How were the LAs prepared to intervene with student learning through questioning as suggested by their performed and narrated PCK-Q?

II. THEORETICAL FRAMEWORK

A. Learning assistants and practice-based teacher education

The LA program originated from the University of Colorado at Boulder [6]. Since then, the program has been expanded to undergraduate STEM education in institutions around the nation. LAs are undergraduate students with relatively strong content knowledge who are hired to assist class instruction in the course that they have taken [14]. Unlike graduate teaching assistants who participate in course design and grading, LAs serve to increase the teacher-student ratio during class instruction for students to receive timely support. Empirical studies have supported the effectiveness of LA programs in promoting students' conceptual understanding measured by conceptual inventories [14,25,26] and positive attitudes toward physics such as a stronger physics identity [12,27]. Similar positive results were found in other disciplines, such as biology [28,29], chemistry [30,31], and integrated STEM [32].

Most of these studies follow an experimental or quasi-experimental design by comparing pre-post differences in student learning between groups with and without LAs. The consistent findings consolidate that the presence of LAs has positive impacts on student learning, but the mechanism remains unclear. Scholars have ascribed the success of LA programs to that LAs make class instruction more interactive, engaging, and personal [29,33]. Knight and colleagues [34] found that students who interacted regularly with LAs asked more questions for information or feedback and spent more time in discussion. However, students articulating their thoughts or participating in discussion is insufficient to guarantee productive learning outcomes such as student investigation proceeding along an expected direction. There is inadequate knowledge about how LAs intervene with student exploration, especially when it goes off on a tangent. Attention needs to be cast on content knowledge in LA-student interaction to examine whether effective inquiry teaching from LAs takes place.

Such knowledge is critical to pedagogical instruction in LA preparation. LA preparation is composed of four aspects, instructional methods courses [9], clinical review [18,26], LA learning community [35], and weekly preparation sessions with course instructors [14]. LAs are expected to transfer educational literature addressed in methods courses [35] into informed pedagogical strategies for practice. Unfortunately, researchers have observed the theory-practice gap as in other science teacher preparation programs [36]. For example, it was found that direct instruction from LAs was most common in their interactions with students even though questioning was advocated [37]. There requires a more efficient approach for LAs to assimilate educational theories and integrate them in practice. In this study, we adopted the model of practice-based teacher education [38] to pilot an approach.

According to this model, teaching practice rather than theoretical instruction lies at the core of teacher preparation because educational theories do not impact student learning directly. This model suggests a cycle of identification, modeling, implementation, and reflection of core practices in a discipline [38]. Our instruments were designed to contribute to implementation and reflection of the practice of questioning. Given that LAs are not education majors or prospective teachers, they may not be fully committed to theories of physics teaching after they enter a LA program. Practices in specific contexts may be better learning objectives because they are more transferable to LAs' own competencies. Our instruments, including video coding and written questions, are to engage LAs in the cycle of practice instruction. For example, video coding could be used for LAs to reflect on their implementation of

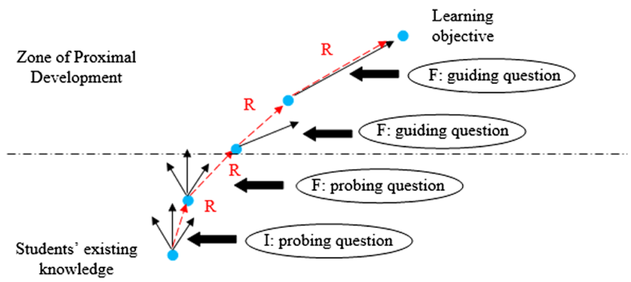

FIG. 1. The functions of probing and guiding questions in a teacher's effort of scaffolding student learning with a questioning chain.

questioning. Written questions could engage LAs in the discussion and reflection of questioning in scenarios of teaching carefully tailored for specific purposes, which compensate for LAs' lack of teaching experiences and limited exposure to authentic teaching practices before they serve in classes. Details will be discussed in the following sections.

B. Questioning in inquiry teaching

Despite its various formats, inquiry teaching shares a key feature that students expand their knowledge scheme to the zone of proximal development (Fig. 1) with the support of more knowledgeable others like LAs [8]. Questioning is an important means of support that enables teachers to solicit students' ideas, prompt productive discourses, and promote student thinking [20,21]. Questioning in traditional lecturing follows the pattern of I (a teacher initiates a question)-R (students respond to that question)-E (the teacher evaluates students' responses). In inquiry teaching, teacher questioning follows a chain of I-R-F (the teacher asks a follow-up question)-R-F [20]. The differences are twofold: (i) questions in inquiry teaching serve to hold students accountable to their knowledge construction in addition to evaluating students' understanding; (ii) questions in inquiry teaching are interconnected and lead students step by step to the learning objective. Questions in an I-R-F-R-F chain fall into two categories based on their functions in leveraging student learning. One category is probing questions that prompt students to articulate their thoughts for the purpose of collecting information about students' understanding. Probing questions are typically open ended without restrictions on students' answers. The other category is guiding questions that embed teacher support by referring to specific learning sources. Guiding questions are narrow-ended questions with restrictions on students' answers. As shown in Fig. 1, black solid arrows are the paths teachers expect students to follow with their questions and red dashed arrows are the actual paths of student learning. Because of a lack of restraints, probing questions are less likely to direct students toward a learning objective than guiding questions, especially when they venture off on a tangent.

Existing studies tend to favor open-ended questions as being more advantageous for inquiry teaching. For example, open-ended questions are found to be more emblematic of inquiry teaching and they entail more active and extended answers from students [39]. However, overemphasizing open-ended questions may underestimate the importance of teacher intervention to student learning and mislead LAs to a hands-off mindset of inquiry teaching where students are expected to develop conceptual understanding merely from articulating their thoughts. Students' conceptual difficulties are ignored or addressed through lecturing. Kawalkar and Vijapurkar [22] specified the progression of questioning from exploring prerequisites, to generating ideas and explanations, probing further, refining conceptions and explanations, guiding the entire class toward scientific conceptions, and eventually achieving the intended teaching goal. Similarly, Smith and Hackling [40] specified a sequence of questions for inquiry teaching as a teacher using open-ideas questions to initiate class discussion, following up with open-description questions to elicit student observation, then using open-explanation questions to prompt student explanation, and ending with closed-ended questions to help students draw a valid conclusion. Both models suggest a sequence of questions from being divergent (i.e., probing questions) to convergent (i.e., guiding questions) and the significance of directing student learning toward a designated direction based on the probed information of students' existing knowledge.

C. Pedagogical content knowledge of questioning

PCK is pedagogical knowledge applied to a specific learning context [23,41]. In this study, PCK-Q describes LAs' skill of asking effective guiding questions when students need LAs' support. We did not deny the importance of probing questions but believed that guiding questions require more sophisticated skills. We adopted the existing framework of PCK [23,41] composed of five components, which are described in the first two columns in Table I. We applied the first four components of O, C, S, and I in the context of questioning, which is defined in the third column of Table I. The four components are associated with four competencies that LAs need to ask effective guiding questions. After gauging student understanding with probing questions, LAs need to identify students' strengths and difficulties in their understanding (S) based on their own physics content knowledge (C). They need to determine whether they would use questions to respond to students (O). Finally, they would select effective responses (I) that are more likely to bridge the gap between students' existing knowledge and learning objectives. We did not incorporate the component of assessment because it is about a teacher's knowledge of different methods of assessment. LAs typically are not involved in curriculum design or the determination of approaches to assessment. Besides, we focused exclusively

TABLE I. Existing framework of PCK [23,41] and its application in the analysis of questioning.

Component	Definition in the existing framework	Application in PCK-Q for this study
Orientation (O)	Teachers' beliefs about the purposes and goals for teaching science at different grade levels	LAs' awareness of the importance of questioning in inquiry-oriented physics teaching, which is evidenced by their preference for using questions during their interaction with students.
Knowledge of curriculum (C)	Teachers' knowledge about curriculum materials available for teaching particular subject matter as well as about both the horizontal and vertical curricula for a subject	LAs' knowledge of physics content associated with a lesson objective.
Knowledge of students (S)	Teachers' knowledge about what students know about a topic and areas of likely difficulty	LAs' knowledge of students' strengths and difficulties in their understanding of physics concept(s).
Knowledge of instructional strategies (I)	Teachers' knowledge of general approaches to instruction that are consistent with the goals of science teaching and specific strategies that apply to teaching particular topics within a domain of science	LAs' knowledge of appropriate responses to students (including questions) that could effectively help students make a progress toward a learning objective.
Knowledge of assessment	Teachers' knowledge of the dimensions of science learning important to assess and knowledge of the methods by which that learning can be assessed	Not applicable

on questioning as one method of formative assessment. It was beyond the scope of this study to assess LAs' competencies of other assessment strategies.

LAs' PCK-Q can be derived from their performance (performed knowledge) or articulation (narrated knowledge). Thus far, teacher questioning is primarily analyzed from their use of questions (i.e., performance) in practice through discourse analysis [42]. These studies yield rich information about teacher questioning in their interaction with students but are typically conducted with a small sample size, such as one lesson from each of the five teachers [43]. Since questioning is contingent to the context of a discourse [44], it may be biased to derive LAs' PCK-Q from limited scenarios. For example, LAs who are more knowledgeable about kinematics than Newton's laws may ask high-quality questions more frequently while teaching kinematics. Thus, the LAs' questions used in both topics should be considered while analyzing their PCK-Q. Besides, existing methods analyzing teachers' questioning mainly concentrate on the features or functions of discrete questions rather than the connection between questions and the role of a questioning chain in scaffolding student learning. In this study, we designed instructor-friendly instruments for quantitative assessment of LAs' performed and narrated PCK-Q from multiple scenarios of teaching. With the instruments, we described how LAs' questioning chains leveraged student learning in an inquiry-oriented setting.

III. METHODS

A. Participants and context

The participants of this study were four LAs in two consecutive introductory physics courses about classical mechanics and electromagnetism, respectively, which were both taught online due to the COVID pandemic. Both courses were calculus-based and inquiry-oriented where students worked in groups to construct physics knowledge. The instructor and LAs met with students online via zoom where students worked in breakout rooms. Hands-on physics experiments were replaced with virtual labs. The LAs moved around different breakout rooms to facilitate student exploration. Toward the end of a lab, all the students returned to the main room on zoom to reflect on their exploration and summarize the target knowledge. The background information of the participating LAs is summarized in Table II. Like many other LA programs, pedagogical and content preparation for the LAs was separated. Before this project, the four LAs had taken an instructional methods course regarding pedagogy for inquiry teaching such as scaffolding and questioning. In both courses, the LAs met with their instructor online during weekly preparation sessions to preview physics

TABLE II. Background information of the participating LAs.

Course	Participant	Gender	Race	LA experience prior to this course
Fall 2020, classical mechanics	LA-1	Female	Hispanic	0 yr
	LA-2	Female	Black	1.5 yr
	LA-3	Male	Hispanic	1.5 yr
Spring 2021, electromagnetism	LA-3	Male	Hispanic	2 yr
	LA-4	Female	Hispanic	1 yr

content knowledge for the coming week. The data collection lasted for 4 weeks in the fall semester and 10 weeks in the following spring semester, during which the participating LAs were requested to videotape their interaction with students once per week for a length of 30–60 min using the zoom camera that followed them. Meanwhile, they were requested to answer 1–2 PCK-Q questions each week. The videos and written responses were analyzed by the two instruments, respectively, to assess the LAs' performed and narrated PCK-Q following the framework of O-C-S-I (Table I).

B. Performed PCK-Q from class videos measured by Instrument 1

Performed PCK-Q of LAs could not be directly observed but inferred from their interaction with students, especially the ways LAs intervene with student learning. LAs with strong performed PCK-Q should be more likely to contribute to student learning through questioning. The coding scheme for LAs' performed PCK-Q was composed of two parts, codes of LAs' utterances (Table III) and LA intervention indicated by code patterns (Table IV). We first segregated a video into episodes. An episode was a video segment where a LA interacted with a fixed group of students. An episode ended when the LA rotated to another group. Each episode might contain multiple vignettes of LA-student interaction regarding different learning tasks. For example, a LA might help a group of students with two questions in an episode, which were calculating the acceleration of a block on a frictionless slope and drawing the free-body diagram of a box sliding on the ground with friction. Then there were two vignettes in this episode. Vignette shift was marked by sentences that indicated the change of learning objectives, such as "let us look at the next question" and "what is your plan for the next step?". Within each vignette, we focused exclusively on a LA's utterance in the I-(R)-F-(R)-F model and coded them based on Table III. The column of "Code" contained the two-letter codes (e.g., pq and gq) used for video coding, which were represented by longer descriptions in "Text" (e.g., ProQ and GuiQ) for readers to easily comprehend them.

TABLE III. Codes of a LA's utterance during their interaction with students (Instrument 1, Part 1).

	Code	Text	Description	Example
Questioning 1: Checking question	cq	CheQ	LAs ask questions unrelated to specific physics content for temperature checking or to collect superficial information about students' learning progress. The answer is normally simple words or short statements that do not require deep thinking of physics.	(1) Are you with me? (2) Does this make sense to you? (3) Do you agree that it is an elastic collision?
Questioning 2: Probing question	pq	ProQ	LAs ask content-related questions to collect information about students' understanding by having them clarify their thoughts or elaborate on their reasoning. Probing questions do not convey information to students regarding physics content as guidance.	(1) What do you mean that the work is the same? (2) Tell me what you have done so far with this lab. (3) Can you tell me why you place the motion sensor here?
Questioning 3: Guiding question	gq	GuiQ	LAs ask content-related questions with the purpose of guiding students in a certain direction when students demonstrate conceptual difficulties (e.g., misconceptions) that probably need LA support. Guiding questions would refer to a specific source of physics content as LAs' intervention with student learning.	(1) Student: The force from the truck to the car is larger in a collision. LA: What is the Newton's 3rd law pair here? (2) Student: Is my a-t graph correct? LA: Is it aligned with your v-t graph?

(Table continued)

TABLE III. *(Continued)*

	Code	Text	Description	Example
Statement 1: Lecturing	le* or le	Lect* or Lect	LAs impart information about physics content directly to students, such as the answer to a question or a definition or theory in physics. If the information lectured is directly related to the vignette task, it is coded as le*, otherwise it is coded as le.	(1) You are supposed to find out that the velocity is halved because the two carts have the same mass. (2) Use a stiff spring and control the elongation to 2 cm, that's how you get a constant force.
Statement 2: Transition	tr	Tran	LAs use connecting statements so the communication with students would proceed more smoothly or naturally, such as acknowledging students' ideas and paraphrasing students' explanation. There is little new information about physics content imparted to students from LAs.	(1) So you mean that the ball would spin faster when you pull the string down. Sounds good, make sure the force you exert is constant. (2) OK, we have done with the x direction, like you said, it's a constant-v motion, let's look at y direction now.
Statement 3: Error	er	Erro	LAs demonstrate errors in their statement about the content or technique of physics that would potentially intervene with student learning in a negative way.	(1) Content: The gravitational force down balances out the centripetal force up. 2) Technique: Just ignore the lab data, you know the answer already, move on to the next activity.
Questioning or statement: Direction	di	Dire	LAs give direction about the norms, expectations, or procedure of an instructional activity in the form of either questioning or statement.	(1) Now we are on Question 2, later we will move on to Question 3. (2) You need to draw the free-body diagram of both objects in the collision.

Based on the patterns of the codes in a vignette, we defined nine vignette levels associated with the LAs' potential intervention in student learning (Table IV). Take a vignette of Newton's third law in collision, for example, a LA first used a probing question (ProQ) to realize students' misconception that a truck exerted a larger force than a car did in their collision. If the LA chose to correct students' misconception by directly lecturing that *"the forces are equal in magnitude between the truck and the car because of Newton's third law*." This lecturing would be coded as "Lect*" because it was directly related to the learning objective. Then the vignette had a pattern of "ProQ-Lect*," which corresponded to the level of "Db" indicating teacher-dominant lecturing. If the LA asked guiding or probing questions while they started lecturing, such as *"what is the Newton's third law pair in this case?"* (ProQ), the pattern became "ProQ-Lect*-ProQ." Then the vignette level was "Da," which indicated interactive lecturing. Instead of lecturing, the LA could guide students with a sequence of questions, like *"Could you recall Newton's third law?"*(ProQ), *"If we call the force from the truck to the car the action force, what is the reaction force?"* (GuiQ), and *"what does Newton's third law say about action and reaction forces?"* (GuiQ). The vignette had a level of "Qa" because the pattern was "ProQ-ProQ-GuiQ-GuiQ." If students confused action-reaction forces with balanced forces, the LA could keep prompting students with a question like *"Do action-reaction forces act upon the same object or different objects?"* (GuiQ), i.e., "ProQ-ProQ-GuiQ-GuiQ-GuiQ." The vignette would remain "Qa." Alternatively, the LA could lecture the prerequisite knowledge that *"Balanced forces act upon the same object and action-reaction forces act upon two different objects"* (Lect). This lecturing was coded as "le" because it was not directly related to the learning objective. Then the vignette had a level of "Qb" because of the pattern "ProQ-ProQ-GuiQ-GuiQ-Lect," which indicated that students constructed the target knowledge by themselves with some delivery of prerequisite knowledge from the LA.

The levels from Qa to Ne in Table IV describe three categories of LA intervention to students' knowledge construction. The category of Q (including Qa, Qb, and Qe) is marked by the presence of "GuiQ" without "Lect*" in the utterance codes of a vignette, which describes a LA's intervention of indirect support through questioning. The category of D (including Da, Db, and De) is marked by the presence of "Lect*" regardless of other codes, which describes a LA's intervention of direct support through lecturing the knowledge regarding the learning objective of a vignette. The category of N (including Na, Nb, and Ne)

TABLE IV. Levels of a LA's potential intervention to student learning (Instrument 1, Part 2).

	Level	Description	Salient pattern	Pattern explanation
Questioning oriented intervention (Q)	Qa	A LA uses guiding questions to support students to accomplish the learning task of a vignette. The LA may use probing questions before or as transitions after guiding questions but uses little lecturing.	gq	Presence of GuiQ without Lect* or Lect
	Qb	A LA uses guiding questions to support students to accomplish the learning task of a vignette. The LA uses some lecturing. However, the content knowledge lectured is not directly related to the vignette task.	gq-le	Presence of GuiQ with some Lect but no Lect*
	Qe	Qa and Qb levels with identifiable errors from a LA	gq-er	Presence of GuiQ and Erro without Lect*
Direct-instruction oriented intervention (D)	Da	A LA accomplishes the learning task of a vignette by directly imparting to students the necessary content knowledge. The LA may try guiding questions but eventually resort to lecturing to accomplish the vignette task. During lecturing, the LA interacts with students with probing or guiding questions.	le*-pq/gq	Presence of Lect* and ProQ or GuiQ
	Db	A LA accomplishes the learning task of a vignette by directly imparting to students the necessary content knowledge. The LA may use probing questions before or checking questions after lecturing. The LA barely uses any guiding or probing questions during lecturing but presents content knowledge with little attention to students' understanding.	le*-(cq)	Presence of Lect* with probably some CheQ but little ProQ or GuiQ
	De	Da and Db levels with identifiable errors from a LA.	le*-er	Presence of Lect* and Error
Vignette task not accomplished (N)	Na	A LA intervenes with student learning by providing instruction or guidance about an important step in the task of a vignette, but the task is incomplete. One example is a LA imparting to students prerequisite knowledge so they can continue their exploration with the vignette task.	le-tr/di	Presence of Lect without GuiQ or Lect*
	Nb	A LA does not intervene with students' learning with either guiding questions or lecturing when students work on the vignette task. One example is a LA checking students' work using probing or checking questions.	tr/di	No GuiQ, Lect*, or Lect
	Ne	Na and Nb levels with identifiable errors from a LA.	(le)-er	Presence of Erro without Lect* or GuiQ

describes the situation where a LA entails little intervention or the intervention is not directly related to the vignette task. Another criterion to separate the three categories is the agent of knowledge construction as being students themselves (Q and N) or LAs (D). A LA's performed PCK-Q was derived from all their vignettes from the videos collected during the research period. This was to diminish the bias of deriving PCK-Q from one vignette due to the context-contingent nature of questioning. The corresponding justifications of these equations represented our theoretical deduction of LAs' performance when they possessed a certain aspect of PCK-Q. It is important to emphasize that the four components of O-C-S-I could not be directly observed from videos and they worked in an integrated manner to shape a LA's performance. The four questions were our initial attempt to separate O-C-S-I and would be subject to adjustment in reference to practical data. More details will be discussed later in the examination of instrument validity.

Orientation (O): 1-(Da% + Db% + De%)

Justification: The total percentage of D levels represents the frequency of LA interventions through direct instruction. A questioning-oriented LA may use direct instruction but not frequently.

Knowledge of curriculum (C): 1-(Qe% + De% + Ne%)

Justification: The levels of Qe, De, and Ne are vignettes where a LA demonstrates errors in physics. A LA with strong physics content knowledge may still demonstrate misconceptions in physics but not frequently.

Knowledge of students (S): Qa% + Qb% + Da% + Na%

Justification: Qa, Qb, Da, and Na (Table IV) are interactive vignettes where a LA gauges students' understanding. A LA who gauges students' understanding more frequently is more likely to develop stronger knowledge of students' understanding.

Knowledge of instructional strategy (I): Qa% + Qb% + Da% + Db%

Justification: A LA with a stronger knowledge of instructional strategies is more likely to successfully support students to accomplish a vignette task (i.e., Qa, Qb, Da, and Db).

We started developing the coding scheme (Instrument 1) in the year of 2019. Since questioning was the focus of our project, we first categorized LAs' utterances into questioning and nonquestioning. Based on the function of a question as information collection or intervention (Fig. 1), we subcategorized questioning into probing and guiding questions. We labeled the category of nonquestioning as "statement" and started with three subcategories based on whether a statement contained correct physics content (i.e., lecturing), incorrect physics content (i.e., error), and no physics content (i.e., direction). Then we applied the prototype of the coding scheme with the videos collected from another inquiry-oriented physics course given in a face-to-face setting. We collected the videos with a SWIVL set that included a microphone carried by a LA for audio recording and an iPad on a robotic set following the microphone for video recording. We collected multiple videos from each voluntary LA in a week and randomly selected one video for coding. During exploratory coding, we transcribed video segments, coded them separately, and met to reflect on the coding. In this process, we spotted another type of questioning not related to specific physics content. It was close to probing questions but functioned mainly for temperature checking rather than information collection, which was labeled as "checking questions." We also separated the code of transition from lecturing because it was a LA paraphrasing students' answers even though physics content was involved. Different from lecturing, transition entailed limited intervention to student learning. Since each vignette contained a specific learning objective, we divided lecturing into lecturing target knowledge directly related to the learning objective (i.e., "Lect*") and lecturing prerequisite knowledge not directly related to the learning objective (i.e., "Lect"). This division was necessary to determine whether it was the LA or students who achieved the learning objective. After finalizing the coding scheme, we did not transcribe videos but directly coded videos on a spreadsheet while watching them. In addition, the first author who designed the coding scheme trained three graduate students with it and led them to code a total of 114 videos of science teaching from 60 elementary preservice science teachers. The findings were reported in another manuscript. Before this current study, the coding scheme had been tested and finalized in both contexts of college physics instruction and elementary science teaching.

C. Narrated PCK-Q from written responses measured by Instrument 2

LAs' narrated PCK-Q was measured by Instrument 2 composed of written questions and a grading rubric (Table V). The written questions described various scenarios of physics teaching regarding key concepts. Each question contains the background information about a conversation between a LA and students. The conversation proceeds to an extent that there is enough information for respondents to infer students' conceptual difficulties associated with a physics concept. The exemplary written question below is about force decomposition in Newton's second law. The student in the conversation does not understand the purpose of force decomposition but blindly memorizes that x and y axes should be parallel and perpendicular to the ground surface. Thus, the student decomposes the normal force but not the acceleration that is not on any axes. Similarly, the LA could impart the right answer to students or respond with a guiding question such as *"If you draw x and y axes in this way, do you need to decompose the acceleration as well?"* By answering this question, students had a chance to make a progress in their learning in terms of realizing their misconceptions.

Example written question.—Context: Students have learned about free-body diagrams and Newton's second law. A group of students is working on a question about a block sliding down a frictionless inclined plane. They are required to draw a force diagram to explain why the block slides down. You approach that group, notice that they have drawn a gravitational force straight downward and a normal force upward and perpendicular to the slope surface. Then you have a conversation with one of them:

You: So why does the block slide down?

Student: It slides down because we have a x component of the normal force acting on it, so that's why it slides down.

You: What do you mean by x component of the normal force?

Student: Force is a vector, so when there is a force, there should be x and y components.

You: So what is the x component of the normal force?

Student: With respect to the table surface, y is perpendicular, and x is parallel.

You: OK, then which component or which force determines if this block moves or not?

Student: The x component of the normal force. The y component cancels out with gravity.

a. Is this student's answer correct? Why?
b. What can you conclude from the information provided about the students' physics content knowledge, both

TABLE V. Rubric for narrated PCK-Q from LAs' written responses (Instrument 2).

Component	Score	Description
Orientation (O)	2	The LA asks content-specific question(s) that refer to a specific physics concept.
	1	The LA asks content-free question(s) that do not refer to any specific physics concepts.
	0	The LA does not respond with a question.
Knowledge of curriculum (C)	2	There are clear and accurate statements of physics in responses, which suggests that the LA holds strong physics content knowledge about the learning objective.
	1	There are no clear statements of physics in responses, which is insufficient to infer whether the LA holds strong physics content knowledge about the learning objective.
	0	There are erroneous statements of physics in responses, or the statements are unrelated to the learning objective, which suggests that the LA holds misconceptions about the learning objective.
Knowledge of students' understanding (S)	2	The LA accurately identifies both the strengths and difficulties in students' understanding of physics concept(s) associated with the learning objective from the given information.
	1	The LA accurately identifies either strengths or difficulties, but not the other, in students' understanding of physics concept(s) associated with the learning objective from the given information.
	0	The LA fails to identify either the strengths or difficulties in students' understanding of physics concept(s) associated with the learning objective from the given information.
Knowledge of instructional strategies (I)	2	The LA's response(s) to students are aligned with the learning objective and could potentially scaffold students to make thorough progress in their learning in terms of realizing their misconception and the direction for correction.
	1	The LA's response(s) to students are aligned with the learning objective and could potentially scaffold students to make initial progress in their learning in terms of realizing their misconception.
	0	The LA's response(s) to students are unaligned or have no potential to scaffold students to make initial progress in their learning.

the strengths (i.e., what they know) and difficulties (i.e., what they do not know)?

c. How would you respond to the student? Please use a direct quote of what you would say. What is (are) your purpose(s) behind that response?

We used the rubric in Table V to assess a LA's narrated O, C, S, and I from their answers. From the answer to subquestion c, we assigned a score to O based on whether the LA intended to use a content-specific question (2 points), a content-free question (1 point), or no question (0 points) in response to students. From the answer to subquestions a and b, we assigned a score to C based on whether there is clear evidence (2 points), no evidence (1 point), or clear counterevidence (0 points) about the LA's physics content knowledge. From the answer to subquestion b, we assigned a score to S based on whether the LA could pinpoint both (2 points), either (1 point), or neither (0 points) strengths and difficulties in students' understanding. From the answer to subquestion c, we assigned a score to I based on whether the LA's response could significantly (2 points), partially (1 point), or unlikely (0 points) help students bridge the gap in their understanding. A LA's answer to a written question had a maximum score of 8 points. The average score among multiple questions that a LA answers represents their narrated PCK-Q.

This rubric was first developed with two dichotomous levels of evident (1 point) and not evident (0 points) for each component of PCK-Q. For example, evident C and not-evident C described strong and problematic physics content knowledge of a LA, respectively. During exploratory coding in the academic year of 2019, we found that an intermediate level was needed for the situation where there was no evidence supporting either strong physics content knowledge or misconception of a LA. This situation applied to the other three components of C, S, and I. Thus, we converted dichotomous scores of 1 and 0 into three levels of 2, 1, and 0 points. Prior to this study, we had finished revising and finalizing the rubric for written questions. Similarly, the first author followed the same procedure to develop written questions regarding elementary science teaching with the support of experienced science teachers and science teacher educators. He trained three graduate students with Instrument 2 and led them to apply the same rubric to analyze the responses to 20 written questions from 108 elementary preservice science teachers. The findings were reported in a separate manuscript.

D. Research design

Each semester, the participating LAs voluntarily answered the PCK-Q questions from us and shared instructional videos with us. There were five PCK-Q questions

<table>
<tr><th>Coder~Reviewer</th><th>Lesson objective</th><th>Time frame</th><th>Episode 1</th><th>Time frame</th><th>Episode 2</th></tr>
<tr><td>BW~TP</td><td>Being able to apply Newton's second law tp calculate the acceleration of an ojbect on a slope and the forces on an object hung by two strings.</td><td>00.00.00 - 00.11.51</td><td>The LA helps a group of students build the free body diagram of a block on a slope

pq-tr-pq-tr-le(normal force is perpendicular to the ramp)-tr-cq-gq(what is the relationship between net force and acceleration?)-gq(what is the missing piece?)-le/le*(the component of gravity along the ramp makes it move)|
cq(00.02.35)-pq-tr-gq(is the force diagram consistent with your force analysis over here?)-tr-gq(are all forces now on the two axes?)-cq-pq-pq-le*(00.05.19-00.05.57)-le(normal force is also called contact force)-le*(normal force cancels the y component of gravity) -cq-di|
le*(00.07.45-00.09.30)-cq-cq|
cq(00.09.53)-le(00.10.09-00.10.51)-cq(00.10.58)-le*/le (calculate x and y components of gravity)-tr</td><td>00.11.51 - 00.18.10</td><td>The LA helps a group of students with the force diagram of a box hung by two strings with different angles to the ceiling

pq-pq-pq-tr-le*(00.12.40-00.13.54)-le*(no normal force, the object is not touching any surfaces)|
cq(00.14.19)-pq-pq-cq|
(student discussion)-pq(00.15.35)-pq-gq(is the box moving?)/pq-pq-tr-gq(what does it mean for net force?)-pq-tr-di-le(vector decomposition)-cq-pq-pq-cq-tr-le(trignometry for two different triangles)-cq-le(the box doesn't move up and down)-cq-cq/gq(are we using 10 for gravity?)-cq-pq-tr-pq-di</td></tr>
<tr><td></td><td></td><td></td><td>Qb/Da|Da|Db|Db/Na</td><td></td><td>Db|Nb|Qb</td></tr>
</table>

FIG. 2. Example of video coding in black text and video reviewing in orange text.

regarding classical mechanics in the fall semester. LA-1 answered four of the five questions whereas LA-2 and LA-3 answered all the five questions. LA-1 and LA-3 shared three videos and LA-2 shared four videos with us. There were 12 questions regarding electromagnetism in the following spring semester. LA-3 answered all of them and LA-4 answered 9 out of the 12 questions. LA-3 shared ten videos and LA-4 shared eight videos with us. Ideally, videos should be randomly collected from a LA. Due to online instruction, it was the LAs rather than us who captured the videos. Thus, it was possible that the videos shared with us were carefully selected by a LA to represent their best teaching practice. Since there were no criteria to exclude any videos, we used all the videos from each LA for both validity and reliability analyses. We also used all the written responses from each LA for reliability analysis. Considering that the written questions may be of different levels of difficulty, we only used the questions that all LAs in a semester answered for validity analysis in order to make comparison among LAs.

Research question 1: Reliability consideration.—To answer the first research question, we started with the interrater reliability with the coding scheme (Instrument 1). As shown in Fig. 2, BW (initials of a pseudonym) was the coder who coded the video based on Tables III and IV. Reviewer TP (initials of a pseudonym) was the reviewer who watched the same video and labeled with a different color any disagreement in the codes and/or vignette levels. In order for the reviewer to better locate the codes in the video, the coder recorded the time stamps of the range of a vignette in the form of "hour. minute. second." In addition, keywords were added to briefly describe the codes of "le*" (i.e., Lect*), "gq" (i.e., GuiQ), and "le" (i.e., Lect) since they weighted heavier in determining a vignette level. For example, if there were different opinions about whether a vignette had a level of "Qb" or "Da," a coder-reviewer pair would use the time stamps and keywords to locate a LA's utterance coded as "le*" (i.e., Lect*) in the video and discuss whether the LA lectured students the target knowledge directly related to the learning objective. This approach could improve the credibility of video analysis because the discussion for the interrater agreement would be based on objective evidence from videos rather than individuals' subjective impressions, feelings, or preferences of a LA's performance in a vignette. Finally, the coder and reviewer met to discuss and reconcile any disagreement but kept the disagreement documented.

Two of us analyzed all the videos about classical mechanics and another two of us analyzed all the videos about electromagnetism. Both unweighted and weighted Cohen's kappa was used when the vignette levels from Qa to Ne were treated as nominal and ordinal data, respectively. As ordinal data, the hierarchy from Qa to Ne represents a sequence of ideal intervention through guiding questions (Q levels), intervention through direct instruction (D levels), to incomplete intervention (N levels). As for the written questions (Instrument 2), we assessed LAs' responses using the rubric in Table V (maximum = 8). The same pairs who co-analyzed the videos coded the questions regarding classical mechanics and electromagnetism, respectively. Two individuals within a pair independently coded all the written questions and compared the results. We treated the component scores of O-C-S-I as nominal data and treated the PCK-Q score as numeric data. Unweighted and weighted Cohen's kappa was used accordingly to gauge the interrater reliability for written questions. We also paid special attention to video segments closer to teaching scenarios described in written questions, from which we examined the parallel-form reliability of the written questions by checking whether a LA's written responses match their performance in those specific contexts.

Research question 2: Validity consideration.—To answer the second research question, we examined the convergent validity of both instruments by comparing the LAs' narrated and performed PCK-Q measured by the two instruments. As illustrated above, the two instruments measured LAs' competencies of questioning following the same framework of PCK but via different approaches. Thus, the same LAs' performed and narrated PCK-Q should somehow be aligned with each other. After reaching an agreement on video coding and the grading of written responses, we compared the patterns of O-C-S-I derived from the two instruments among all the LAs. We also examined the predictive validity by checking whether the LAs' PCK-Q suggested by the instruments matched their teaching experiences. According to the theory of practice-based teacher education [38], novice teachers with more practical experiences are more likely to develop stronger pedagogical competencies. Since the courses were inquiry oriented, we hypothesized that more-experienced LAs were more likely to demonstrate stronger PCK-Q. In the fall semester, LA-1 was a new LA in the course of classical mechanics whereas LA-2 and LA-3 were experienced LAs. In the spring semester, LA-3 was 1-yr more experienced than LA-4. LA-3 who participated in both semesters should have relatively consistent PCK-Q since there was no intervention between the two semesters. We investigated whether the LAs' PCK-Q measured by the two instruments represents those situations.

Research question 3: Interpretation of the LAs' performed and narrated PCK-Q. To answer the third research question, we analyzed the LAs' interaction with students in the online setting as suggested by the two instruments. We referred to the consistency of a LA's competencies of questioning to validate both instruments. On the other hand, we kept in mind the possible theory-practice discrepancy in teacher education [36]. For example, the written questions might measure LAs' perceptions of ideal responses in a snapshot of teaching without any feedback from students, whereas video coding might measure LAs' authentic performance in an entire vignette of teaching where they consistently received students' reactions to their responses. The differences in performed and narrated PCK-Q could reflect on the advantages and disadvantages of both instruments in application.

IV. FINDINGS

A. Q1. Are the coding scheme and written questions reliable to measure LAs' PCK-Q?

To demonstrate our video coding, we used the vignette where LA-1 worked with students on a question about calculating the acceleration of a box on a frictionless slope when the mass of the box and the angle of the slope were given. The students drew the free-body diagram as shown in Fig. 3(a). They singled out the gravitational force in Fig. 3(b), from which they calculated the net force as $(mg\sin\theta + mg\cos\theta)$ and the acceleration as $(g\sin\theta + g\cos\theta)$. The conversation below showed how LA-1 attempted to intervene with the students' learning, from which we could infer that the students might blindly memorize that force decomposition involves sine and cosine components of a force and x and y axes are horizontal and vertical to the ground surface. They were probably unaware that they actually decomposed the gravitational force along the slope and that vector addition was different from scalar addition. In this vignette, the task was to help students correctly apply Newton's second law with a correct free- body diagram. In the beginning, one representative student elaborated on their work as requested by LA-1. LA-1 seemed to misinterpret the student's words of *"going straight down"* (underlined in the text) as gravity being straight down when the student probably meant the component down the ramp. After LA-1 led a discussion about gravity, the student seemed to perceive this discussion as being unrelated to *"the way I did it*." After the student reiterated how *"the cosine would go into the ramp, and the other component going down*," LA-1 requested the student to draw the force diagram again. The student then drew Fig. 3(b) one more time. After LA-1 asked *"why are you drawing the components like that? What are they for?*," the student became frustrated probably because he felt that LA-1 requested the same explanation that he had given but did not respond to it. Eventually, LA-1 resorted to lecturing. Overall, this vignette had a level of Da (Table IV) because the vignette task was accomplished by LA-1 through direct instruction even though she tried guiding questions. In the course of classical mechanics, there were 66 vignettes from the 10 videos of LA-1, LA-2, and LA-3. The Cohen's Kappa between the two reviewers was 0.61 (unweighted, $n = 66$) and 0.86 (weighted, $n = 66$). In the course of electromagnetism, there were 138 vignettes from the 18 videos of LA-3 and LA-4. The Cohen's Kappa between another two reviewers was 0.67 (unweighted, $n = 138$) and 0.83 (weighted, $n = 138$). The consistent pattern suggests that the interrater agreement was substantial (0.61–0.80, [45]) and excellent (0.81–1) when the vignette levels (Table IV) were taken as nominal and ordinal data, respectively

Da vignette: "Dire-ProQ-GuiQ-Tran-Dire-ProQ-Dire-Lect*"

LA-1: Explain to me what you did, the process **Dire**

Student: OK, I have the box sliding down, but I cannot use the component <u>going straight down</u> because there is a component pushing it into the ramp. So I have to use those two components.

LA-1: Do you mean gravity? **ProQ**

Student: What do you mean? I was trying to find the force going straight down?

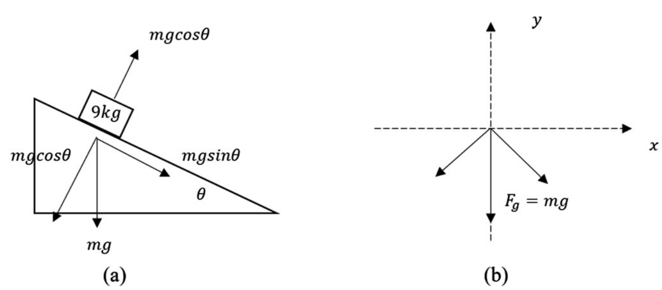

FIG. 3. The reproduction of a student's free-body diagram for a box on a frictionless ramp.

LA-1: What do we know about gravity? Is it completely downward? **GuiQ**

Student: Yes.

LA-1: So that's the force you were talking about here, the force perpendicular to the ground. **Tran**

Student: I guess, but that's not the way I did it. I mean, I was trying to find, I don't know, I thought the cosine [component] would go into the ramp, and the other component going down. It is gravity, but it should have a component, the acceleration in it.

LA-1: Yeah, that's acceleration, let's talk about what you did. **Dire**

[LA-1 told the student to draw the force diagram. After repeating his idea, this student drew Fig. 3(b) again]

Students: I have the force of gravity, and components. No?

LA-1: Why are you drawing the components like that? What are they for? **ProQ**

Student: The box? Let me know if I am right. And it's getting me even more confused. Don't question me. It makes me not understand if I am wrong. I don't know.

LA-1: You need to explain why you are drawing this though. **Dire**

Student[Frustrated]: Because we are supposed to draw the components.

[LA-1 started to demonstrate how to draw the correct force diagram, x-y axes, and the components of gravitational force] **Lect***

To demonstrate our coding of LAs' responses to the written questions, we directly quoted LA-1's answer to the sample written question shown in the previous section. According to Part a, LA-1 demonstrated strong content knowledge about this concept because the analysis of forces was accurate. As shown in Part b, *"this student understands the content at the surface level"* is a vague statement about the strengths in the student's understanding. In addition, it is inaccurate to state that *"only the gravitational force has the x and y components"* because the components of vectors depend on the orientation of x and y axes. In this question, the two axes drawn by the student were not wrong but less convenient because more vectors needed to be decomposed, such as acceleration. Thus, LA-1 might memorize one method of force decomposition as being along the motion without understanding why, which was verified by the postinterview where LA-1 stated that *"you[students] are supposed to put the reference [x and y axes] along the motion, you cannot put it anywhere."* Together, the score of C for LA-1 was 1 because it is unclear whether she possessed robust content knowledge about this concept. The score of S was 0 because the analyses of the student's strengths and difficulties were inaccurate. According to Part c, LA-1 responded to the students with a question requesting them to decompose normal and gravitational forces. The score of O was 2 because this question was narrow ended that referred to two specific forces. However, the score of I was 0 because LA-1's question was unlikely to guide the student to notice the misconception. The dialogue in the question stem suggested that the student already articulated how he would decompose normal and gravitational forces. LA-1's response would probably have the students repeat the illustration already made. In addition, the statement *"they will be a little stumped, because the normal force is completely in the y direction"* indicates that LA-1 probably paid insufficient attention to student reasoning because the normal force would be in the y direction based on her own reference frame but not the student's frame. While being asked in the postinterview *"what if the students draw the components of normal force and gravity in their reference without changing it to be along the slope?"*, LA-1 stated that the follow-up response would be *"I would tell the students that this [students' own reference] is not helpful, you need to draw it along the ramp."* Together, LA-1 received 3 out of 8 ($O = 2$, $C = 1$, $S = 0$, $I = 0$) for this question. In the course of classical mechanics, there were 14 responses from LA-1, LA-2, and LA-3. The unweighted Cohen's Kappa between two reviewers was 1 ($n = 14$) for O, 0.88 for C, 0.54 for S, and 0.65 for I. In the course of electromagnetism, there were 23 responses from LA-3 and LA-4. The unweighted Cohen's Kappa between another two reviewers was 0.92 for O ($n = 23$), 0.85 for C, 0.53 for S, and 0.68 for I. The weighted Cohen's Kappa for narrated PCK-Q was at the excellent level (0.81–1) in both semesters. The data from different pairs of reviewers yielded the same pattern and there was an excellent agreement (0.81–1) in coding O and C, a moderate agreement (0.41–0.60) in S, and a substantial agreement (0.61–0.80) in I. This finding suggests that deriving LAs' knowledge of S and I from their written responses is more subjective and challenging than deriving O and C.

LA-1's responses

a. *This student's answer is not correct. Although their reasoning is almost correct, they are mistaking the normal force with the gravity force. The y component of gravity force cancels out with the normal force, which is both perpendicular to the ramp but in opposite directions; this cancelation is what allows the block not to float nor fall through the ramp. The x component of the gravity force is parallel to the ramp, and is unopposed by other forces (since no other force is acting along that same plane) and is therefore the force that makes this block slide down.*

b. *I think that this student understands the content at the surface level, because they are saying all of the right words, but do not yet understand that only the gravitational force has the x and y component in this case.*

c. *I would say, "well, let's take a look at this diagram. Can you draw for me the x and y components of the normal force, as well as the gravitational force?" When they try to draw the components of normal force, they will be a little stumped, because the normal* *force is completely in the y direction. If they don't know how to proceed from there, I'd say, "well, let's try drawing the components of gravitational force instead," which would lead us to the accurate components, and I'd be able to point out the difference between them, and how they cancel out.*

LA-1 encountered a scenario in teaching similar to the situation described in the sample written question. LA-1 chose the same strategy of requesting students to demonstrate their thinking that had already been articulated, which suggests that LA-1 struggled with inferring students' understanding from their work in the teaching video or the prescribed information in the question stem. LA-1 was aware of using questions, but the questions were ineffective in scaffolding student learning. As shown in the video, the student was still confused after answering LA-1's questions. The student might feel repetitive or interrogated rather than guided by LA-1's questions, which led to his frustration by saying that *"Don't question me. It makes me not understand if I am wrong."* This student's reaction in the video supports our analysis of I as being 0 in LA-1's written response that the question *"well, let's take a look at this diagram. Can you draw for me the x and y components of the normal force, as well as the gravitational force?"* was unlikely to promote student learning. The alignment between LA-1's performance and written responses supports the parallel-form reliability of the written questions.

B. Q2. Are the coding scheme and written questions valid to measure LAs' PCK-Q?

Unfortunately, we could not find one-on-one matches between LAs' all written responses and their performance captured by class videos. Instead, we derived LAs' performed and narrated PCK-Q from multiple videos and written responses, respectively, and examined their alignment. Table VI summarizes the total number of vignettes from each LA (n), the number of each vignette level (i), and the percentage of each vignette level in decimals (i/n). We used the equations introduced in Sec. III B to calculate the LAs' performed O-C-S-I. Each component had a maximum of 1. Performed PCK-Q was the sum of the four components with a maximum of 4. For comparison with written responses, we converted performed PCK-Q into percentage scores in decimals by dividing it by the maximum of 4. Take LA-1 for example, her performed O-C-S-I was calculated as below:

O-Video: $1\text{-}(\mathrm{Da}\% + \mathrm{Db}\% + \mathrm{De}\%) = 1\text{-}(0.24 + 0.16 + 0.12) = 0.48$

C-Video: $1\text{-}(\mathrm{Qe}\% + \mathrm{De}\% + \mathrm{Ne}\%) = 1\text{-}(0.00 + 0.12 + 0.04) = 0.84$

S-Video: $\mathrm{Qa}\% + \mathrm{Qb}\% + \mathrm{Da}\% + \mathrm{Na}\% = 0.00 + 0.00 + 0.24 + 0.16 = 0.40$

I-Video: $\mathrm{Qa}\% + \mathrm{Qb}\% + \mathrm{Da}\% + \mathrm{Db}\% = 0.00 + 0.00 + 0.24 + 0.16 = 0.40$

Performed PCK-Q: $(\mathrm{O} + \mathrm{C} + \mathrm{S} + \mathrm{I})/4 = (0.48 + 0.84 + 0.40 + 0.40)/4 = 0.53$

As for written responses, we averaged the scores of O-C-S-I associated with a LA among the questions that all the LAs answered in a semester. For example, LA-1, LA-2, and LA-3 all answered the first four questions in the fall semester. LA-1's component scores for each question (Table V) are shown below:

Written question 1: $\mathrm{O} = 2$, $\mathrm{C} = 1$, $\mathrm{S} = 0$, $\mathrm{I} = 0$, $\text{PCK-Q} = \mathrm{O} + \mathrm{C} + \mathrm{S} + \mathrm{I} = 3$

Written question 2: $\mathrm{O} = 1$, $\mathrm{C} = 1$, $\mathrm{S} = 1$, $\mathrm{I} = 0$, $\text{PCK-Q} = \mathrm{O} + \mathrm{C} + \mathrm{S} + \mathrm{I} = 3$

Written question 3: $\mathrm{O} = 2$, $\mathrm{C} = 2$, $\mathrm{S} = 1$, $\mathrm{I} = 2$, $\text{PCK-Q} = \mathrm{O} + \mathrm{C} + \mathrm{S} + \mathrm{I} = 7$

Written question 4: $\mathrm{O} = 2$, $\mathrm{C} = 0$, $\mathrm{S} = 1$, $\mathrm{I} = 1$, $\text{PCK-Q} = \mathrm{O} + \mathrm{C} + \mathrm{S} + \mathrm{I} = 4$

Then LA-1 narrated PCK-Q scores became

TABLE VI. Summary of vignette levels from Qa to Ne (Table IV) from the LAs' videos.

Fall 2020 semester, classical mechanics										
	n	Qa	Qb	Qe	Da	Db	De	Na	Nb	Ne
LA-1	25	0	0	0	6	4	3	4	7	1
	%	0.00	0.00	0.00	0.24	0.16	0.12	0.16	0.28	0.04
LA-2	23	0	7	0	4	7	1	1	3	0
	%	0.00	0.31	0.00	0.17	0.31	0.04	0.04	0.13	0.00
LA-3	18	1	4	0	5	2	1	4	1	0
	%	0.06	0.22	0.00	0.27	0.11	0.06	0.22	0.06	0.00

Spring 2021 semester, Electromagnetism										
	n	Qa	Qb	Qe	Da	Db	De	Na	Nb	Ne
LA-3	85	0	11	0	24	16	2	22	10	0
	%	0.00	0.13	0.00	0.27	0.19	0.02	0.26	0.12	0.00
LA-4	56	2	7	0	17	11	1	15	2	1
	%	0.03	0.13	0.00	0.30	0.19	0.02	0.27	0.03	0.02

O-Written: Average O from Q1 to Q4 $= (2+1+2+2)/4 = 1.75$

C-Written: Average C from Q1 to Q4 $= (1+1+2+0)/4 = 1$

S-Written: Average S from Q1 to Q4 $= (0+1+1+1)/4 = 0.75$

I-Written: Average I from Q1 to Q4 $= (0+0+2+1)/4 = 0.75$

Narrated PCK-Q: Average PCK-Q from Q1 to Q4 $= (3+3+7+4)/4 = 4.25$

Similarly, we converted narrated PCK-Q into percentage scores in decimals for the comparison with performed PCK-Q. The four components of O-C-S-I had a maximum of 2 and PCK-Q had a maximum of 8. While being divided by the maximum scores, the narrated PCK-Q of LA-1 became $1.75/2 = 0.88$ for O, $1/2 = 0.50$ for C, $0.75/2 = 0.38$ for S, $0.75/2 = 0.38$ for I, and $4.25/8 = 0.54$ for PCK-Q. We summarized the performed and narrated PCK-Q of all the LAs in Table VII. Three patterns stood out while comparing across all the LAs from the two semesters. First, the scores of O-video from videos were lower than the scores of O-written, which indicated that all the LAs were more questioning-oriented while answering written questions than they were in teaching videos. Second, the component scores of C, S, and I for each LA followed the same order as suggested by both instruments. For example, both instruments suggested the same sequence of $C > S > I$ for LA-1 and $C > I > S$ for LA-2. Generally, all the LAs were most competent in knowledge of curriculum (C) and struggled more with identifying students' understanding (S) and appropriate instructional strategies (I). Third, both instruments suggested that LA-2 and LA-3 possessed comparable PCK-Q which was stronger than LA-1 in the fall semester, and LA-3 and LA-4 possessed comparable PCK-Q in the spring semester. Convergent validity was supported by the consistency of the patterns between PCK-Q measured by the two instruments regardless of the content areas of classical mechanics or electromagnetism.

PCK-Q of the three LAs from the fall semester matched the fact that LA-2 and LA-3 were more experienced (1.5 yr) LAs than LA-1 (0 yr). In the following spring semester, LA-3 was more experienced (2 yr) than LA-4 (1 yr). Such a difference was represented by the narrated PCK-Q from written questions (i.e., LA-3: 0.74 and LA-4: 0.67), but not by the performed PCK-Q from videos (i.e., LA-3: 0.69 and LA-4: 0.71). A close look into the videos revealed that LA-3 shared ten videos of a total length 465 min and 10 s, which doubled the total length of the eight videos shared by LA-4 which was 209 min and 49 s. Besides, LA-4 answered three less written questions than LA-3 did. It is reasonable to assume that LA-4 might be more sensitive to the artifacts shared with us than LA-3 and was more likely to select videos without ineffective teaching scenarios. Thus, the videos from LA-4 might be biased. We controlled the difference in the number of written questions each LA voluntarily answered by using the questions that all LAs in a semester answered. Unfortunately, we could not control the difference in videos. This is a limitation of this study. LA-3 was the only LA who participated in both semesters. His performed PCK-Q was stable, i.e., 0.73 in fall and 0.69 in the following spring. The component scores shared the same pattern of $C > S > I > O$. LA-3's scores in narrated PCK-Q and written O-C-S-I in the spring were lower than those in the fall, especially orientation (i.e., 1.00 in the fall and 0.67 in the spring). This difference might because we used more questions from LA-3 in the spring ($n = 9$) than in the fall ($n = 4$). More questions probably involved a wider range of difficulty levels. An alternate reason could be that LA-3 was generally more competent in teaching classical mechanics than electromagnetism. Overall, the predictive validity of both instruments was supported by the alignment of PCK-Q with the LAs' teaching experiences and the consistency of LA-3's PCK-Q across two semesters.

C. Q3. How were the LAs prepared to intervene with student learning through questioning as suggested by their performed and narrated PCK-Q?

LAs' performed PCK-Q inferred from videos represented their impromptu responses to students contingent on specific situations during class instruction. LAs' narrated PCK-Q inferred from written questions represented their pondered responses when they had sufficient time to analyze a given situation or what they perceived as legitimate responses aligned with educational theories regarding inquiry teaching. According to the patterns in Table VII, all the LAs' O scores from videos were lower than the O scores from written responses. The LAs seemed to be aware of the importance of questioning in an inquiry-oriented setting but were less questioning-oriented in practice. According to the LAs in the interviews, they agreed with the idea of questioning for inquiry teaching. However, they sometimes found questioning ineffective because it aroused students' anxiety and frustration, which would impact their own credibility as "teachers." They would resort to direct instruction because it was more efficient to address students' difficulties. The gap between narrated and performed O echoed the theory-practice gap regarding questioning in LA preparation.

As for the other three components, all the LAs were most competent in C but not that much in S and I. This was reasonable considering that the weekly LA preparation sessions were focused primarily on physics content knowledge. On the other hand, this finding suggested that the pedagogical instruction regarding questioning from the methods course might not prepare LAs with strong knowledge of student understanding (S) and corresponding instructional strategies (I). All the LAs' C scores from videos were noticeably higher than their C scores from written questions, which was probably because the

TABLE VII. Summary of performed and narrated PCK-Q of the LAs from both semesters.

Fall 2020 semester, classical mechanics					
	O-Video	C-Video	S-Video	I-Video	Performed PCK-Q
LA-1	0.48	0.84	0.40	0.40	0.53
LA-2	0.48	0.96	0.52	0.79	0.69
LA-3	0.56	0.94	0.77	0.66	0.73
	O-Written	C-Written	S-Written	I-Written	Narrated PCK-Q
LA-1	0.88	0.50	0.38	0.38	0.54
LA-2	1.00	0.88	0.75	0.88	0.88
LA-3	1.00	0.88	0.75	0.88	0.88

Spring 2021 semester, electromagnetism					
	O-Video	C-Video	S-Video	I-Video	Performed PCK-Q
LA-3	0.51	0.98	0.67	0.60	0.69
LA-4	0.48	0.96	0.73	0.66	0.71
	O-Written	C-Written	S-Written	I-Written	Narrated PCK-Q
LA-3	0.67	0.83	0.78	0.67	0.74
LA-4	0.89	0.67	0.67	0.44	0.67

representation of content knowledge in written questions was different from that in the course curriculum. The LAs' content knowledge might not be as robust as instructors as they struggled with the same physics concepts contextualized in different scenarios. In this sense, the written questions were more sensitive to LAs' misconceptions. Overall, the LAs' performed PCK-Q was lower than their narrated PCK-Q except for LA-4, which is reasonable because written questions were easier to the LAs than teaching. During teaching, the LAs needed to instantly analyze contextual information and take actions based on rapidly changing conversations with students. With written questions, the LAs had sufficient time to analyze given scenarios that were simplified snapshots of teaching. This result suggests that the written questions designed as an assessment tool could also be used for training purposes. Like microteaching, the written questions could be tailored to shield LAs from the complexity of teaching and direct their attention exclusively to the practice of questioning in specific contexts.

We further examined performed and narrated PCK-Q of each LA across multiple video vignettes or written questions, but there were no clear patterns. We failed to use a LA's performance in one written question to predict their performance in another, or to use a LA's vignette levels to predict their performance in another vignette with a similar context (e.g., students struggle with a lab). The LAs' use of questioning as well as its effectiveness was contingent on contextual factors like students' abilities and difficulty levels of physics concepts. This result justified that LAs' PCK-Q should be measured from multiple scenarios representing a wide range of contexts rather than one or two limited cases. The more diverse the videos or written questions from a LA are, the more accurately the instruments could describe the LA's competency of questioning. Overall, the participating LAs were prepared with the physics content knowledge required by the course curricula, which was insufficient for them to pinpoint student understanding involved in extemporaneous responses. The significance of questioning was successfully seeded in the LAs' minds, but the use of guiding questions in response to contingent situations was a challenge due to their inadequate knowledge of S and I. Thus, direct instruction was a more efficient strategy for the LAs to address students' needs. The inquiry-oriented structure of both courses might be affected by minilectures from the LAs to students in group work.

V. DISCUSSION AND IMPLICATIONS

This study was aimed to introduce two instruments for the assessment of LAs' PCK-Q, which were a video coding schema (Tables III and IV) and written questions (Table V). We took the stance that LAs are more knowledgeable others [8] who use inquiry-oriented pedagogies like questioning to scaffold student learning (Fig. 1). For both instruments, we prioritized the pedagogy of questioning over direct instruction through the component of O (Table I) because questioning holds students' accountable to their learning [20,21]. We also valued direct instruction through the component of I as long as it could potentially address students' needs. Thus, the four components of O-C-S-I target different aspects of PCK-Q that a LA needs to ask effective guiding questions. Both instruments accommodate the context-contingent nature of questioning by deriving LAs' PCK-Q from multiple scenarios [44], which could enhance the accuracy of the assessment of LAs' PCK-Q. In addition to the types of questions, we examined the physics content knowledge involved in questions and the possible impact of a questioning chain to students' conceptual learning. Specifically, the vignette levels from Qa to Ne in the video coding schema (Table IV) suggest how LAs would possibly intervene with student learning in their interaction. Our findings have supported the convergent and predictive validity as well as the interrater and the parallel-form reliability of the instruments.

We juxtaposed the two instruments to illustrate their applications. The coding schema measures LAs' practice of questioning directly and infer their PCK-Q from the frequencies of vignette levels. The equations for performed O-C-S-I were derived from our theoretical deductions. The components of O and C demonstrated better consistency than S and I. While analyzing videos, we found it especially difficult to separate S and I because they worked together to shape LAs' impromptu responses to students. We hypothesized a model of performed PCK-Q as shown in Fig. 4. The components of O and C (i.e., the rectangles) are more fundamental competencies of LAs to support the context-contingent competencies of S and I (i.e., the circles). This is probably why O and C could be more reliably inferred. We are confident that the overlap among O-C-S-I is the percentage of Qa and Qb vignettes combined because effective guiding questions require all four components. The equation for each component needs to be further examined and may be subject to adjustment in light of practical data. Educators could continue refining the model and equations for performed PCK-Q with more LAs or only use the coding schema (Tables III and IV) to describe LAs' interaction with students without quantifying their performed PCK-Q. The written questions measure LAs' narrated PCK-Q directly and predict their performance based on average scores of O-C-S-I. The written questions enable the separation of O-C-S-I because they explicate LAs' reasoning process from their articulation. Thus, the written questions are more advantageous to measure LAs' PCK-Q. To better predict LAs' performance in a specific course, educators could adjust teaching scenarios in question stems to align with their curriculum, such as switching between vector and scalar equations or between calculus and algebra representations, so LAs would be more familiar with the context.

The patterns of the LAs' PCK-Q also shed light on the effectiveness of LA preparation in terms of an instructional methods course and weekly preparation sessions [9,14]. Both instruments returned consistent results about the gap in orientation (O) between performed and narrated PCK-Q, the LAs' strength in physics content knowledge (C), and their struggles with students' understanding (S) and corresponding responses (I). The LAs' strong O in written questions suggests that the methods course regarding educational theories successfully seeded the importance of questioning in the LAs' minds. In practice, a LA might find ineffective the questions that he or she intended to use as they put in their written responses. Thus, the LAs' weaker O in videos does not necessarily mean that they used questions less frequently but that they were less questioning-oriented in supporting students to accomplish a learning task. In other words, the measurement of O from videos inevitably involved the measurement of I, which again demonstrates the challenge to separate overlapping components (i.e., O and I) in performed PCK-Q. The LAs' relatively strong C suggests the effectiveness of the weekly sessions in preparing the LAs with physics content knowledge involved in the course curriculum. The LAs were equipped with the fundamental competencies of O and C (Fig. 4) necessary for them to be effective LAs, but not the context-contingent competencies of S and I necessary for them to be questioning-oriented LAs.

Teaching experience seemed to be able to compensate for S and I because more experienced LAs (i.e., LA-2 and LA-3) demonstrated stronger competencies in the two components. Unfortunately, LAs are undergraduate non-education majors with limited teaching experiences and their term of LA service is typically short [10,14]. Thus, the challenge boils down to how LAs can be efficiently prepared with a repertoire of guiding questions for various teaching scenarios. One possible solution is practice-based LA preparation [38] where LAs are engaged in concrete cases of LA-student interaction. In methods courses, theoretical instruction could be accompanied with video segments analyzed by our coding scheme to dissect how effective questions support student learning. In weekly preparation sessions, LAs could work on our written questions regarding the target concept for the following week to preview students' possible misconceptions and ponder candidate guiding questions. Both instruments that we designed for assessment could be used for pedagogical education as well. Another solution is rehearsal [46] where LAs engage in role-playing activities that imitate class interaction with students. Rehearsal would allow LAs to refer back to their own learning experiences and

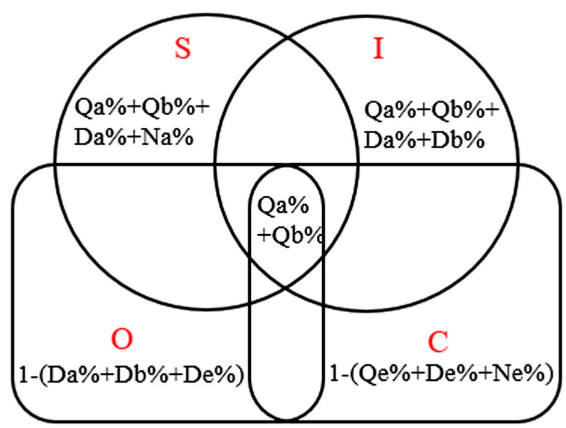

FIG. 4. The hypothesized model of LAs' performed PCK-Q derived from teaching videos.

receive instantaneous feedback from peers or instructors. Meanwhile, LAs could be encouraged to position themselves as more knowledgeable others who share their learning experience and struggle together with students rather than authority obligated to convey correct information to students (i.e., lecturing oriented) or to ask high-quality guiding questions (i.e., inquiry oriented). LAs could be prompted by students' guiding questions as well.

VI. LIMITATIONS

We admitted two limitations of this study. First, the LAs' videos were not randomly captured due to online instruction during the COVID pandemic. Thus, the videos might be biased in representing a LA's teaching practice. Second, only four LAs agreed to participate in this study. We compensated for the small sample size by collecting multiple videos and written questions from each LA. Yet, the findings may not be generalizable to a large population of LAs. Both instruments need to be further examined with more LAs or in another instructional context.

ACKNOWLEDGMENTS

The research reported here is supported by the National Science Foundation, through Grant No. 1838339 to Texas Tech University. The opinions, findings, and conclusions or recommendations expressed are our own and do not necessarily reflect the views of the National Science Foundation.

[1] Joanne Gainen, Barriers to success in quantitative gatekeeper courses, New Dir. Teach. Learn. **1995**, 5 (1995).
[2] David E. Meltzer and Valerie K. Otero, A brief history of physics education in the United States, Am. J. Phys. **83**, 447 (2015).
[3] Valerie K. Otero and David E. Meltzer, 100 years of attempts to transform physics education, Phys. Teach. **54**, 523 (2016).
[4] Eugenia Etkina, David T. Brookes, and Gorazd Planinsic, *Investigative Science Learning Environment* (Morgan & Claypool Publishers, San Rafael, CA, 2019), https://iopscience.iop.org/book/mono/978-1-64327-780-6.
[5] Eric Brewe, Laird Kramer, and George O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with class, Phys. Rev. ST Phys. Educ. Res. **5**, 013102 (2009).
[6] Anthony P. Barrasso and Kathryn E. Spilios, A scoping review of literature assessing the impact of the learning assistant model, Int. J. STEM Educ. **8**, 12 (2021).
[7] Barbara A. Crawford, From inquiry to scientific practices in the science classroom, in *Handbook of Research on Science Education, Volume II* (Routledge, New York, London, 2014), pp. 529–556, https://www.taylorfrancis.com/chapters/edit/10.4324/9780203097267-36/inquiry-scientific-practices-science-classroom-barbara-crawford.
[8] Lev Vygotsky, Interaction between learning and development, Read. Dev. Child. **23**, 34 (1978), https://ia.eferrit.com/ea/a6589cd862231ed3.pdf.
[9] Renee Michelle Goertzen, Eric Brewe, Laird H Kramer, Leanne Wells, and David Jones, Moving toward change: Institutionalizing reform through implementation of the learning assistant model and open source tutorials, Phys. Rev. ST Phys. Educ. Res. **7**, 020105 (2011).
[10] Cody Sandifer and Eric Brewe, *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices* (American Physical Society, College Park, MD, 2015), https://www.compadre.org/portal/items/detail.cfm?ID=13728.
[11] Ben Van Dusen, Laurie Langdon, and Valerie Otero, Learning assistant supported student outcomes (LASSO) study initial findings, arXiv:1509.05358.
[12] Eleanor W. Close, Jessica Conn, and Hunter G. Close, Becoming physics people: Development of integrated physics identity through the learning assistant experience, Phys. Rev. Phys. Educ. Res. **12**, 010109 (2016).
[13] Eleanor W. Close, Jessica Conn, and Hunter G. Close, Learning assistants' development of physics (teacher) identity, in *Proceedings of PER Conf. 2013, Portland, OR*, 10.1119/perc.2013.pr.010.
[14] Valerie Otero, Steven Pollock, and Noah Finkelstein, A physics department's role in preparing physics teachers: The colorado learning assistant model, Am. J. Phys. **78**, 1218 (2010).
[15] Mica Estrada, Myra Burnett, Andrew G. Campbell, Patricia B. Campbell, Wilfred F. Denetclaw, Carlos G. Gutiérrez, Sylvia Hurtado, Gilbert H. John, John Matsui, Richard McGee *et al.*, Improving underrepresented minority student persistence in stem, CBE Life Sci. Educ. **15**, es5 (2016).
[16] Jada-Simone S. White, Ben Van Dusen, and Edward A. Roualdes, The impacts of learning assistants on student learning of physics, arXiv:1607.07469.
[17] Nafis I. Karim, Alexandru Maries, and Chandralekha Singh, Do evidence-based active-engagement courses reduce the gender gap in introductory physics?, Eur. J. Phys. **39**, 025701 (2018).
[18] Thomas Pentecost, Laurie Langdon, Margaret Asirvatham, Hannah Robus, and Robert Parson, Graduate teaching assistant training that fosters student-centered instruction and professional development, J. Coll. Sci. Teach. **41**, 68 (2012), https://sei.ubc.ca/bitstream/seima/2139/1/Pentecost_TA-Training_JCST2012.pdf.
[19] Adrian Madsen, Sarah B. McKagan, and Eleanor C. Sayre, How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).

[20] Christine Chin, Teacher questioning in science classrooms: Approaches that stimulate productive thinking, J. Res. Sci. Teach. **44**, 815 (2007).

[21] Christine Chin and Jonathan Osborne, Students' questions: a potential resource for teaching and learning science, Stud. Sci. Educ. **44**, 1 (2008).

[22] Aisha Kawalkar and Jyotsna Vijapurkar, Scaffolding science talk: The role of teachers' questions in the inquiry classroom, Int. J. Sci. Educ. **35**, 2004 (2013).

[23] Shirley Magnusson, Joseph Krajcik, and Hilda Borko, Nature, sources, and development of pedagogical content knowledge for science teaching, in *Examining Pedagogical Content Knowledge* (Springer, New York, 1999), pp. 95–132, 10.1007/0-306-47217-1.

[24] Lee S. Shulman, Those who understand: Knowledge growth in teaching, Educ. Res. **15**, 4 (1986), http://depts.washington.edu/comgrnd/ccli/papers/shulman_ThoseWhoUnderstandKnowledgeGrowthTeaching_1986-jy.pdf.

[25] Paul M. Miller, Jeffrey S. Carver, Aniketa Shinde, Betsy Ratcliff, and Ashley N. Murphy, Initial replication results of learning assistants in university physics, AIP Conf Proc. **1513**, 30 (2013).

[26] Steven J. Pollock and Noah D. Finkelstein, Sustaining educational reforms in introductory physics, Phys. Rev. ST-Phys. Educ. Res. **4**, 010110 (2008).

[27] Zahra Hazari, Gerhard Sonnert, Philip M. Sadler, and Marie-Claire Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47**, 978 (2010).

[28] Nadia Sellami, Shanna Shaked, Frank A. Laski, Kevin M. Eagan, and Erin R. Sanders, Implementation of a learning assistant program improves student performance on higher-order assessments, CBE Life Sci. Educ. **16**, ar62 (2017).

[29] Robert M. Talbot, Laurel M. Hartley, Katrina Marzetta, and Bryan S. Wee, Transforming undergraduate science education with learning assistants: Student satisfaction in large-enrollment courses, J. Coll. Sci. Teach. **44**, 24 (2015), https://bud-t.com/wp-content/uploads/2018/12/Talbot-et-al.-2015-Transforming-undergraduate-science-education-with-learning-assistants-Student-satisfaction-in-large-enrollment-courses.pdf.

[30] Kathleen A. Jeffery, Samantha Marie Frawley Cass, and Ryan David Sweeder, Comparison of students' readily accessible knowledge of reaction kinetics in lecture-and context-based courses, J. STEM Educ. **19**, 5 (2019), https://www.jstem.org/jstem/index.php/JSTEM/article/view/2338.

[31] Kari L. Stone, Sarah E. Shaner, and Carol M. Fendrick, Improving the success of first term general chemistry students at a liberal arts institution, Educ. Sci. **8**, 5 (2018).

[32] Jessica L. Alzen, Laurie S. Langdon, and Valerie K. Otero, A logistic regression investigation of the relationship between the learning assistant model and failure rates in introductory stem courses, Int. J. STEM Educ. **5**, 56 (2018).

[33] Carolyn P. Schick, Trying on teaching: Transforming stem classrooms with a learning assistant program, in *Strategies Promoting Success of Two-Year College Students* (ACS Publications, Washington D.C., 2018), pp. 3–27, 10.1021/bk-2018-1280.ot001.

[34] Jennifer K. Knight, Sarah B. Wise, Jeremy Rentsch, and Erin M. Furtak, Cues matter: Learning assistants influence introductory biology student interactions during clicker-question discussions, CBE Life Sci. Educ. **14**, ar41 (2015).

[35] Kara E. Gray, David C. Webb, and Valerie K. Otero, Effects of the learning assistant model on teacher practice, Phys. Rev. Phys. Educ. Res. **12**, 020126 (2016).

[36] Ken Zeichner, Rethinking the connections between campus courses and field experiences in college-and university-based teacher education, J. Teach. Educ. **61**, 89 (2010).

[37] Amreen Nasim Thompson, Robert M. Talbot, Leanne Doughty, Hannah Huvard, Paul Le, Laurel Hartley, and Jeffrey Boyer, Development and application of the action taxonomy for learning assistants (Atlas), Int. J. STEM Educ. **7**, 1 (2020).

[38] James Hiebert and Anne K. Morris, Teaching, rather than teachers, as a path toward improving classroom instruction, J. Teach. Educ. **63**, 92 (2012).

[39] Sami Lehesvuori, Umesh Ramnarain, and Jouni Viiri, Challenging transmission modes of teaching in science classrooms: Enhancing learner-centredness through dialogicity, Res. Sci. Educ. **48**, 1049 (2018).

[40] Prudence M. Smith and Mark W. Hackling, Supporting teachers to develop substantive discourse in primary science classrooms, Aust. J. Teach. Educ. (Online) **41**, 151 (2016), https://files.eric.ed.gov/fulltext/EJ1098139.pdf.

[41] Soonhye Park and J. Steve Oliver, Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals, Res. Sci. Educ. **38**, 261 (2008).

[42] Lori A. Reinsvold and Kathryn F. Cochran, Power dynamics and questioning in elementary science classrooms, J. Sci. Teach. Educ. **23**, 745 (2012).

[43] Antti Lehtinen, Sami Lehesvuori, and Jouni Viiri, The connection between forms of guidance for inquiry-based learning and the communicative approaches applied—a case study in the context of pre-service teachers, Res. Sci. Educ. **49**, 1547 (2019).

[44] Wolff-Michael Roth, Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses, J. Res. Sci. Teach. **33**, 709 (1996).

[45] J. Richard Landis and Gary G. Koch, The measurement of observer agreement for categorical data, Biometrics **33**, 159 (1977).

[46] Magdalene Lampert, Megan Loef Franke, Elham Kazemi, Hala Ghousseini, Angela Chan Turrou, Heather Beasley, Adrian Cunard, and Kathleen Crowe, Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching, J. Teach. Educ. **64**, 226 (2013).