How to help students learn: An investigation of how in- and pre-service physics teachers respond to students' explanations

Danijela Dodlek,[1,2,*] Gorazd Planinsic,[2] and Eugenia Etkina[3]

[1]*Department of Physics, Josip Juraj Strossmayer University of Osijek,*
Trg Ljudevita Gaja 6, HR-31000 Osijek, Croatia
[2]*Faculty of Mathematics and Physics, University of Ljubljana,*
Jadranska ulica 19, 1000 Ljubljana, Slovenia
[3]*Graduate School of Education, Rutgers University, New Brunswick, New Jersey 08904, USA*

(Received 22 August 2023; accepted 5 January 2024; published 5 April 2024)

Research carried out through the last 20 years gave us undeniable evidence that to learn anything we need to be active participants, not passive observers. One of the important aspects of learning physics is constructing explanations of physical phenomena. To support and guide students toward constructing their explanations, teachers need to be attentive and responsive to students' explanations. To learn how physics teachers interpret and respond to students' explanations we investigated pre- and in-service physics teachers' responses to students' written explanations of their answers to a complex physics problem. The survey administered to the participants included the problem statement and four authentic student explanations. The participants were asked to identify each student's strengths and weaknesses and to provide a response to that student. We found that while the participants were successful in identifying productive and problematic aspects of student reasoning, they rarely built on student reasoning when responding to the students, mostly focusing on addressing problematic aspects. The paper discusses why this finding is important for physics teacher preparation programs and professional development programs.

DOI: 10.1103/PhysRevPhysEducRes.20.010120

I. INTRODUCTION

Through the last two decades, we obtained a great deal of indisputable evidence that to learn anything students need to be active participants in the learning process. This evidence comes from studying how the human brain works and how people learn [1,2], as well as from numerous studies that focus on specific subjects such as physics [3–6]. One of the aspects of active learning is students' construction of their explanations. Students' self-explanations refer to the practice of students articulating their understanding of a concept or problem-solving process, and it involves explaining the steps they took, the reasoning behind their choices, and connecting new information to prior knowledge. The constructive process of self-explaining was found to lead to better learning and understanding [7,8]. While recognizing that the role students themselves have in the learning process is important, we must also acknowledge the role that teachers play in facilitating students' learning.

When teachers actively listen and respond to students' explanations, they play a crucial role in guiding and shaping the learning process. The teachers gain insights into students' thinking processes, identify any difficulties or gaps in understanding, assess the effectiveness of their teaching methods, and modify their lessons based on what students are saying. Therefore, teachers need to identify productive and problematic aspects of students' explanations to foster their learning. While recognizing and addressing problematic aspects is important, recognizing the seeds of the right ideas when they are not transparent is also very important. Research on the functioning of the human brain [1,2] emphasizes that new ideas can only be constructed when they are linked to previously held beliefs. Consequently, identifying the strengths in student explanations allows teachers to build on these strengths to help their students move forward.

While many studies focused on investigating how pre-service and in-service (elementary, middle, or high school) science and mathematics teachers interpret and respond to student explanations [9–17], none dealt with physics teachers exclusively. Therefore, the purpose of this study is to address this research gap by studying how physics teachers interpret and respond to students' written explanations of their answers to a complex physics problem. Additionally, we wanted to learn if different levels of

---

[*] ddodlek@fizika.unios.hr

experience affect the way teachers respond to students' needs. Therefore, we also investigate the differences in the responses to students' explanations among different groups of in-service and pre-service physics teachers.

To analyze participants' responses we used the recently developed Content Knowledge for Teaching Energy (CKT-E) framework [18]. Specifically, we used one of the components of CKT-E called the tasks of teaching (ToTs). The ToTs embody all activities in which physics teachers engage during instruction. By employing this framework, we aimed to address the following research questions:

1. How do teachers interpret and respond to students' incorrect explanations?
2. How do teachers interpret and respond to students' correct explanations?
3. What are the differences in responses to students' explanations among different groups of in-service and pre-service teachers?

II. LITERATURE REVIEW

A. Models of teacher knowledge

If teachers want to help their students learn, it is not enough for them to know the subject matter [19,20]. Teachers also need to possess special knowledge that combines their understanding of a particular subject matter (content knowledge) with their knowledge of effective teaching strategies and methods (pedagogical knowledge) to be able to transform content in a way that is accessible and meaningful to students.

In the 1980s Shulman introduced the concept of pedagogical content knowledge (PCK) [19,20] as a way to identify and illustrate a distinct type of professional knowledge that sets teachers apart from other professionals. For Shulman, PCK entails the integration of content and pedagogy, enabling teachers to understand how certain topics, problems, or issues are organized, presented, and tailored to learners' diverse skills and interests and subsequently framed for instruction [20].

Due to its vital role in both teacher education and assessment, over the past three decades, there has been a significant body of research dedicated to the exploration of what constitutes and how to assess pedagogical content knowledge (PCK) [21–25]. The initial attempt to define this knowledge operationally was in the Magnusson *et al.*'s model of science PCK [26] in which they identified five key aspects of PCK: (1) orientation toward science teaching, (2) knowledge and beliefs about science curriculum, (3) knowledge and beliefs about students' understanding of specific science topics, (4) knowledge and beliefs about assessment in science, and (5) knowledge and beliefs about instructional strategies for teaching science. Etkina [27] further demonstrated the applicability of these five aspects to the teaching of physics, highlighting the subject-specific nature of PCK.

In 2015, the consensus model of PCK emerged, positioning PCK as just one element within teachers' broader professional knowledge and practice, with a shift towards topic-specific knowledge [28]. However, this model lacked specificity in how it represented and addressed the complexities of PCK. In response, the refined consensus model of PCK was introduced in 2019, categorizing PCK into three domains: collective PCK, personal PCK and enacted PCK [29]. These domains "describe the specialized professional knowledge held by multiple educators in a field, to the personalized professional knowledge held by an individual teacher in science, and the unique subset of knowledge that a teacher draws on to engage in pedagogical reasoning during the planning of, teaching of, and reflecting on a lesson" (p. 82).

Two issues are important here: (1) do content experts need PCK to be able to help students learn or their content knowledge is enough, and (2) can PCK be developed only through the practice of teaching [30], or is it possible to develop some aspects of it during a teacher preparation program? Research by Seeley *et al.* shows that while undergraduate physics majors in the last year of studies could successfully solve problems given to high school students, they were not able to help those students who had difficulties solving a problem [31]. Etkina described a physics teacher preparation program focused on the development of PCK and provided evidence of its success [27].

Although Shulman introduced the concept of Pedagogical Content Knowledge [20], it was Ball *et al.* who conceptualized this distinct knowledge as used in the *practice* of teaching [32]. What is necessary for teachers to know to be able to effectively implement complex tasks (referred to as the tasks of teaching) that promote student learning? Ball *et al.* created a list of what constitutes such knowledge and grouped those elements into two main categories: subject matter knowledge and pedagogical content knowledge—similar to Shulman's PCK [32]. Drawing on Shulman's ideas, Ball *et al.* [32] developed and defined the model of mathematical knowledge for teaching (MKT). MKT refers to "the mathematical knowledge needed to carry out the work of teaching mathematics" (p. 395). Essentially, Ball *et al.* introduced a more comprehensive concept called content knowledge for teaching (CKT) [32]. CKT framework has a broader scope than PCK because it encompasses both subject matter knowledge reconceptualized for the purpose of teaching and the pedagogical implementation of that knowledge.

While most of the research on describing and evaluating CKT was done in mathematics (see, e.g., [32–34]) and reading (see, e.g., [35–37]), recently this line of research was extended to the field of physics by Etkina *et al.* [18]. Their objective was to design an assessment tool for physics teachers' CKT in the context of energy. However, before developing the assessment, they needed to establish the construct to be measured. To achieve this goal, the authors developed the framework of content knowledge for teaching energy in the context of mechanics in high schools (CKT-E) [18]. This framework enabled them to design an assessment to evaluate physics teachers' content knowledge of this domain. The CKT-E framework consists of two elements. The first is the *tasks of teaching* (ToTs) which represents the tasks physics teachers do during instruction and the second element is the *Student energy targets* which "describe disciplinary core ideas, science practices, and cross-cutting concepts that are important for student learning of energy in the context of mechanics" (p. 14).

Our research uses the theoretical lens of ToTs to analyze participants' responses to the explanations written by the students. There are six ToTs in the CKT-E framework:

I. anticipating student thinking around science ideas;
II. designing, selecting, and sequencing learning experiences and activities;
III. monitoring, interpreting, and acting on student thinking;
IV. scaffolding meaningful engagement in a science learning community;
V. explaining and using examples, models, representations, and arguments to support students' scientific understanding; and
VI. using experiments to construct, test, and apply concepts [18].

Each of these tasks consists of several specific subtasks (for a complete list together with the explanations see Ref. [18]). Some of the subtasks found in Tasks II (subtasks a–j) and III (subtasks a–g), which teachers implement in the classroom, align closely to our study in which we investigated how participants interpret students' productive and problematic ideas, the kind of feedback, if any, participants give to students, whether participants' responses engage students in metacognition and whether teachers build on students' productive ideas and address problematic ones. The specific subtasks we used in our analysis of the participants' responses to student explanations are as follows:

— II. (d) addressing learners' actual learning trajectories by building on productive elements and addressing problematic ones,
— III. (b) interpreting productive and problematic aspects of student thinking and mathematical reasoning,
— III. (e) identifying specific cognitive and experiential needs or patterns of needs and providing students with descriptive feedback and
— III. (f) engaging students in metacognition and epistemic cognition.

B. Models of student ideas

Research shows that student thinking is complex, context-dependent, and progressive [38]. To help students learn, teachers must accept that students come to classrooms with prior knowledge based on their life experiences [1,2]. This prior knowledge significantly influences students' learning [39–41].

One of the models that describes students' existing knowledge is the Knowledge in Pieces framework introduced by diSessa in the 1980s [42]. In his view, students' prior knowledge "consists of a rather large number of fragments…" which he calls phenomenological primitives or "p-prims" (p. 52). The name illustrates that p-prims are connected to how events in the real world (phenomena) are interpreted and that p-prims are basic elements of memory [43]. Essentially, p-prims are small and context-dependent student ideas that are neither correct nor incorrect in and of themselves. Some examples of diSessa's p-prims include: (1) *Ohm's p-prim*—the stronger the cause, the stronger its effect; the stronger the resistance or impediment, the stronger its effect (bigger potential difference means bigger current, bigger force means higher velocity…) and (2) *constancy/conservation*—some quantity staying the same.

As can be seen from the examples above, p-prims are sometimes suitable for the situation and sometimes they are not, depending on the students' prior experiences and the language that we use. If we could identify the p-prim on which the students based their reasoning, we could use it as a productive resource on which to build students' new knowledge by adjusting the language or the context [38].

Hammer *et al.* expanded on the concept of resources [44–46]. They argue resources are bigger than p-prims pieces of existing knowledge that students possess. Resources can be activated individually or together with other resources when a student reasons about a physics concept. Resources can be categorized as conceptual resources and epistemological resources.

Conceptual resources differ from p-prims not only in size (they are larger) but also in scope. Many of them are physics-related. While p-prims reflect primitive knowledge, conceptual resources represent more advanced, content-specific knowledge [47]. For example, a conceptual resource of energy as a substance might help a student correctly explain how a lightbulb is powered by a battery but might lead to an incorrect answer that current is used up when analyzing what happens to electric current in a circuit. Epistemological resources, on the other hand, explain how students approach learning [45]. For example, students might think that knowledge comes from authority, or that physics knowledge is memorizing facts and formulas.

Like p-prims, conceptual and epistemological resources are activated when we ask students questions or when they are interpreting reading materials. P-prims and resources occasionally make the students give answers that are "wrong" or seem wrong, but it should not stop us from trying to determine the source of their ideas and navigating this source in a productive direction.

In this research, we adopt the KiP model as the foundation of our understanding of student ideas and we emphasize the significance of recognizing productive "seeds" in students' explanations and building on them.

C. Responsive Teaching and Ambitious Science Teaching

Responsive Teaching (RT) is an instructional approach to education that focuses on tailoring instruction to meet the diverse and evolving needs of individual students.

In RT teachers begin with the premise that students' ideas are both resourceful and productive, and they actively involve and enhance these ideas [48–51]. Responsive teachers focus on understanding the interpretations that their students derive from their disciplinary experiences [50,52,53]. Teachers listen carefully to what their students are saying so that they can build on their ideas instead of correcting them. Responsive instruction also *recognizes disciplinary connections within students' ideas*, meaning that the teachers attentively listen for emerging links between students' interpretations and the discipline [48,50]. These "disciplinary progenitors" [54] or "seeds of science" [51] may include, for example, "children's puzzlement over a phenomenon, their citing evidence to support an idea, their efforts toward precision, their using mechanistic reasoning (or the beginnings of it) to support their predictions or explanations, or their devising an informal experiment or suggesting an explanation…" [55] (p. 2). Responsive teaching *takes up and pursues the substance of student thinking* [48,50].

The ideas of responsive teaching applied to science are known as the Ambitious Science Teaching (AST) [56]. The goal of AST pedagogy is to elevate the quality of science education by emphasizing authentic scientific practices, complex tasks, and meaningful engagement.

Ambitious Science Teaching aims to empower students to become critical thinkers, problem solvers, and lifelong learners in the field of science by having students explain their thinking to others (metacognitive practice), construct and compare theories, and justify different claims [57]. In this process, students have support from their teachers who inform them about their current level of understanding and provide additional guidance on areas they may need to improve. Such targeted feedback is beneficial for all students.

Acting like a guide for both of these pedagogical approaches, Responsive Teaching and Ambitious Science Teaching, is the knowledge in pieces compass. KiP serves as the common thread that highlights the importance of acknowledging, recognizing and building on productive aspects of students' ideas in both RT and AST.

D. Responding to students' ideas

In this section, we turn our attention to the concepts of feedback and metacognition as important features of both Responsive Teaching and Ambitious Science Teaching.

To help students learn, teachers need to recognize and respond to their ideas [58]. We can think of recognizing and responding to student ideas as formative assessment [59]. Research shows that implementing formative assessment practices is inherently challenging for teachers as they need to identify and interpret student ideas at the moment and subsequently build on those ideas to foster students' progress [60].

Research on teachers' interpretation and responses to student ideas produced mixed findings. While the study of elementary mathematics teachers conducted by Ebby *et al.* [14], showed that the majority of K-5 mathematics teachers focused primarily on the strategies that students used to solve simple math problems as a sign of growing understanding, Kristinsdóttir *et al.* found that upper secondary mathematics teachers looked primarily for mistakes and the things that are missing in the students' work [11]. Additionally, Gotwals and Birmingham [10] found that preservice science teachers (biology and chemistry majors) tended to conceptualize high school and middle school students' ideas in a binary way—either correct or incorrect, which they interpreted as having misconceptions.

Brookhart [61] provides instructions for teachers on how to give effective feedback to their students. The book highlights the features of feedback that research identified as important for enhancing students' learning such as descriptive rather than judgmental feedback, and positive feedback that includes suggestions for improvement. It is also important to note that providing effective feedback is not a "one-size-fits-all" approach. What works for one student may not work for the others. Therefore, teachers need to listen carefully to what their students are saying and adjust their feedback accordingly.

Hattie and Timperley reviewed several meta-analyses regarding feedback to synthesize a model of effective feedback. According to their work, effective feedback addresses three key questions: Where am I going?, How am I going? and Where to next?. These questions align with the concepts of feed up, feedback, and feed forward [62]. The feed up refers to the goals of learning, feedback refers to the progress that is being made toward achieving the goal, and feed forward focuses on guiding students' future performance and improvement rather than solely evaluating their past performance and aims to provide constructive suggestions, strategies, and resources to help students enhance their future work [62].

The research conducted by Brookhart [61], as well as the findings presented by Hattie and Timperley [62], align well with the principles of feedback in Responsive Teaching (RT) [55] and Ambitious Science Teaching (AST) [56]. As responsive and ambitious teachers constantly monitor students' progress and tailor their instruction in response to students' needs, they need to recognize and build on students' ideas. Teachers' feedback informs students of their current status, guides them toward the learning goals, lets students know about productive ideas in their reasoning and helps them address their weaknesses. Teachers can respond to students' ideas during class time or they can respond to students' written work. Though feedback on students' written work might involve some delay, recent research [63] shows that it is possible to respond to students' written work in real time.

Other research related to responding to student ideas showed a significant positive correlation between teachers' beliefs and students' problem-solving achievement [64]. Peterson *et al.* found that students of teachers who believed that they should build on students' existing knowledge achieved higher scores for solving simple addition and subtraction problems than did children of teachers whose beliefs were less cognitively based [64]. On the other hand, students from both types of classes did equally well on addition and subtraction facts.

Recognizing the importance of being responsive to students and students' ideas, Responsive Teaching and Ambitious Pedagogy also support students in articulating and explaining their thinking (metacognition), as research found that metacognition plays a significant role in promoting students' learning [65–67]. In his work, Ader [68] found that secondary mathematics teachers primarily fostered students' metacognitive skills by encouraging them to articulate and elaborate on their ideas and responses. Examples of metacognitive questions are as follows: *Can you explain more what you mean when talking about resistance to motion?*, *How does this affect your answer? Based on what did you make that assumption?, Can you explain how you know that the smaller mass of car 1 results in the smaller kinetic energy of that car?, Is your limiting case analysis consistent with constancy of momentum?*, etc.

Possessing metacognitive skills empowers individuals to analyze, solve problems, self-monitor, make informed decisions and enhance their performance [69]. Similar to the concept of feed forward, asking metacognitive questions actively engages students in the construction of their knowledge which leads to significant positive effects on students' learning [1–6]. Therefore, teachers need to cultivate students' metacognitive skills to promote effective learning.

E. Investigative science learning environment

The investigative science learning environment (ISLE) approach is a pedagogical approach to learning and teaching physics that helps students construct physics concepts by themselves by following processes similar to those that physicists use when they construct and apply new knowledge [70]. Students working in groups observe carefully selected experiments and describe patterns using their own words, not scientific terminology at first. They then construct their explanations of those patterns and design and conduct new experiments to test those explanations. When the outcomes of these testing experiments do not match predictions based on the explanations under test, they revise the explanations. The ISLE approach[1] naturally builds on the ideas of Responsive Teaching and Ambitious Science Teaching as it is the students who develop their explanations and design experiments to test those ideas through carefully designed activities. The role of a teacher in the ISLE approach is to guide students through these processes. Such guidance is impossible without listening to the students and adjusting instruction based on what they are saying and doing. Additionally, in the ISLE approach students are encouraged to improve and resubmit their work based on the feedback of the teacher without their grade being lowered for multiple attempts. This element of the ISLE approach allows the teachers to meet the needs of diverse students who need different amounts of time and guidance to learn physics [72].

ISLE-based physics teacher preparation programs such as those described in Ref. [73] help pre-service teachers learn how to play the role of a guide while teaching high school physics. The programs have multiple physics teaching method courses all based on the ISLE approach and closely connected to the coursework clinical practice. Preservice teachers develop content knowledge for teaching physics by systematically participating in the activities that their future high school students will do while learning physics. After performing such activities, they reflect on their experiences. When planning their lessons during microteaching experiences (microteaching is teaching lessons to one's peers who play the role of K-12 students, see Ref. [73]) ISLE pre-service teachers learn about student ideas relevant to the topic which were found through research and learn how to anticipate and build on those ideas in their lesson plans [74]. Enacting microteaching, they practice listening to student responses and changing their lesson based on those responses [73]. Finally, they learn to provide oral and written feedback to the students building on productive student ideas through specially constructed exercises. Exercises and activities focused on the development and responses to student ideas are built into multiple physics teaching methods courses in such programs [27].

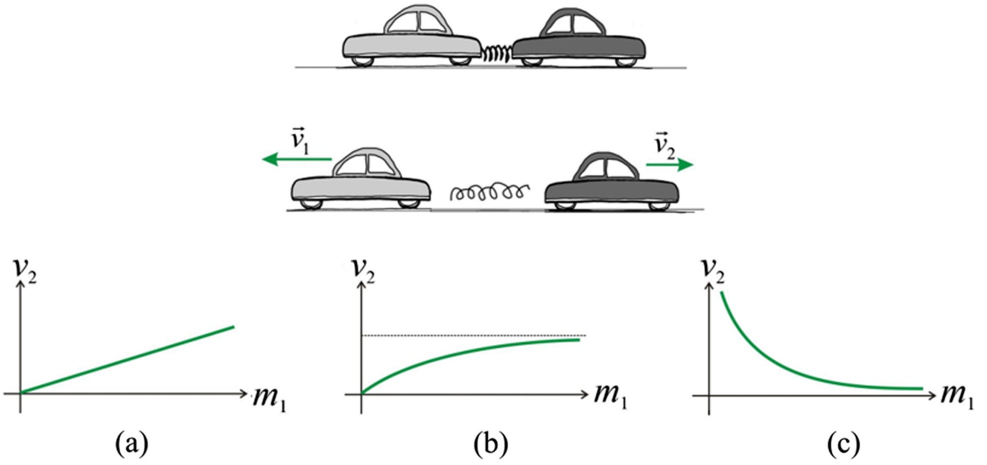
III. DESCRIPTION OF THE STUDY

To investigate how physics teachers interpret and respond to student answer choices for a physics problem and their explanations of the choices, we made a survey using Google Forms. In the survey we gave the participants the problem statement and four students' answer choices and explanations and asked them to comment on the student's strengths and weaknesses, and to provide their hypothetical responses to the student.

---

[1] Another example of an interactive engagement curriculum approach in physics that builds on students' ideas is Modeling Instruction [71].

You place two toy cars on a horizontal table and connect them with a light compressed spring as shown in figure. The spring tries to push the cars apart, but they are tied together by a thread. When the thread is burned, the spring pushes the cars apart. You decide to investigate how the final speed of car 2 depends on the mass of car 1. You run several experiments changing m_1 and measuring v_2 while keeping the compression of the spring and the mass of car 2 constant. Which of the v_2-versus-m_1 graphs do you expect to obtain? Evaluate the graphs by analysing limiting cases. Explain your reasoning. Solve the problem without deriving the equations.



(a) (b) (c)

FIG. 1. The problem students were solving (published in Ref. [76]). The correct answer choice is (b).

To create our survey, we used students' answers and explanations of a nontraditional problem (see Fig. 1) that Emilija Simonović collected as a part of her master thesis [75]. Emilija was comparing how students from different learning environments evaluated proposed solutions. Students had to evaluate the solution by doing the limiting case analysis and to explain their reasoning.

The research team chose four students' answers and explanations in such a way that each explanation was representative of different student ideas. One explanation was correct, while three were incorrect. Out of the incorrect ones, two were for the wrong answer choices and one was for the correct answer choice (the student chose the correct answer but provided an incorrect explanation).

We collected the participants' responses to the anonymous survey over 9 months (April 2022–December 2022). Out of 70 responses, answers from 66 participants were suitable for our study. Responses from four participants were not considered in our analysis because they were either from an elementary in-service teacher (one response) or from pre-service physics teachers who did not have any general education or physics-specific education courses (three responses).

The participants of the study were in-service (high school and university level) (42) and pre-service physics teachers (24). All participants participated voluntarily. We had three groups of in-service teachers: two groups were from two European countries—A and B, and one group was international—teachers were from Europe, Asia, and North America. We also had three groups of pre-service teachers from countries A and B—undergraduate pre-service teachers A, graduate pre-service teachers A and graduate pre-service teachers B. Table I provides a more detailed description of the groups. All groups represent convenience samples.

IV. DATA ANALYSIS

To answer research questions, we analyzed participants' responses to the explanations written by high school students. To analyze responses provided by the participants, we developed coding schemes using one of the components of the content knowledge for teaching energy (CKT-E) framework—tasks of teaching and expert responses.

We used task of teaching II. Designing, selecting, and sequencing learning experiences and activities, and task of teaching III. Monitoring, interpreting, and acting on student thinking [18]. Each of these ToTs has several subtasks, and four were relevant to our study:

1. II. (d) addressing learners' actual learning trajectories by building on productive elements and addressing problematic ones,
2. III. (b) interpreting productive and problematic aspects of student thinking and mathematical reasoning,
3. III. (e) identifying specific cognitive and experiential needs or patterns of needs and providing students with descriptive feedback, and
4. III. (f) engaging students in metacognition and epistemic cognition.

TABLE I. Description of the groups of participants.

Name of the group	Number of participants in the group	Years of teaching physics (number of participants)	Comments
International in-service teachers	13	0–4 (0) 5–9 (1) 10–14 (4) 15–19 (1) 20–24 (4) 25–30 (2) more than 30 (1)	Voluntarily participate in every day discussions and attend monthly workshops through the Facebook group "Exploring and Applying Physics" which was created to help teachers who use or plan to use the investigative science learning environment (ISLE) [70] approach in their classrooms.
In-service teachers A	17	0–4 (0) 5–9 (3) 10–14 (3) 15–19 (2) 20–24 (4) 25–30 (1) more than 30 (4)	Voluntarily participate in monthly professional development workshops by attending a continuous education program in their country.
In-service teachers B	12	0–4 (1) 5–9 (4) 10–14 (2) 15–19 (2) 20–24 (1) 25–30 (1) more than 30 (1)	We do not have information about their professional development activities.
Experienced pre-service teachers A	5	⋯	Graduate pre-service teachers who were about to finish their master's degree. They had three semesters of Didactics of Physics courses and clinical practice. All of them had a bachelor's physics degree. All Didactics of Physics courses and clinical practice were consistent with the ISLE approach.
Less experienced pre-service teachers A	15	⋯	Undergraduate pre-service teachers who were about to finish their bachelor's physics degree. They had one semester of Didactics of Physics ISLE-based course (the first course in the sequence taken by experienced pre-service teachers A).
Less experienced pre-service teachers B	4	⋯	Graduate pre-service teachers who were about to finish their master's degree. All of them had a bachelor's physics degree. They had one or two semesters of general education courses (Pedagogy, Psychology, Didactics) and only 8 weeks of a subject specific pedagogical course called Physics Education. The course was not ISLE based.

We broke these subtasks further down into six categories listed below[2]:

1. Ia—interpreting productive reasoning,
2. Ib—interpreting problematic reasoning,
3. II—providing descriptive feedback,
4. III—engaging in metacognition,
5. IVa—building on productive elements, and
6. IVb—addressing problematic elements.

Before administering the final version of the survey to the participants, four experts in physics education worked at first individually and then collaboratively to respond to all of the questions for each student's explanation. None of the experts individually wrote all of the things that were later used as the expert responses for the coding purposes. Details of expert responses are in Appendix A.

Using the above-mentioned categories, which gave names to our codes, and the expert responses, we developed descriptive coding schemes to analyze participants' responses. We compared what experts and participants wrote for each of the codes and quantified the differences between their responses by assigning numerical scores (0, 1, or 2) to each of the codes. A detailed description of the coding schemes is in Appendix B. We used the same

---

[2] For example, Etkina *et al.* [18] list the subtask III. b) as Interpreting productive and problematic aspects of student thinking and mathematical reasoning. As it was impossible for us to code for both aspects (productive and problematic) together, we split this task into interpreting productive reasoning and interpreting problematic reasoning to arrive at two codes: Ia and Ib.

coding scheme to score participants' responses to students' incorrect explanations and the correct one. The only scheme that was slightly different was the one for interpreting problematic reasoning in the correct explanation where the participants received a score of 0 if they did not find anything problematic in the student's explanation. The maximum total score that each participant could receive for each of the explanations was 12.

To establish the interrater reliability for the scoring, three experts independently scored responses from six participants for one of the given student answers [answer (b)]. At first, the level of agreement on the scores was relatively low. After clarifying the criteria for the scores for all of the codes, the experts again scored another set of responses from another six participants, but this time for the explanation of answer (a), and achieved a much better agreement. We achieved interrater reliability of 90% or greater after several iterations of revising the criteria for the scores for all of the codes. One of the authors then scored all of the participants' responses again.

A caveat is in order. The participants wrote a lot of different things for codes Ia and Ib (interpreting productive and problematic reasoning, respectively) and to establish reliability for our coding schemes we decided that for the participants to receive a score of 2 for the above-mentioned codes, they had to write all of the things from the experts' list of productive and problematic aspects of student reasoning. To receive a score of 2 for code II (providing descriptive feedback) the participants had to give feedback for at least one correct and at least one incorrect or problematic element from the student's explanation. Since quantifying the levels of metacognition is very complicated, we decided to score code III only with 0 (no signs of engaging in metacognition) or 2 (signs of engaging in metacognition were present). To receive a score of 2, the participants had to have at least one question or statement in their response to the student that could be interpreted as metacognitive. As for the codes IVa and IVb (building on productive and addressing problematic elements, respectively)—to receive a score of 2, the participants had to explicitly build on at least *one* productive element in the student's explanation that they have identified and to address *all* of the problematic elements that they identified in the student's explanation, respectively. The participants were not "punished" with lower scores if they did not address all problematic elements that experts identified in the student's explanation. For example, let us say that the experts identified five problematic elements in the student's explanation. If the participant identified two problematic aspects but addressed both of them in their response to the student, then the participant would receive a score of 2 for addressing problematic elements. Examples of participants' coded responses are in Appendix B.

V. FINDINGS

As we mentioned before, we scored participants' responses based on the six codes. In this section we first show our findings of all participants' interpretations and responses to students' explanations, and then we show our findings for the interpretation and responses to students' explanations for each group of participants.

A. Participants' interpretation and responses to incorrect explanations

In this section, we describe the results of scoring participants' interpretation and responses to students' incorrect explanations (as we said above, two explanations were for the wrong answer choices and one incorrect explanation was for the correct answer choice). In Fig. 2, the vertical bars represent the scores that all participants received for specific codes for their responses to all incorrect explanations combined. There were three incorrect explanations in our survey and we used responses from 66 participants (a total of 198 responses for all incorrect explanations). To calculate the percentage of the total number of responses that, for example, received a score of 1, we divided the number of responses that received a score of 1 with 198 and multiplied it by 100. We did this procedure for each of the six codes.

From the histogram in Fig. 2, it is evident that the majority of participants' responses received a score of 1 for *interpreting productive and problematic reasoning* (codes Ia and Ib) for all incorrect explanations. This indicates that the participants mentioned some productive and some problematic aspects but not all aspects that the experts identified as important in the students' explanations. Receiving a score of 1 is a very good result because it is virtually impossible for one person to mention everything that the experts identified by working together.

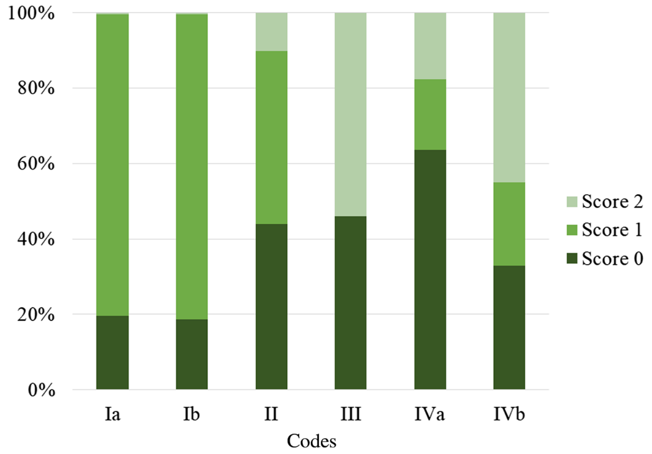


FIG. 2. The distribution of the scores for each of the codes for all participants' responses to all incorrect explanations combined.

When participants *interpreted productive reasoning* (code Ia), they mainly emphasized that the student recognized the constancy or conservation of momentum and/or energy. The participants mentioned the choice of the system only when it was explicitly mentioned in the student's explanation. It should be noted that when analyzing a situation from the momentum and/or energy perspective, it is crucial to always begin by selecting a system of interest. Another noteworthy observation is that most participants did not mention limiting cases at all, whether in terms of strengths or weaknesses of the student's response even though there was evidence of limiting case analysis in the students' explanations. This finding is particularly intriguing because the problem itself explicitly states, "Evaluate the graphs by analyzing limiting cases".

When participants *interpreted problematic reasoning* (code Ib), they mainly noted that the student failed to mention energy or the spring in their explanation. Additionally, participants successfully noted that the student did not realize that their explanation contradicted the conservation of momentum or energy. Finally, participants highlighted that the student incorrectly assumed that the speed of car 1 remains constant.

The majority of the participants' responses received a score of 0 or 1 for *providing descriptive feedback* to students (code II). Score 0 means that, in their responses, the participants did not provide feedback for either correct or the incorrect/problematic aspects of the student's reasoning or that the participants wrote something that could not be interpreted as feedback. Receiving a score of 1 means that the participants provided somewhat deficient descriptive feedback to the students, which means that the participants provided feedback only on one of the aspects of student reasoning (correct or problematic) but not on both. For the correct aspects of the student's reasoning, participants focused on ideas such as momentum and/or energy and constancy or conservation of momentum and/or energy. For the incorrect aspects of the student's reasoning, participants primarily emphasized that the speed of car 2 could not be infinite (which would contradict the conservation of energy), that the speed of car 1 is not constant and that the smaller mass of car 1 does not mean smaller (kinetic) energy of that car (which would violate the constancy of momentum).

As can be seen in Fig. 2, slightly more than half of the participants' responses received a score of 2 for *engaging the student in metacognition* (code III). This indicates that the participants' responses prompted the students to reflect on their thinking processes and that the participants' responses reflected the students' thoughts and considerations.

The majority of participants' responses received a score of 0 when it came to *building on productive elements* (code IVa) found in the students' explanations. In the rare cases when the participants built on the productive elements, they typically built (implicitly or explicitly) on the conservation laws, specifically the conservation of momentum, as well as the student's understanding of momentum. On the other hand, Fig. 2 shows that almost half of the participants' responses received a score of 2 for *addressing problematic elements* (code IVb). They acknowledged and dealt with all of the problematic elements that they had identified in students' explanations. When evaluating how the participants addressed those problematic elements, we found that they primarily focused on the fact that the student did not mention energy or the spring in their explanation, the student's incorrect idea that the speed of car 2 can be infinitely large, the student's wrong assumption that the speed of car 1 is constant, and the student's incorrect assertion that the smaller mass of car 1 meant a smaller (kinetic) energy of that car.

B. Participants' interpretations and responses to the correct explanation

In this section, we describe the results of scoring participants' interpretations and responses to the student's correct explanation. In Fig. 3 the vertical bars represent the scores that all participants received for specific codes for their responses to the correct explanation. There was one correct explanation in our survey and we collected responses from 66 participants (a total of 66 responses for the correct explanation). To calculate the percentage of the number of responses which, for example received a score of 1, we divided the number of responses which received a score of 1 with 66 and multiplied it by 100. We did this procedure for each of the six codes.

From Fig. 3, it is evident that the majority of participants' responses received a score of 1 when it came to *interpreting productive reasoning* (code Ia) in the student's correct explanation meaning that the participants mentioned some productive aspects but not all aspects that the experts identified in the student's explanation. While interpreting

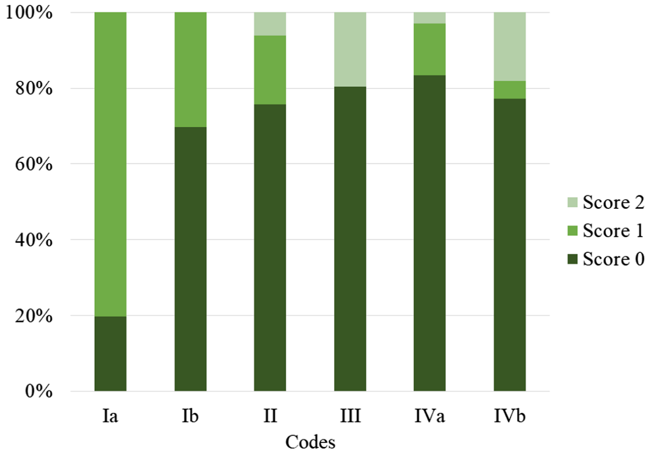


FIG. 3. The distribution of the scores for each of the codes for all participants' responses to the correct explanation.

productive reasoning, participants typically acknowledged that the student realized that both total momentum and total (mechanical) energy of the system were constant, and some participants mentioned that the student correctly analyzed the limiting cases.

On the other hand, the majority of participants' responses received a score of 0 when it came to *interpreting problematic reasoning* (code Ib) which means that, in their responses, participants did not mention any of the problematic aspects that the experts identified in the student's explanation. A smaller number of participants, who mentioned some problematic things, focused on the fact that the student did not explain why the smaller mass of car 1 results in a larger transfer of energy from the spring to that car.

Figure 3 also shows that the majority of participants' responses received a score of 0 for *providing descriptive feedback* (code II) as they did not provide feedback for neither correct nor the problematic aspects of the student's reasoning. A few who provided somewhat good feedback mainly focused on the correct aspects of the student's reasoning. Their feedback included ideas such as constancy/conservation of momentum and energy, and that the student correctly stated that a smaller mass of car 1 results in a larger (kinetic) energy of that car. The majority of participants did not identify any problematic aspects in the student's explanation. Instead, the participants predominantly praised the student for their comprehensive explanation (e.g., "Excellent work." or "Well done!").

The majority of participants' responses received a score of 0 for *engaging the student in metacognition* (code III), meaning that the participants did not prompt the student to reflect on their thinking. Those who demonstrated metacognitive responses focused on making the student think about how they know that the smaller or larger mass of car 1 results in the larger or smaller (kinetic) energy of car 1.

The majority of participants' responses received a score of 0 for *building on productive elements* (code IVa) found in the student's explanation, meaning that the participants did not build on the productive ideas that they identified in the student's explanation. However, in the responses from a small number of participants, we saw evidence of building on the student's productive ideas. In particular, they usually built on the correct statement that the smaller/larger mass of car 1 results in the larger/smaller (kinetic) energy of that car.

We found that the majority of participants' responses received a score of 0 for *addressing problematic elements* (code IVb) in the student's explanation as they either provided irrelevant or incorrect responses (e.g., "Take friction into account.") or they did not find anything problematic in the student's explanation. A few of those who identified some problematic elements focused their attention on the fact that the student did not explain why the smaller or larger mass of car 1 results in a larger or smaller (kinetic) energy of that car, and that, although the student mentioned that momentum is constant, they did not incorporate that into their explanation.

C. Similarities and differences in participant groups interpreting and responding to students' explanations

In this section, we describe the results of scoring participant groups' interpretations and responses to students' explanations. We compared the results among the six groups of participants. Three of these groups were in-service teachers. We did not group all in-service teachers together and all pre-service teachers together in the findings as these groups turned out to be somewhat different. As we said in Sec. III, we have limited information about the in-service teacher participants. We know that the international in-service teachers participate in daily ISLE-based professional development discussions and monthly workshops through the Facebook group and in-service teachers A participate in monthly professional development workshops, a few of which are ISLE-oriented. We do not have information about the professional development activities of in-service teachers B. Therefore, we could not group all in-service teachers together. Similarly, we could not combine all pre-service teachers together. While both pre-service teachers A enrolled in the ISLE-based physics teaching methods courses, their exposure to these methods was different. Pre-service teachers B had minimal exposure to physics teaching method courses and none of it was ISLE based.

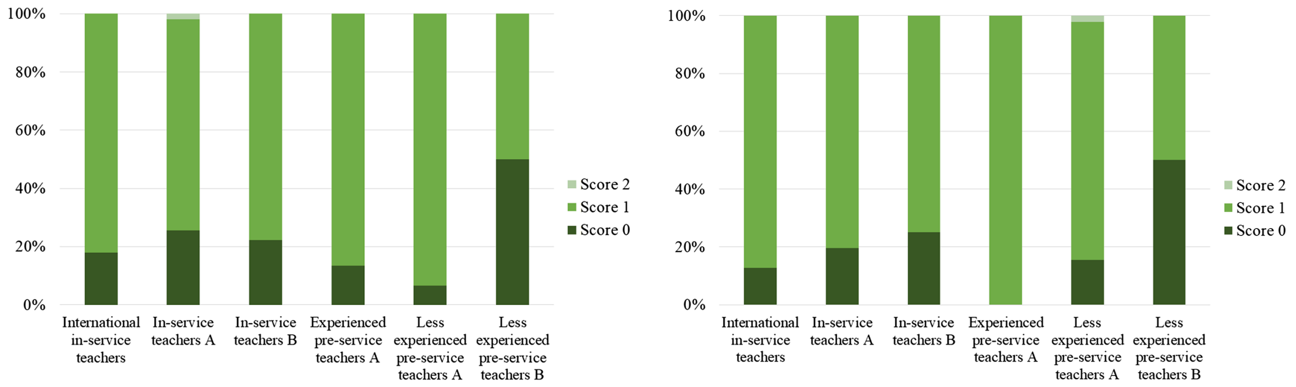
For incorrect explanations, the results of the scoring for each group of participants are shown in Figs. 4–6. We counted how many responses from each group of participants received a certain score (for all three incorrect explanations) and divided that number by the number of responses from a certain group multiplied by 3, and multiplied by 100 to get a percentage. For example, we had 13 participants in the international in-service teachers' group. In total, that was 39 (13×3) responses for all 3 incorrect explanations. Let us say that there were 12 responses out of those 39 that received a score of 1. To calculate the percentage of responses from the international in-service teachers who received a score of 1, we divided 12 by 39 and multiplied it by 100.

For the correct explanation, the results of the scoring for each group of participants are shown in Figs. 7–9. We counted how many responses from each group of participants received a certain score, divided that number by the total number of responses from a certain group, and multiplied by 100 to get a percentage.

We first show the results of scoring for each group of participants for students' incorrect explanations and then for the student's correct explanations.

1. Interpreting productive and problematic reasoning in incorrect explanations

Figure 4 shows the distribution of the scores for interpreting productive and problematic reasoning for



(a) (b)

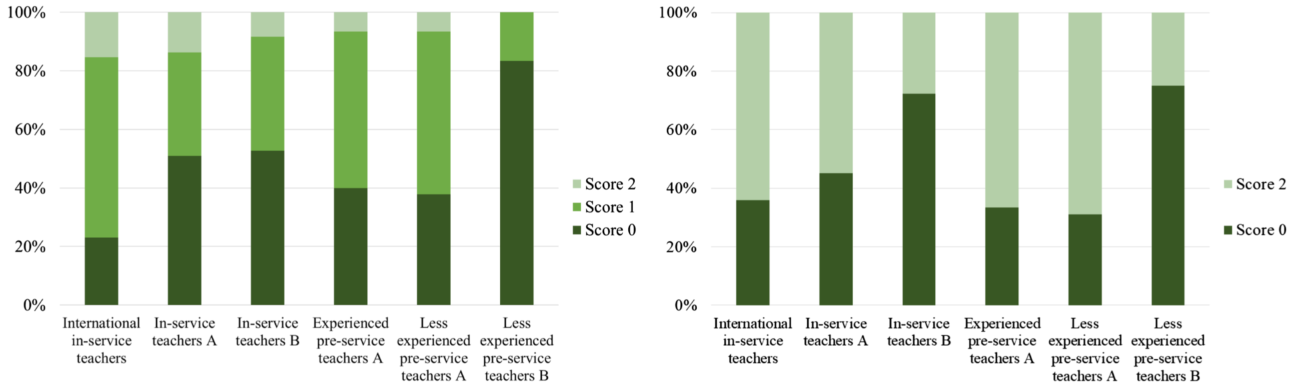
FIG. 4. The distribution of the scores for interpreting productive (a) and problematic (b) reasoning for students' incorrect explanations combined. Statistical significance of the differences is analyzed in Appendix C.

students' incorrect explanations combined. Figures 4(a) and 4(b) show similarities in the responses for the first five groups for interpreting productive and problematic reasoning as the majority received a score of 1. The participants mentioned some productive and problematic aspects but not all aspects that the experts identified in the students' explanations. The only group that was somewhat weaker was the group of less experienced pre-service teachers B whose responses equally received a score of 0 and 1. However, even they were moderately successful.

Notice that none of the experienced pre-service teachers A received a score of 0 for interpreting problematic reasoning, demonstrating better performance than all of the other groups. It is also worth mentioning that pre-service teachers A (both groups) show similar, and at times better results for interpreting productive and problematic reasoning compared to in-service teachers in this survey.

2. *Providing descriptive feedback and engaging in metacognition for incorrect explanations*

Figure 5 shows the distribution of the scores for providing descriptive feedback and engaging students in metacognition for students' incorrect explanations combined. Figure 5(a) shows notable differences in the percentages of responses of individual groups for each of the scores for providing descriptive feedback. Here, the responses of international in-service teachers received more scores 1 and 2 than other groups. Score 1 means that the participants provided feedback only on one of the aspects of student reasoning, correct or incorrect/problematic, but not on both, while score 2 means that they provided feedback for both of those aspects. Both groups of pre-service teachers A have practically identical distribution of scores. The majority of their responses received a score of 1. It is also worth mentioning that the percentages of responses from both groups of pre-service teachers A who received a score of 1 are comparable to the percentage of responses of international in-service teachers who received a score of 1 for providing descriptive feedback. While in-service teachers A and in-service teachers B also show a very similar distribution of scores, their score distribution is different from international in-service teachers and pre-service teachers A as their responses received more scores of 0. This means that, in their responses, the participants did



(a) (b)

FIG. 5. The distribution of the scores for providing descriptive feedback (a) and engaging students in metacognition (b) for students' incorrect explanations combined. Statistical significance of the differences is analyzed in Appendix C.

not provide feedback for either correct or the incorrect/problematic aspects of the student's reasoning or that the participants wrote something that could not be interpreted as feedback. Less experienced pre-service teachers B showed rather poor responses as the majority of their responses received a score of 0.

Figure 5(b) shows that international in-service teachers, in-service teachers A and both groups of pre-service teachers A are practically indistinguishable when it comes to engaging the student in metacognition. Those groups of participants mostly received a score of 2 which means that their questions and/or comments would engage students in metacognition. Likewise, there are no major differences between in-service teachers B and less experienced pre-service teachers B as the percentages of scores for their responses are almost identical. Their responses mostly received a score of 0 indicating that there were no metacognitive questions and/or comments in their responses to students.

3. Building on productive and addressing problematic elements in incorrect explanations

Figure 6 shows the distribution of the scores for building on productive and addressing problematic elements for students' incorrect explanations combined. From Fig. 6(a) we can see that all participants have difficulties building on students' productive ideas that they identified before. While the majority of participants' responses received a score of 1 for interpreting productive ideas in students' explanations, they were much less successful when building on those productive ideas. Even in the best performing groups, experienced and less experienced pre-service teachers A, only about 50% of their responses received scores of 1 and 2, while the rest received the score of 0. All groups of in-service teachers performed worse, receiving a score of 1 for over 60% of their responses, and the less experienced pre-service teachers B received mostly a score of 0 for their responses. Receiving a score of 0 for building on productive elements indicates that the participants did not build on students' productive ideas that they identified.

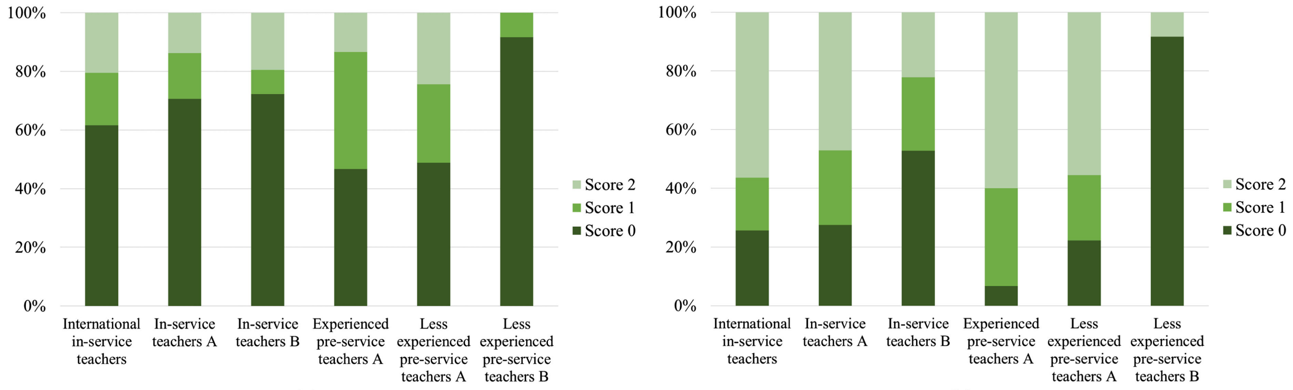
Though there were no statistically significant differences between the groups for either of the scores, we must point out that responses from both groups of pre-service teachers A received a score of 1 more often than responses from other groups of participants. This means that they were building on students' productive ideas but only implicitly instead of explicitly which would render them a score of 2.

The situation was much better for addressing problematic elements in students' explanations as many participants' responses received a score of 2 for addressing problematic elements [see Fig. 6(b)]. The score of 2 means that the participants addressed all of the problematic ideas they found in students' explanations.

Figure 6(b) shows notable differences in the percentages of responses of individual groups for each of the scores for addressing problematic elements. Here, experienced pre-service teachers A demonstrated better performance than international in-service teachers, in-service teachers A and less experienced pre-service teachers A as their responses received more scores 1 and 2. Receiving a score of 1 or 2 for addressing problematic elements means that the participants addressed some or all of the problematic ideas found in students' explanations, respectively. In-service teachers B and less experienced pre-service teachers B showed rather poor responses having mostly received a score of 0 for their responses.

4. Interpreting productive and problematic reasoning in the correct explanation

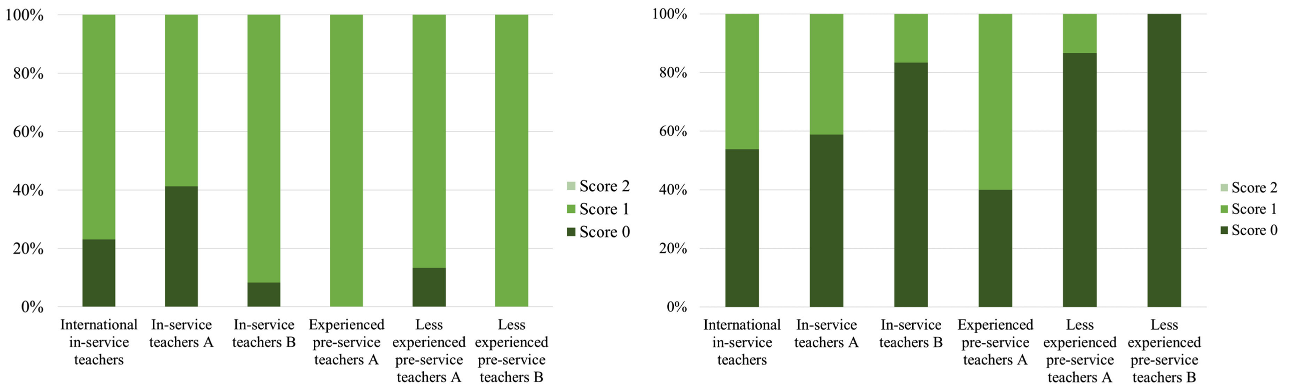
Figure 7 shows the distribution of the scores for interpreting productive and problematic reasoning for the student's correct explanation. Figure 7(a) shows that all of the groups mostly received a score of 1. We can also see that less experienced pre-service teachers A and less experienced pre-service teachers B did not receive any score of 0. All of their responses received a score of 1 for



(a)

(b)

FIG. 6. The distribution of the scores for building on productive (a) and addressing problematic elements (b) for students' incorrect explanations combined. Statistical significance of the differences is analyzed in Appendix C.



(a) (b)

FIG. 7. The distribution of the scores for interpreting productive (a) and problematic (b) reasoning for the student's correct explanation.

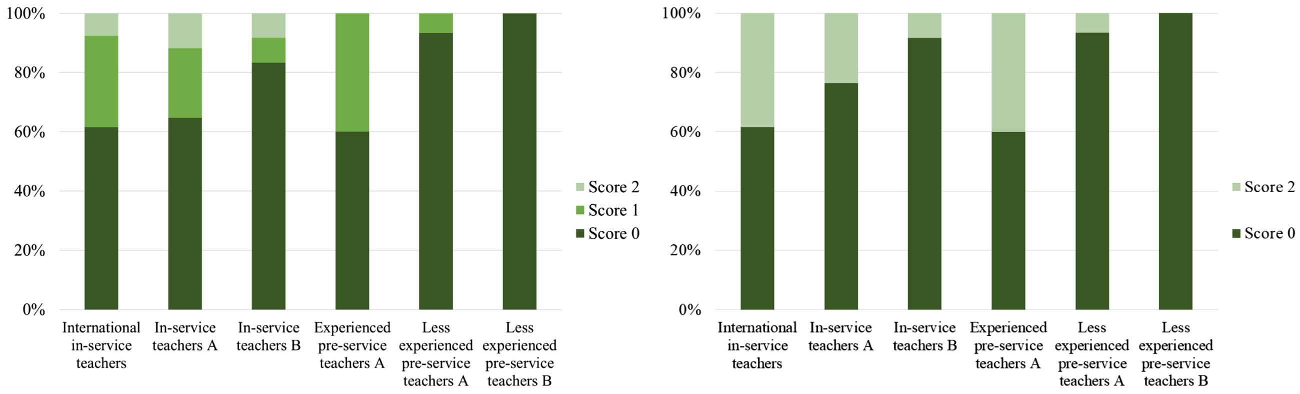
interpreting productive reasoning indicating that the participants mentioned some productive aspects but not all aspects that the experts identified in the student's explanation. International in-service teachers and in-service teachers A received a score of 0 more than other groups. This result might lead someone to conclude that those groups performed worse than others. However, due to our way of scoring, it only appears that they did not interpret the student's productive ideas when in fact they just did not explicitly state the student's productive ideas. They would usually say that "The answer contains everything you would expect from a student" or "I like everything that the student states."

Figure 7(b) shows notable differences in the percentages of responses of individual groups for each of the scores for interpreting problematic reasoning. Experienced pre-service teachers A showed better performance than other groups. Their responses mostly received a score of 1 which means that they found some problematic aspects in the student's explanation. All other groups mostly received a score of 0 for interpreting problematic reasoning which means that they did not find anything problematic in the student's explanation. The distribution of the scores is similar for international in-service teachers and in-service teachers A, as well as for in-service teachers B and less experienced pre-service teachers A. Less experienced pre-service teachers B showed poor results as all of their responses received a score of 0.

Despite the visible differences among the groups, statistical analysis showed no significant differences between the groups for interpreting productive nor for interpreting problematic reasoning.

5. *Providing descriptive feedback and engaging in metacognition for the correct explanation*

Figure 8 shows the distribution of the scores for providing descriptive feedback and engaging students in metacognition for the student's correct explanation. Figure 8(a) shows notable differences in the percentages of responses of individual groups for each of the scores. Only in-service teacher groups received a score of 2 for some of their responses which means that they gave feedback for the correct and for some problematic aspects



(a) (b)

FIG. 8. The distribution of the scores for providing descriptive feedback (a) and engaging students in metacognition (b) for the student's correct explanation.

of the student's reasoning. From the pre-service teacher groups, the percentage of responses that received a score of 1 is highest for the experienced pre-service teachers A. However, Fig. 8(a) also shows that all of the groups mostly received a score of 0 for providing descriptive feedback to the student indicating that in the majority of their responses there was no feedback for the correct nor the problematic aspects of the student's reasoning. The participants mostly praised the student for their thorough explanation.

Figure 8(b) shows notable differences in the percentages of responses of individual groups for each of the scores for engaging the student in metacognition. Here, international in-service teachers, in-service teachers A and experienced pre-service teachers A showed better results than other groups as their responses received more scores 2. This means that their questions and/or comments would engage the student in metacognition. One should have in mind that those are the same groups that showed better results for interpreting problematic reasoning. However, Fig. 8(b) also shows that all of the groups mostly received a score of 0 which means that there were no metacognitive questions and/or comments in their responses to the student.

Even though there were visible differences among the groups, statistical analysis showed no significant differences between the groups in providing descriptive feedback or engaging the student in metacognition.

6. Building on productive and addressing problematic elements in the correct explanation

Figure 9 shows the distribution of the scores for building on productive and addressing problematic elements for the student's correct explanation. Figure 9(a) shows a very similar distribution of scores for international in-service teachers, in-service teachers A, in-service teachers B, and experienced pre-service teachers A. Out of those groups, international in-service teachers and in-service teachers A are the only groups where some of their responses received a score of 2 for building on productive elements which means that the participants from those groups explicitly built on found productive ideas. Though some of the responses from the first five groups received a score of 1 or even 2, Fig. 9(a) shows that all of the groups mostly received a score of 0 indicating that there was no evidence of either implicit or explicit building on student's productive ideas.

Statistical analysis showed no significant differences between the groups for building on productive elements.

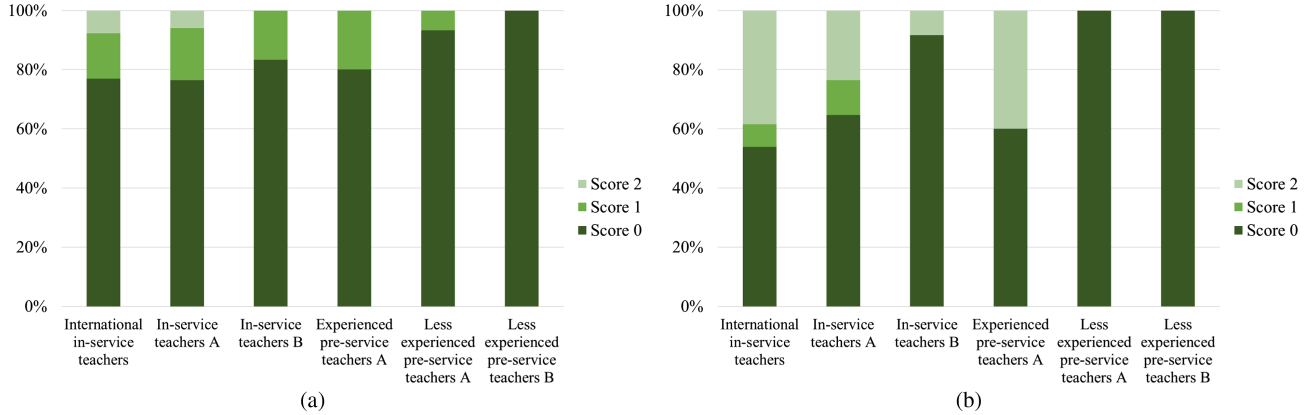
Figure 9(b) shows visible differences in the percentages of responses of individual groups for each of the scores for addressing problematic elements. Here, international in-service teachers, in-service teachers A, and experienced pre-service teachers A demonstrated better performance than other groups as their responses received more scores 1 and 2. We must point out that these groups are the same ones whose responses received more scores 1 for interpreting problematic reasoning and scores 2 for engaging the student in metacognition. Receiving a score of 1 or 2 for addressing problematic elements means that the participants addressed some or all of the problematic ideas found in students' explanations, respectively. Less experienced pre-service teachers A and less experienced pre-service teachers B show poor performance having received only a score of 0 for their responses.

VI. DISCUSSION

With this study we sought to investigate how in-service and pre-service physics teachers interpret and respond to students' incorrect and correct explanations. We will refer to them as participants, except when we want to highlight the differences between the in-service and pre-service physics teachers.

A. Research question 1

The first research question we set out to answer was "How do teachers interpret and respond to students' incorrect explanations?"



(a) (b)

FIG. 9. The distribution of the scores for building on productive (a) and addressing problematic elements (b) for the student's correct explanation. Statistical significance of the differences is analyzed in Appendix C.

1. Interpreting productive and problematic reasoning

We found that the participants were successful in identifying productive and problematic aspects of student reasoning. Although they did not mention all productive and problematic aspects that the experts identified as important, we need to remember that the experts worked as a group and the participants worked individually. Therefore, we can interpret this finding as very positive. When interpreting productive and problematic aspects of students' reasoning, the participants usually focused on the momentum and energy aspects. A possible explanation for this result is that the participants recognized this problem as a "momentum and energy" problem and focused on those aspects in the students' explanations.

Although the majority of the participants focused on the physics aspects when interpreting students' reasoning, and did it rather well, they usually did not mention the limiting case analysis and the choice of the system in the students' explanations. Not mentioning the limiting case analysis is especially interesting because there is an explicit instruction in the problem to evaluate the graphs by analyzing limiting cases. It is possible that as the participants recognized the problem as a "momentum and energy" problem, they focused their interpretations of students' strengths and weaknesses primarily on these aspects, missing limiting case analysis. It is also possible that the participants may not have been familiar with utilizing limiting case analysis as a problem-solving approach or were not aware of its benefits for student learning (see research done by Warren [77]). Research also shows that students need explicit instruction to develop expertise in evaluating limiting cases [78].

Another finding was that unless it was explicitly specified in the student's explanation, the majority of the participants did not mention the choice of the system. This finding aligns with previous research conducted by Seeley *et al.* [31] where they reported that teachers have difficulties in identifying the implied system in the student's response.

Our participants' interpretation of productive and problematic reasoning is consistent with previous research on teachers' interpretation of students' work [11,14]. Similar to Ebby *et al.* [14] who found that teachers focused primarily on the strategies that the students used to solve math problems, our participants focused on the ideas of momentum conservation that students were using to solve the problem and on the missing energy conservation that the students did not bring up. The latter finding is consistent with the work of Kristinsdóttir *et al.* [11] who found that teachers looked primarily for the things that were missing in the students' work (they studied trigonometry problems).

However, it is important to note that our analysis was confined to the information provided in the participants' responses, and we can only reflect on the content they wrote. The participants may not have mentioned all of the productive and problematic elements that they observed in the students' explanations, either because they considered them as details that were not so important or due to a lack of motivation to write them. Unfortunately, due to the nature of the study's format, we were unable to investigate this particular matter deeper.

2. Providing descriptive feedback and engaging in metacognition

When analyzing participants' responses, we focused on the type of feedback (if any) that the participants provided to the students and whether there were indications of engaging students in metacognition.

As we said in Sec. V, the majority of the participants provided somewhat deficient feedback as their feedback did not usually address both the correct and incorrect aspects of the student's reasoning, being mostly focused on the incorrect aspects of the student's reasoning. This finding is consistent with the research conducted by Kristinsdóttir *et al.* [11] who found that teachers tend to focus on students' mistakes and what is missing in their work. A possible explanation is that the participants may not have been familiar with what constitutes effective feedback and therefore they usually omitted providing feedback on the productive things (for more details on effective feedback, see Refs. [61,62]) which is very important because once students know exactly what was correct in their explanation, they can use the same strategy when solving other problems.

We found that slightly more than half of the participants' responses to students' incorrect explanations encouraged students to think about their thinking through the use of metacognitive questions and comments. The participants tried to assist the students in figuring out the right answer on their own instead of offering them the correct answer and explanation. In their responses, the participants encouraged students to explain in more detail their thinking which is consistent with Ader's work [68] where he found that teachers primarily fostered students' metacognitive skills by encouraging them to articulate and elaborate on their ideas and responses.

3. Building on productive and addressing problematic elements

In our research, we sought to identify how the participants were building on the productive elements that they found in students' explanations and how they were addressing problematic ones.

Although RT, AST, and the ISLE approach all emphasize the importance of building on productive aspects of students' ideas, our findings suggest that participants still encounter challenges in effectively building on students' existing ideas. The cases where participants built on productive elements in students' explanations were rather rare. The participants usually built (implicitly or explicitly) on the conservation laws, specifically the conservation of momentum, as well as the student's understanding of momentum, even when the student did not explicitly say that they were using conservation of momentum. It is possible that the participants recognized the *conservation* p-prim in students' explanations even when it was implicit in students' work.

We found that the participants were more successful in addressing perceived problematic elements than in building on students' productive ideas (similar to the results reported by Kristinsdóttir *et al.* [11]. It is possible that the participants recognized several p-prims and resources (denoted in italics) in students' explanations which were not suitable for the situation at hand. Therefore, participants addressed those problematic or incorrect elements in their responses to students. We found that the participants primarily focused on the fact that students did not mention energy or the spring in their explanation (*conservation p-prim* [47]), the student's incorrect idea of infinitely large speed of car 2 (*Ohm's p-prim*), the student's wrong assumption about constant speed of car 1 (*constancy p-prim*) and the student's incorrect assertion that the smaller mass of car 1 meant a smaller (kinetic) energy of that car (*less cause means less effect p-prim*).

Consistent with the feed forward concept, Responsive Teaching and Ambitious Science Teaching, our participants showed that they were responsive to the students' needs. While the majority of the participants addressed problematic things they found, a few of them also built on students' productive ideas. By implicitly building on the concept of conservation of momentum by mentioning another crucial concept of conservation, conservation of energy, that was missing in some students' explanations, our participants guided students to improve their work. In this way, participants provided students with strategies and resources that would enhance their future work.

B. Research question 2

The second research question we set out to answer was "How do teachers interpret and respond to students' correct explanations?"

1. Interpreting productive and problematic reasoning

Similar to the interpretation of incorrect explanations, we found that the participants were successful in identifying productive aspects of students' reasoning.

Most of the participants acknowledged that the student realized that both total momentum and total (mechanical) energy of the system were constant, and some participants also mentioned that the student correctly analyzed the limiting cases. This is an excellent result, but we should have in mind that this student explicitly mentioned momentum and energy in their explanation.

However, even in the correct and very thorough student's explanation the majority of the participants did not mention the student's implicit choice of the system. This result aligns again with the previous research conducted by Seeley *et al.* [31] where they reported that teachers have difficulties in identifying the system that is implied in the student's response.

We found that the majority of the participants did not identify any problematic aspects in the correct explanation. It is possible that once the participants recognized that the student's answer and explanation were correct, they did not thoroughly examine the finer details of the explanation that could reveal potential problematic elements. This result aligns with the research conducted by Gotwals and Birmingham [10] who found that teachers seldom look for nuances in students' ideas. A smaller number of participants, who mentioned some problematic things, stated that the student did not explain why the smaller mass of car 1 results in a larger transfer of energy from the spring to that car. This is consistent with the work of Kristinsdóttir *et al.* [11] who found that teachers tend to focus on what is missing in students' work.

It is possible that the participants may not have mentioned all of the productive and problematic elements they observed in the students' explanations because they considered them too obvious, minor details, or simply because they were not motivated to write them. Unfortunately, due to the nature of our study, we were not able to investigate this particular matter deeper.

2. Providing descriptive feedback and engaging in metacognition

We found that the majority of participants did not provide feedback for either the correct or the problematic aspects of the student's reasoning when the student's explanation was correct. Instead of providing specific and constructive feedback, the majority of participants wrote some sort of generic praise, such as "Well done!" or "Excellent work!". It is possible that the participants omitted providing feedback on the productive things because they were not familiar with what constitutes effective feedback and why it is important to tell students what they are doing correctly. Providing feedback on the correct aspects is important for students' learning because once they know exactly what was correct in their explanation, they can use the same strategy when solving other problems.

Since the student's explanation was indeed correct, there were no clear incorrect or problematic elements for the teachers to address. However, there were some minor problematic aspects that the participants did not mention. It is possible that they did not thoroughly examine the details of the explanation that could reveal potential problematic elements once they recognized that the student's answer and explanation were correct. This is again consistent with the research conducted by Gotwals and Birmingham [10], who found that teachers rarely look at nuances in students' ideas.

As the majority of participants did not identify any problematic aspects in the correct explanation, it is not surprising that there were not many metacognitive questions in the participants' responses. However, it is important to note that those few participants who mentioned some problematic aspects addressed them by posing metacognitive questions and asking the student to articulate his reasoning in more detail. This is consistent with the findings of Ader's work [68] that teachers primarily foster students' metacognitive skills by encouraging them to articulate and elaborate on their ideas and responses.

3. Building on productive and addressing problematic elements

Similar to responding to incorrect explanations, we found that participants still have difficulties with effectively building on students' productive ideas. Only a handful of participants built (implicitly or explicitly) on the student's correct statement that the smaller or larger mass of car 1 results in larger or smaller (kinetic) energy of that car. When the participants identified many productive ideas, they used these productive ideas to push the students further.

Situation was slightly better for addressing problematic elements in the student's explanation. Remember that the majority of the participants found nothing problematic with the student's explanation. However, those few participants who found some problematic elements addressed them in their responses. They focused their attention on the fact that the student did not explain why the smaller or larger mass of car 1 results in a larger or smaller (kinetic) energy of that car, and that, although the student mentioned that momentum is constant, they did not incorporate that into their explanation.

C. Research question 3

The third research question we set out to answer was "What are the differences in responses to students' explanations among different groups of in-service and pre-service teachers?" We anticipated that in-service teachers, having prior experience working with students, would score higher on all codes than pre-service teachers in our study. Some of our findings were surprising.

For the majority of codes, for all four students' explanations, international in-service teachers showed better performance than in-service teachers A and B. International in-service teachers voluntarily participate in everyday discussions in the Facebook group "Exploring and Applying Physics" which was created to help teachers who use or plan to use the Investigative Science Learning Environment (ISLE) [70] approach in their classrooms. This means that these teachers represent dedicated and concerned educators. It is also possible that their knowledge of the ISLE approach due to the engagement in this group (not only discussions but also the workshops) and/or self-study (research papers are regularly posted in this group) contributed to their better results compared to the results of other groups of in-service teachers. This allows us to hypothesize that not only the teaching practice, but also participation in targeted professional development affects how in-service teachers respond to the students.

The surprising finding here is that the interpretation and responses to all students' explanations from pre-service teachers A (both groups) were often comparable, and at times better than the interpretation and responses of in-service teachers indicating that certain aspects of teachers' content knowledge for teaching can be developed through teacher preparation programs. Such aspects as providing effective feedback, engaging students in metacognition, identifying and building on students' productive ideas are at the heart of the ISLE approach which is the philosophical foundation of the pre-service teachers' A teacher preparation program. Participating in activities that attend to the above-mentioned aspects might have contributed to the pre-service teacher A's performance being comparable, and at times better than the performance of in-service teachers.

The differences between two groups of pre-service teachers A were minimal when they interpreted and responded to students' incorrect explanations. However, when it came to interpreting and responding to the student's correct explanation, experienced pre-service teachers A exhibited better performance in nearly all of the codes than less experienced pre-service teachers A. A possible explanation might lie in the fact that experienced pre-service teachers A had three semesters of the Didactics of Physics course based on the ISLE approach, while the less experienced group only had one semester. Additionally, experienced pre-service teachers A had clinical practice in ISLE-based settings. This longer engagement in the courses based on ISLE, combined with practicing the skills developed through coursework, might have contributed to their development of productive habits of providing feedback on both correct and incorrect responses (see Ref. [79]) and thus led to their better performance when the student's explanation was correct.

For all codes, both groups of pre-service teachers A scored higher than pre-service teachers B. The explanation might lie in the experiences of both groups prior to taking the survey, that is different levels of preparation. While all three groups had finished (or about to finish) their physics training and supposedly had similar knowledge of the physics required to solve the problem, their exposures to the content knowledge for teaching physics were different. Pre-service teachers A attended courses and participated in clinical practice where they had purposeful and systematic training in the content knowledge for teaching physics practicing the tasks of teaching that were used to evaluate their performance in this study, while pre-service teachers B attended only generic pedagogical courses prior where they had no training in content knowledge for teaching physics. This result clearly demonstrates that having pure content knowledge is not enough to help the students who are struggling with solving the problem. This is consistent with the work of Seeley *et al.* [31] who found that undergraduate physics majors who knew how to solve problems given to high school students, were not able to help those students who had difficulties solving a problem. Since all groups of pre-service teachers are familiar with the subject matter of the problem students were solving, it is possible that the differences in their results stem from the teacher preparation program.

D. Limitations of the study

It is important to acknowledge several limitations that should be considered when drawing conclusions from this study.

First, the relatively small number of participants restricts the generalizability of the findings. With a small sample size, there may be limited diversity and variability among the participants, which can reduce the robustness of the conclusions. Therefore, we should be careful when drawing conclusions about the differences between pre-service and in-service teachers based on our findings.

Second, all participants in the study were recruited using convenience sampling. This sampling method, based on availability and accessibility, may result in a sample that is not representative of the broader population and lacks diversity. Consequently, the generalizability of the findings may be limited.

Third, it is crucial to recognize that the study's findings are derived solely from the responses provided on the survey and do not provide insight into teachers' actual classroom practices.

Fourth, due to the anonymous nature of our survey, certain findings could not be explored in greater depth. For example, the reasons why very few participants mentioned all of the productive and/or problematic elements in student explanations could not be examined.

Fifth, to score participants' responses, we used numerical scores 0, 1, and 2. This kind of quantitative scoring, which was rather crude, sometimes prevented us from differentiating between satisfactory *(even though accurate, there are no explicit remarks based on physics)* and good *(the participant makes explicit remarks that are based on physics)* responses. It was a trade-off in which we lost fine grained analysis of our data but gained reliability for our coding schemes.

Finally, the instruction in our survey said: "For each student explanation please describe what you think the student's strengths and weaknesses are and how you would respond to this student if they were a student in your class." This means that the participants knew that their feedback would not be going directly to the students. Because of that, it is possible that some teachers interpreted the research task as an assessment of students' understanding rather than an evaluation of the quality of the feedback itself. This misinterpretation may have influenced the nature and depth of participants' responses. As we did not interview the participants after they completed the survey (which was anonymous), the lack of clarity in how teachers perceived the task and the potential impact on their feedback presents a limitation in our ability to draw precise conclusions about the quality and nature of the feedback provided.

E. Implications for instruction and future research

Although this research was conducted using a convenience sample of participants and has many limitations, there are several instructional implications for both pre-service and in-service teachers that can be derived from the findings of this study. Our findings offer valuable insights into the areas where teachers may encounter difficulties and emphasize the need for additional investigations and support in addressing these challenges.

First, if we wish that teachers would focus more on the productive ideas in students' reasoning, it would be useful to incorporate activities that attend to students' needs similar to the ones described in the paper into their preparation (both coursework and clinical practice) and professional development. Our findings indicate that teachers are more responsive to students' weaknesses than strengths. It is possible that by giving teachers more opportunities to respond to students' productive ideas, they become more consistent with building on those ideas, even when those ideas are hidden behind incorrect answers.

Second, to help teachers learn to provide effective feedback, it might be useful to give them student work that contains not just answers but also explanations and explicitly ask to focus on both problematic and productive aspects of student reasoning. Previous research showed that students perform better if their teachers believe that students' existing knowledge should be built on [64]. Similar research showed that students performed better when their teachers participated in professional development programs where they discussed the importance of building on students' existing knowledge, and where they worked on implementing this approach [80]. Future research can also help find the most efficient ways of incorporating such training in teacher preparation and professional development.

Third, while our study was about teachers' written responses to students' written work, we hypothesize that the above recommendations are valid for oral responses in real-time class interactions as well. We acknowledge that it is harder to provide feedback in real time—that is why the teachers need to develop specific habits of responding to the students [79]. Similarly to practicing responses to students' written work, the pre- and in-service teachers can practice responding to the students orally at the moment during microteaching episodes [73].

Like we mentioned above, because of the nature of our study, it is not clear how the participants interpreted the task of providing a response to the student. Future research could address this limitation by incorporating additional measures (such as interviews) to understand teachers' perceptions and motivations when participating in similar studies.

While we gathered data on how teachers interpret and respond to student ideas, we did not collect their actual classroom data. By "actual data," we mean observing how teachers provide oral feedback during lessons or written feedback on homework assignments and how this feedback affects subsequent student performance. To investigate how physics teachers interpret and respond to student ideas in authentic contexts, we would need video recordings of lessons and examples of students' written assignments to analyze how teachers provide feedback in these specific situations and how students respond to this feedback. Future studies could address this matter.

It could also be interesting to look for the relationships between teachers' reasoning about a certain question and their analysis of students' reasoning about the same question.

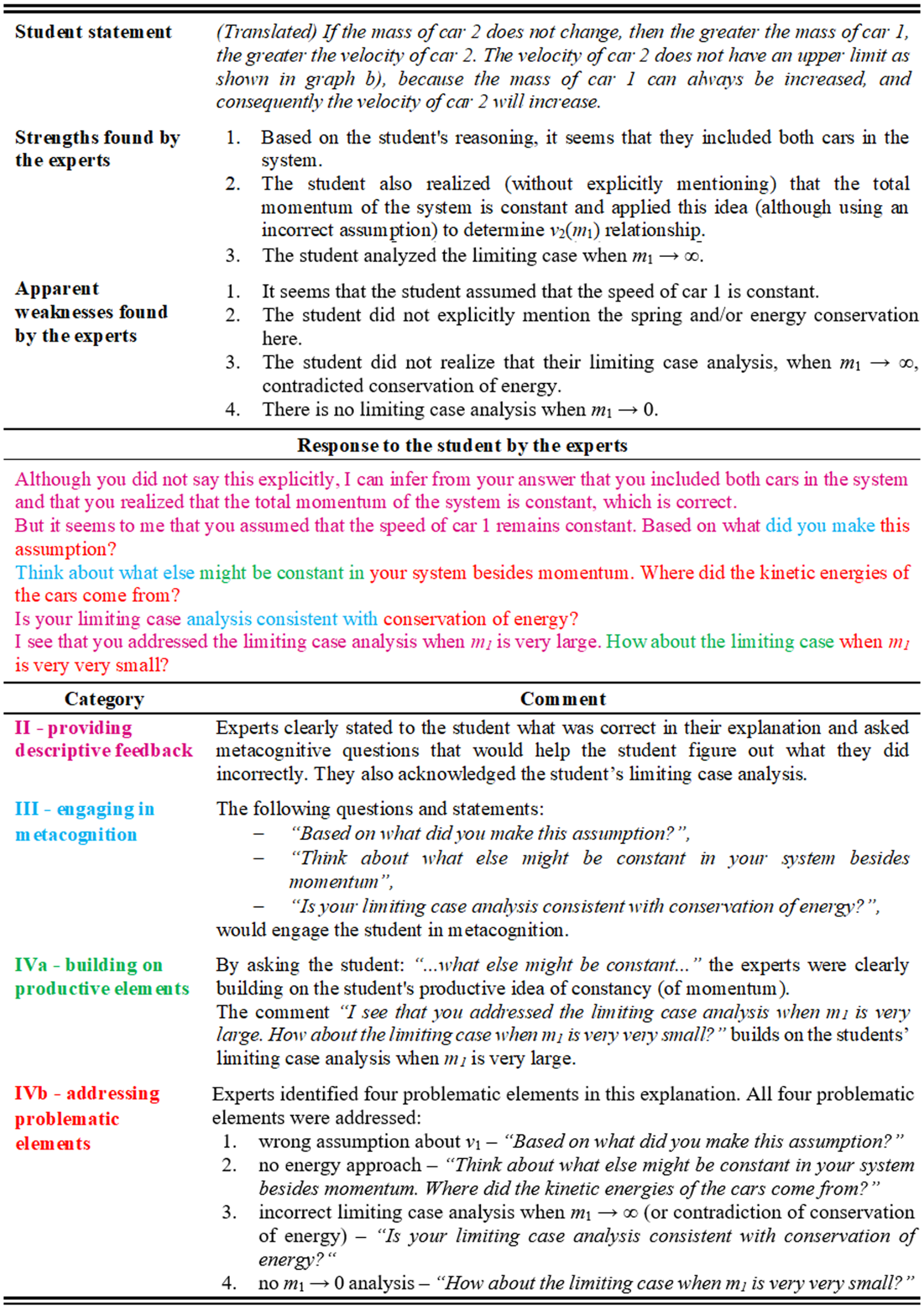
Future research on how teachers interpret and respond to student ideas in different contexts, such as experimental work or other types of problems, could also prove valuable. This broader exploration can provide deeper insights into the variations in teachers' responses based on the specific context or task.

ACKNOWLEDGMENTS

We are grateful to Professor Lane Seeley for his valuable feedback that led to the present version of this paper and to Ivan Dodlek and Professor Darko Dukić who helped with the statistical analysis. We also thank the three anonymous referees for their helpful comments and suggestions. This work was partly supported by the Slovenian Research Agency (Grant No. P1-0060).

APPENDIX A

In our study, we assumed that the experts identified all productive and problematic ideas in the student explanations. Here, in Figs. 10–13, we show, for each of the four students' explanations, which productive and problematic ideas experts mentioned and their response to the student broken down by codes. The response to the student is color-coded.



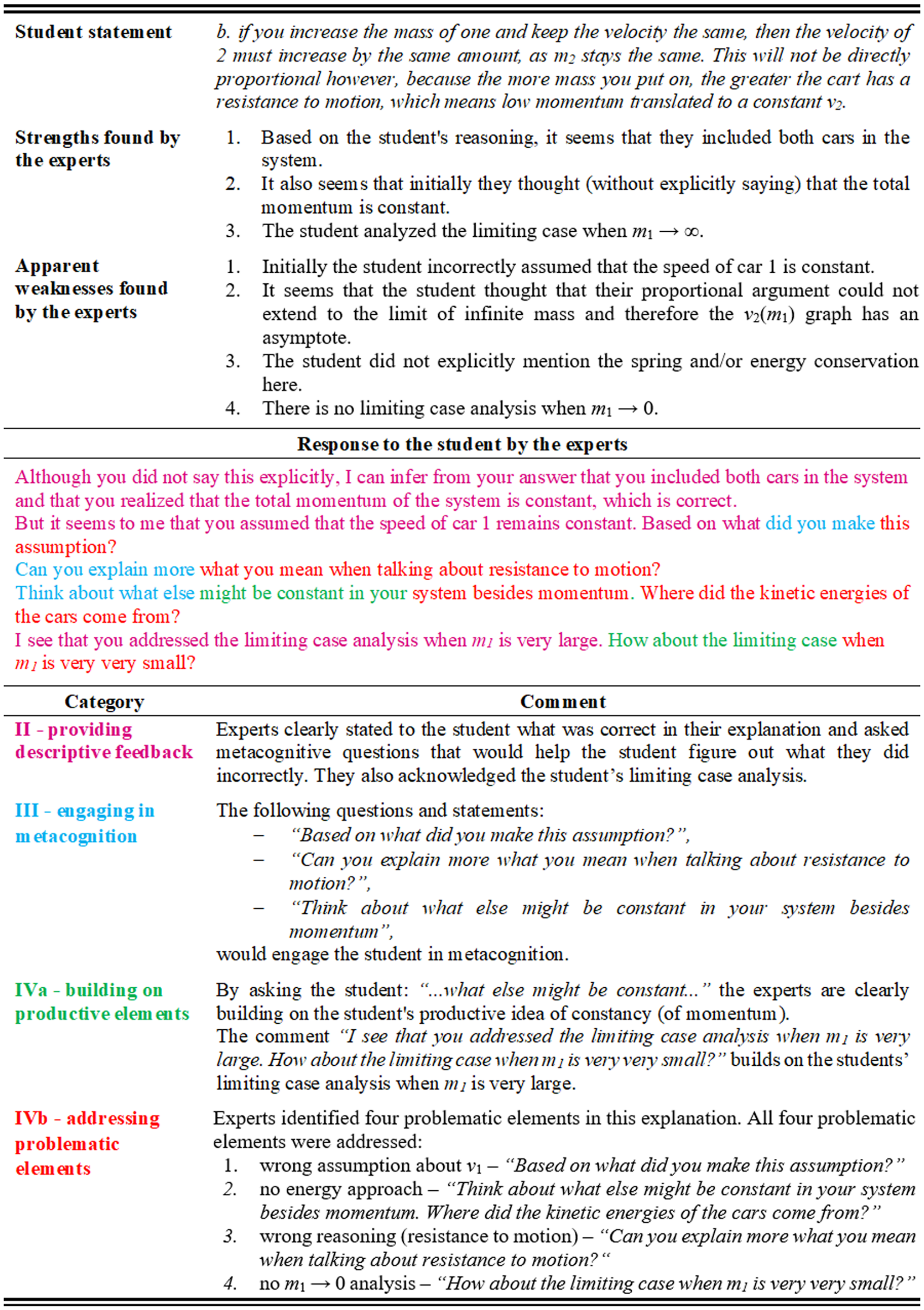
Student statement	*(Translated) If the mass of car 2 does not change, then the greater the mass of car 1, the greater the velocity of car 2. The velocity of car 2 does not have an upper limit as shown in graph b), because the mass of car 1 can always be increased, and consequently the velocity of car 2 will increase.*
Strengths found by the experts	1. Based on the student's reasoning, it seems that they included both cars in the system. 2. The student also realized (without explicitly mentioning) that the total momentum of the system is constant and applied this idea (although using an incorrect assumption) to determine $v_2(m_1)$ relationship. 3. The student analyzed the limiting case when $m_1 \rightarrow \infty$.
Apparent weaknesses found by the experts	1. It seems that the student assumed that the speed of car 1 is constant. 2. The student did not explicitly mention the spring and/or energy conservation here. 3. The student did not realize that their limiting case analysis, when $m_1 \rightarrow \infty$, contradicted conservation of energy. 4. There is no limiting case analysis when $m_1 \rightarrow 0$.

Response to the student by the experts

Although you did not say this explicitly, I can infer from your answer that you included both cars in the system and that you realized that the total momentum of the system is constant, which is correct.
But it seems to me that you assumed that the speed of car 1 remains constant. Based on what did you make this assumption?
Think about what else might be constant in your system besides momentum. Where did the kinetic energies of the cars come from?
Is your limiting case analysis consistent with conservation of energy?
I see that you addressed the limiting case analysis when m_1 is very large. How about the limiting case when m_1 is very very small?

Category	**Comment**
II - providing descriptive feedback	Experts clearly stated to the student what was correct in their explanation and asked metacognitive questions that would help the student figure out what they did incorrectly. They also acknowledged the student's limiting case analysis.
III - engaging in metacognition	The following questions and statements: – *"Based on what did you make this assumption?"*, – *"Think about what else might be constant in your system besides momentum"*, – *"Is your limiting case analysis consistent with conservation of energy?"*, would engage the student in metacognition.
IVa - building on productive elements	By asking the student: *"...what else might be constant..."* the experts were clearly building on the student's productive idea of constancy (of momentum). The comment *"I see that you addressed the limiting case analysis when m_1 is very large. How about the limiting case when m_1 is very very small?"* builds on the students' limiting case analysis when m_1 is very large.
IVb - addressing problematic elements	Experts identified four problematic elements in this explanation. All four problematic elements were addressed: 1. wrong assumption about v_1 – *"Based on what did you make this assumption?"* 2. no energy approach – *"Think about what else might be constant in your system besides momentum. Where did the kinetic energies of the cars come from?"* 3. incorrect limiting case analysis when $m_1 \rightarrow \infty$ (or contradiction of conservation of energy) – *"Is your limiting case analysis consistent with conservation of energy?"* 4. no $m_1 \rightarrow 0$ analysis – *"How about the limiting case when m_1 is very very small?"*

FIG. 10. All productive and problematic ideas experts identified in the student's explanation of answer (a) and the analysis of the experts' response to the student.



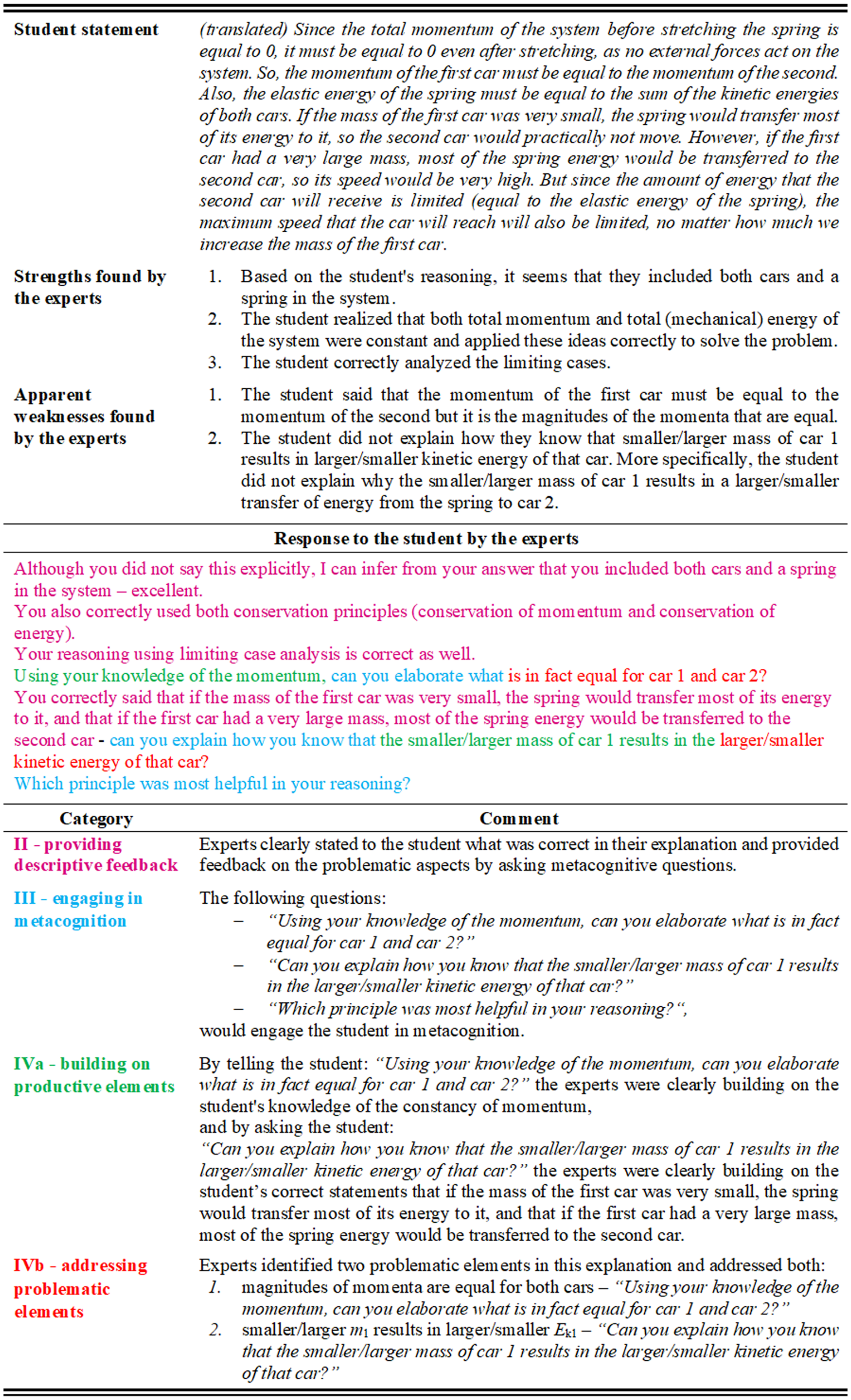
Student statement	*b. if you increase the mass of one and keep the velocity the same, then the velocity of 2 must increase by the same amount, as m_2 stays the same. This will not be directly proportional however, because the more mass you put on, the greater the cart has a resistance to motion, which means low momentum translated to a constant v_2.*
Strengths found by the experts	1. Based on the student's reasoning, it seems that they included both cars in the system. 2. It also seems that initially they thought (without explicitly saying) that the total momentum is constant. 3. The student analyzed the limiting case when $m_1 \to \infty$.
Apparent weaknesses found by the experts	1. Initially the student incorrectly assumed that the speed of car 1 is constant. 2. It seems that the student thought that their proportional argument could not extend to the limit of infinite mass and therefore the $v_2(m_1)$ graph has an asymptote. 3. The student did not explicitly mention the spring and/or energy conservation here. 4. There is no limiting case analysis when $m_1 \to 0$.

Response to the student by the experts

Although you did not say this explicitly, I can infer from your answer that you included both cars in the system and that you realized that the total momentum of the system is constant, which is correct.
But it seems to me that you assumed that the speed of car 1 remains constant. Based on what did you make this assumption?
Can you explain more what you mean when talking about resistance to motion?
Think about what else might be constant in your system besides momentum. Where did the kinetic energies of the cars come from?
I see that you addressed the limiting case analysis when m_1 is very large. How about the limiting case when m_1 is very very small?

Category	Comment
II - providing descriptive feedback	Experts clearly stated to the student what was correct in their explanation and asked metacognitive questions that would help the student figure out what they did incorrectly. They also acknowledged the student's limiting case analysis.
III - engaging in metacognition	The following questions and statements: – *"Based on what did you make this assumption?",* – *"Can you explain more what you mean when talking about resistance to motion?",* – *"Think about what else might be constant in your system besides momentum",* would engage the student in metacognition.
IVa - building on productive elements	By asking the student: *"...what else might be constant..."* the experts are clearly building on the student's productive idea of constancy (of momentum). The comment *"I see that you addressed the limiting case analysis when m_1 is very large. How about the limiting case when m_1 is very very small?"* builds on the students' limiting case analysis when m_1 is very large.
IVb - addressing problematic elements	Experts identified four problematic elements in this explanation. All four problematic elements were addressed: 1. wrong assumption about v_1 – *"Based on what did you make this assumption?"* 2. no energy approach – *"Think about what else might be constant in your system besides momentum. Where did the kinetic energies of the cars come from?"* 3. wrong reasoning (resistance to motion) – *"Can you explain more what you mean when talking about resistance to motion?“* 4. no $m_1 \to 0$ analysis – *"How about the limiting case when m_1 is very very small?"*

FIG. 11. All productive and problematic ideas experts identified in the student's incorrect explanation of answer (b) and the analysis of the experts' response to the student.



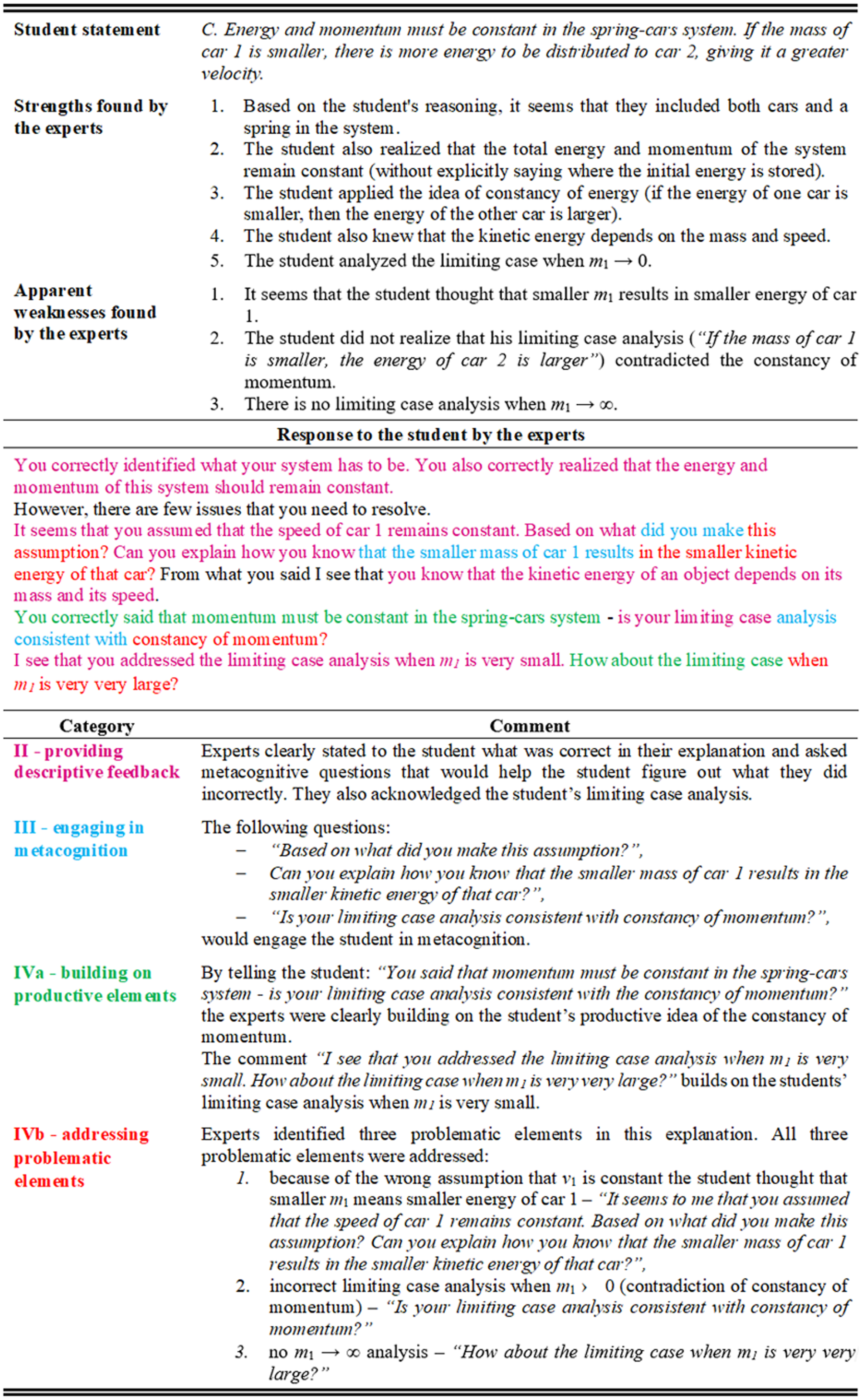
Student statement	*(translated) Since the total momentum of the system before stretching the spring is equal to 0, it must be equal to 0 even after stretching, as no external forces act on the system. So, the momentum of the first car must be equal to the momentum of the second. Also, the elastic energy of the spring must be equal to the sum of the kinetic energies of both cars. If the mass of the first car was very small, the spring would transfer most of its energy to it, so the second car would practically not move. However, if the first car had a very large mass, most of the spring energy would be transferred to the second car, so its speed would be very high. But since the amount of energy that the second car will receive is limited (equal to the elastic energy of the spring), the maximum speed that the car will reach will also be limited, no matter how much we increase the mass of the first car.*
Strengths found by the experts	1. Based on the student's reasoning, it seems that they included both cars and a spring in the system. 2. The student realized that both total momentum and total (mechanical) energy of the system were constant and applied these ideas correctly to solve the problem. 3. The student correctly analyzed the limiting cases.
Apparent weaknesses found by the experts	1. The student said that the momentum of the first car must be equal to the momentum of the second but it is the magnitudes of the momenta that are equal. 2. The student did not explain how they know that smaller/larger mass of car 1 results in larger/smaller kinetic energy of that car. More specifically, the student did not explain why the smaller/larger mass of car 1 results in a larger/smaller transfer of energy from the spring to car 2.

Response to the student by the experts

Although you did not say this explicitly, I can infer from your answer that you included both cars and a spring in the system – excellent.
You also correctly used both conservation principles (conservation of momentum and conservation of energy).
Your reasoning using limiting case analysis is correct as well.
Using your knowledge of the momentum, can you elaborate what is in fact equal for car 1 and car 2?
You correctly said that if the mass of the first car was very small, the spring would transfer most of its energy to it, and that if the first car had a very large mass, most of the spring energy would be transferred to the second car - can you explain how you know that the smaller/larger mass of car 1 results in the larger/smaller kinetic energy of that car?
Which principle was most helpful in your reasoning?

Category	**Comment**
II - providing descriptive feedback	Experts clearly stated to the student what was correct in their explanation and provided feedback on the problematic aspects by asking metacognitive questions.
III - engaging in metacognition	The following questions: – *"Using your knowledge of the momentum, can you elaborate what is in fact equal for car 1 and car 2?"* – *"Can you explain how you know that the smaller/larger mass of car 1 results in the larger/smaller kinetic energy of that car?"* – *"Which principle was most helpful in your reasoning?"*, would engage the student in metacognition.
IVa - building on productive elements	By telling the student: *"Using your knowledge of the momentum, can you elaborate what is in fact equal for car 1 and car 2?"* the experts were clearly building on the student's knowledge of the constancy of momentum, and by asking the student: *"Can you explain how you know that the smaller/larger mass of car 1 results in the larger/smaller kinetic energy of that car?"* the experts were clearly building on the student's correct statements that if the mass of the first car was very small, the spring would transfer most of its energy to it, and that if the first car had a very large mass, most of the spring energy would be transferred to the second car.
IVb - addressing problematic elements	Experts identified two problematic elements in this explanation and addressed both: *1.* magnitudes of momenta are equal for both cars – *"Using your knowledge of the momentum, can you elaborate what is in fact equal for car 1 and car 2?"* *2.* smaller/larger m_1 results in larger/smaller E_{k1} – *"Can you explain how you know that the smaller/larger mass of car 1 results in the larger/smaller kinetic energy of that car?"*

FIG. 12. All productive and problematic ideas experts identified in the student's correct explanation of answer (b) and the analysis of the experts' response to the student.



Student statement	*C. Energy and momentum must be constant in the spring-cars system. If the mass of car 1 is smaller, there is more energy to be distributed to car 2, giving it a greater velocity.*
Strengths found by the experts	1. Based on the student's reasoning, it seems that they included both cars and a spring in the system. 2. The student also realized that the total energy and momentum of the system remain constant (without explicitly saying where the initial energy is stored). 3. The student applied the idea of constancy of energy (if the energy of one car is smaller, then the energy of the other car is larger). 4. The student also knew that the kinetic energy depends on the mass and speed. 5. The student analyzed the limiting case when $m_1 \rightarrow 0$.
Apparent weaknesses found by the experts	1. It seems that the student thought that smaller m_1 results in smaller energy of car 1. 2. The student did not realize that his limiting case analysis (*“If the mass of car 1 is smaller, the energy of car 2 is larger”*) contradicted the constancy of momentum. 3. There is no limiting case analysis when $m_1 \rightarrow \infty$.

Response to the student by the experts

You correctly identified what your system has to be. You also correctly realized that the energy and momentum of this system should remain constant.
However, there are few issues that you need to resolve.
It seems that you assumed that the speed of car 1 remains constant. Based on what did you make this assumption? Can you explain how you know that the smaller mass of car 1 results in the smaller kinetic energy of that car? From what you said I see that you know that the kinetic energy of an object depends on its mass and its speed.
You correctly said that momentum must be constant in the spring-cars system - is your limiting case analysis consistent with constancy of momentum?
I see that you addressed the limiting case analysis when m_1 is very small. How about the limiting case when m_1 is very very large?

Category	Comment
II - providing descriptive feedback	Experts clearly stated to the student what was correct in their explanation and asked metacognitive questions that would help the student figure out what they did incorrectly. They also acknowledged the student’s limiting case analysis.
III - engaging in metacognition	The following questions: – *“Based on what did you make this assumption?”*, – *Can you explain how you know that the smaller mass of car 1 results in the smaller kinetic energy of that car?”*, – *“Is your limiting case analysis consistent with constancy of momentum?”*, would engage the student in metacognition.
IVa - building on productive elements	By telling the student: *“You said that momentum must be constant in the spring-cars system - is your limiting case analysis consistent with the constancy of momentum?”* the experts were clearly building on the student’s productive idea of the constancy of momentum. The comment *“I see that you addressed the limiting case analysis when m_1 is very small. How about the limiting case when m_1 is very very large?”* builds on the students’ limiting case analysis when m_1 is very small.
IVb - addressing problematic elements	Experts identified three problematic elements in this explanation. All three problematic elements were addressed: *1.* because of the wrong assumption that v_1 is constant the student thought that smaller m_1 means smaller energy of car 1 – *“It seems to me that you assumed that the speed of car 1 remains constant. Based on what did you make this assumption? Can you explain how you know that the smaller mass of car 1 results in the smaller kinetic energy of that car?”*, 2. incorrect limiting case analysis when m_1 › 0 (contradiction of constancy of momentum) – *“Is your limiting case analysis consistent with constancy of momentum?”* *3.* no $m_1 \rightarrow \infty$ analysis – *“How about the limiting case when m_1 is very very large?”*

FIG. 13. All productive and problematic ideas experts identified in the student’s explanation of answer (c) and the analysis of the experts’ response to the student.

APPENDIX B

Here, in Tables II–VIII, we provide coding schemes that we used to analyze the participants' responses. Since we used almost the same coding schemes to score participants' responses to incorrect explanations and the correct one, in Tables IX–XI, we only provide examples of how we scored several participants' responses for the correct explanation, without the description of the criteria.

TABLE II. Coding scheme for code Ia—interpreting productive reasoning.

Score	Description of the criteria	Example of a teachers' responses for answer (c)
0	The teacher did not write anything.	
	or	
	The teacher wrote something that was inappropriate/irrelevant or incorrect.	*"The student is thinking."* (irrelevant)
1	The teacher mentioned only some of the productive ideas in the student's explanation.	*"Correct identification of both conserved quantities."* (for a complete list of productive ideas see Fig. 13)
	or	
	Some of the things that the teacher wrote were correct but the teacher's response was too vague.	*"Noticed terms that are key to the development of the situation."*
2	The teacher mentioned all of the productive ideas in the student's explanation and the teacher's answer was clear.	*"1) The student clearly chooses/defines the observed system.* *2) The student takes into account the conservation laws in the chosen system.* *3) The student adheres to mathematical reasoning: if* $m_1 = 0$, *then he cannot have speed. All the energy can only go to the car 2. The student does not take into account that the energy can stay in the spring."*

TABLE III. Coding scheme for code Ib—interpreting problematic reasoning.

Score	Description of the criteria	Example of a teachers' responses for answer (b) (incorrect explanation)
0	The teacher did not write anything.	
	or	
	The teacher wrote something that was inappropriate/irrelevant or incorrect.	*"The student "remembers" too many details and in such a situation can easily leave something out of consideration."* (irrelevant)
1	The teacher mentioned only some of the problematic ideas in the student's explanation.	*"The student makes an assumption (incorrectly) about* v_1 *staying constant. Does not include thinking about energy."* (for a complete list of problematic ideas see Fig. 11)
	or	
	Some of the things that the teacher wrote were correct but the teacher's response was too vague.	*"He poorly argues his answer."*
2	The teacher mentioned all of the problematic ideas in the student's explanation and the teacher's answer was clear.	*"In the second sentence, he thinks about what would happen if we kept* v_1 *the constant, but we don't do that in our experiment, and so the reasoning isn't necessarily relevant.* *In the third sentence, he says "less momentum," although he probably means a slight increase in momentum.* *He also makes no mention of conserving energy, which is otherwise crucial to the calculation itself if he wanted to test his intuition.* *He does not analyse the limit case when* m_1 *goes towards 0."*

TABLE IV. Coding scheme for code II—providing descriptive feedback.

Score	Description of the criteria	Example of a teachers' responses for answer (c)
0	The teacher did not provide feedback for neither correct nor the incorrect aspects of the student's reasoning.	*"Interesting, explain to me why you chose these particular assumptions?"* (for a complete list of productive and problematic ideas see Fig. 13)
	or The teacher wrote something that would not help the student realize they were wrong.	*"Think about conservation of momentum, think about conservation of energy."*
1	The teacher only provided feedback on the correct aspect of student reasoning or only on the incorrect. For the correct aspect the teacher clearly stated to the student what was correct in their reasoning and for the incorrect aspect the teacher either clearly stated to the student what was incorrect in their reasoning or asked metacognitive questions that would help the student figure out what they did incorrectly.	*"What if the mass of car 1 is extremely small? How would the energy be distributed between the cars then?"* (this is feedback for the incorrect aspects only; see a complete list of problematic ideas in Fig. 13)
2	The teacher provided feedback for both correct and incorrect aspects. For the correct aspect the teacher clearly stated to the student what was correct in their reasoning and for the incorrect aspect the teacher either clearly stated to the student what was incorrect in their reasoning or asked metacognitive questions that would help the student figure out what they did incorrectly.	*"You have correctly considered what quantities are constant. If the spring pushes a heavier and lighter car, which one will get more speed? Think again about limiting cases."* (for a complete list of productive and problematic ideas see Fig. 13)

TABLE V. Coding scheme for code III—engaging the student in metacognition.

Score	Description of the criteria	Example of a teachers' responses for answer (a)
0	The teacher wrote something that would not engage the student in metacognition as the teacher's response did not reflect what the student was thinking about.	*"Think about energy."* (for the student's explanation see Fig. 10)
2	The teacher wrote a response that would engage the student in metacognition—teacher's response clearly reflected what the student was thinking about.	*"You say," … the greater the mass 1, the greater the velocity 2. "Why?"* (for the student's explanation see Fig. 10)

TABLE VI. Coding scheme for code IVa—building on productive elements.

Score	Description of the criteria	Example of a teachers' responses for answer (c)
0	The teacher did not build on what they found to be productive reasoning.	*"The initial speed of the car is zero, not some value."*
	or The teacher wrote something that was inappropriate/ irrelevant or incorrect or too vague.	*"Think about conservation of momentum, think about conservation of energy."* (vague)
1	The teacher's response can be interpreted as building on the productive elements they found (it was implicit).	*"…you are making statements" …car 1 is smaller, there is more energy… car 2 "that are not supported. You need to make the connection to the physics principle that shows this statement to be true."*
2	The teacher's response clearly showed building on the productive elements they found (it was explicit).	*"You identified momentum as being constant but do not apply that in your answer. How does that affect things? …"*

TABLE VII. Coding scheme for code IVb—addressing problematic elements.

Score	Description of the criteria	Example of a teachers' responses for answer (a)
0	The teacher wrote something that was inappropriate/irrelevant or incorrect.	*"Take friction into account."* (irrelevant)
	or	
	The teacher did not address any of the found problematic elements.	No example for answer (a)
1	The teacher addressed some of the problematic elements they found (explicitly or implicitly).	*"Consider limiting cases. Write down all the quantities that occur in the conservation of momentum and for each one, consider how it changes."*
	or	
	The teacher's response might (implicitly) address problematic elements but the response is too vague.	*"Consider the basic laws that apply to this experiment."*
2	The teacher addressed all of the problematic elements they found (explicitly or implicitly).	*"If* m_1 *is very, very large, the experiment is the same as for a recoil from a rigid wall. Does the speed of the car go in this case really increase with no limits? Where does a car get its energy to move? How is energy related to its speed?"*

Building on productive elements and addressing problematic ones can only be interpreted based on the productive and problematic aspects each teacher found. Each teacher's response from coding schemes for building on productive and addressing problematic elements needs more context in order to fully understand why each score was assigned. Therefore, in addition to the teacher's response, we show below what a certain teacher identified as productive and problematic reasoning.

TABLE VIII. Examples of what certain teachers identified as productive (code Ia) and problematic (code Ib) reasoning and how their responses were scored for building on productive (code IVa) and addressing problematic elements (code IVb).

Code Ia	Code IVa	Code Ib	Code IVb
	Score 0		***Score 0***
"Respects the law of conservation of energy and momentum."	*"The initial speed of the car is zero, not some value."*	*"The speed cannot be increased continuously."*	*"Take friction into account."*
"The theorem about both conservations is correct"	*"Think about conservation of momentum, think about conservation of energy."* (vague)		
	Score 1		***Score 1***
"Thinking about momentum and energy conservation"	*"…you are making statements:" … car 1 is smaller, there is more energy… car 2 "that are not supported. You need to make the connection to the physics principle that shows this statement to be true."*	*"He chose the wrong graph. He silently assumed that* $v_1 = \text{const}$. *He did not think about a single limiting case and consequently did not recognize the problem (of his solution) at high* m_1.*"*	*"Consider limiting cases. Write down all the quantities that occur in conservation of the momentum and for each one consider how it changes."*
		"The student sees no problem in how the speed of something can become infinitely large"	*"Consider the basic laws that apply to this experiment"*
	Score 2		***Score 2***
"Constancy of energy and momentum are correctly identified."	*"You identified momentum as being constant but do not apply that in your answer. How does that affect things? …"*	*"The idea of* v_2 *increasing with no limits makes no sense. The student has no idea about energy."*	*"If* m_1 *is very, very large, the experiment is the same as for a recoil from a rigid wall. Does the speed of the car go in this case really increase with no limits? Where does a car get its energy to move? How is energy related to its speed?"*

TABLE IX. Example 1 of how we scored the participant's response to the correct explanation.

	Participant's response	Code	Score
Student's strengths	*"Analyses looking at limiting cases, uses energy when useful, and momentum when useful, easily switches between the two."*	Ia	1
Student's weaknesses	*"None."*	Ib	0
Response to the student	*"Excellent work!"*	II	1
		III	0
		IVa	0
		IVb	0

TABLE X. Example 2 of how we scored the participant's response to the correct explanation.

	Participant's response	Code	Score
Student's strengths	*"The answer contains everything you would expect from a student."*	Ia	0
Student's weaknesses	*"1) If we are very meticulous, the student did not strictly explain the choice of the observed systems, but this did not hinder his reasoning.* *2) I don't know how he knows that in the case of a small mass, the spring would transfer most of the energy to the first cart. He needs to know this from somewhere, because without derivation, in my opinion, he can't know it."*	Ib	1
Response to the student	*"Great. Good thinking and good justification. You did not have any problems with this task, but in the future, I advise you to be more clear about what you choose for the observed system. This can help you in more complex cases."*	II	1
		III	0
		IVa	0
		IVb	1

TABLE XI. Example 3 of how we scored the participant's response to the correct explanation.

	Participant's response	Code	Score
Student's strengths	*"This makes very clear and correct use of the principles of conservation of momentum and energy. It is clear that the student is not only able to articulate the principles but can connect them to this situation."*	Ia	1
Student's weaknesses	*"My only gripe would be that the student doesn't offer an explanation of why a large mass discrepancy leads to most of the energy being transferred to the smaller object."*	Ib	1
Response to the student	*"Excellent work! you did a great job of explaining conservation of energy and momentum and why they are relevant to this situation. One way to strengthen the explanation would be to add a little detail about why the lighter object receives more of the energy."*	II	2
		III	2
		IVa	1
		IVb	2

APPENDIX C

Here, we describe the results of the statistical analysis. To determine whether the differences among the groups for each of the codes were statistically significant, we performed statistical analyses using SPSS 25.0 (IBM, Armonk, NY, USA). As we expressed categorical variables as percentages for descriptive statistics, we used Fisher's exact test and chi-square test for between-group analyses. We agreed to call results significant for $p < 0.05$. If the chi-square test and Fisher's exact test were significant, we continued the analysis with the multiple comparisons of column proportions with Bonferroni correction.

Though participants' results for interpreting productive reasoning in students' incorrect explanations were very similar for the first five groups and different for less experienced pre-service teachers B [see Fig. 4(a)], a statistically significant difference was only found between this group and less experienced pre-service teachers A. The percentage of responses of less experienced pre-service teachers A (93%) who received a score of 1 was significantly higher than the percentage of responses of less experienced pre-service teachers B (50%) who received a score of 1, $p = 0.004$. We believe that the sizes of individual groups significantly influenced the results of the statistical analysis (the group of in-service teachers A was the largest).

Likewise, though participants' results for interpreting problematic reasoning were very similar for the first five groups and different for less experienced pre-service teachers B [see Fig. 4(b)], and even though the chi-square test showed significant differences in the percentages of responses of individual groups who received a score of 0 and a score of 1, the multiple comparisons of column proportions with Bonferroni correction were not possible (proportions 0 or 1 in the tables when spss does the statistical analysis).

Statistical analysis revealed a significant difference between international in-service teachers and less experienced pre-service teachers B who received a score of 0 for providing descriptive feedback when students' explanations were incorrect [see Fig. 5(a)]. The multiple comparisons of column proportions with Bonferroni correction showed that the percentage of responses of less experienced pre-service teachers B (83%) who received a score of 0 was significantly higher than the percentages of responses of international in-service teachers (23%) who received a score of 0, $p = 0.002$.

Despite visible similarities between international in-service teachers, in-service teachers A and both groups of pre-service teachers A, and similarities between in-service teachers B and less experienced pre-service teachers B for engaging the students in metacognition when students' explanations were incorrect [see Fig. 5(b)], the multiple comparisons of column proportions with Bonferroni correction only showed that the percentages of responses of international in-service teachers (64%) and less experienced pre-service teachers A (69%) who received a score of 2 were significantly higher than the percentage of responses of in-service teachers B (28%) who received a score of 2, $p = 0.024$ and $p = 0.004$, respectively.

We found several statistically significant differences among and between the groups for addressing problematic elements in students' incorrect explanations [see Fig. 6(b)], The percentages of responses of international in-service teachers (56%) and less experienced pre-service teachers A (56%) who received a score of 2 were significantly higher than the percentage of responses of in-service teachers B (22%) who received a score of 2, $p = 0.038$ and $p = 0.036$, respectively. Recall that the participants were not "punished" with lower scores if they did not address all problematic elements that experts identified in the student's explanation. Instead, the participants' responses, for addressing problematic elements, were scored based on their list of identified problematic elements.

The percentage of responses of less experienced pre-service teachers B (92%) who received a score of 0 was significantly higher than the percentages of responses of international in-service teachers (26%, $p = 0.001$), in-service teachers A (28%, $p = 0.001$), experienced pre-service teachers A (7%, $p < 0.001$) and less experienced pre-service teachers A (22%, $p < 0.001$) who received a score of 0. Likewise, the percentage of responses of in-service teachers B (53%) who received a score of 0 was significantly higher than the percentage of responses of experienced pre-service teachers A (7%) who received a score of 0, $p = 0.032$.

Even though Fisher's exact test showed significant differences between the percentages of responses of individual groups who received a score of 2, $p = 0.042$, the multiple comparisons of column proportions with Bonferroni correction showed that there were no significant differences between the percentages of responses of individual groups ($p > 0.05$) for addressing problematic elements when the student's explanation was correct [see Fig. 9(b)]. Fisher's exact test showed no significant differences between the percentages of responses of individual groups who received a score of 1, $p = 0.668$. Fisher's exact test showed significant differences between the percentages of responses of individual groups who received a score of 0, $p = 0.010$. Unfortunately, the multiple comparisons of column proportions with Bonferroni correction were not possible (proportions 0 or 1 in the tables when spss does the statistical analysis).

[1] J. E. Zull, *The Art of Changing the Brain: Enriching Teaching by Exploring the Biology of Learning* (Stylus Publishing, Sterling, VA, 2002).

[2] S. Dehaene, *How We Learn: The New Science of Education and the Brain* (Penguin Publishing Group, Toronto, 2021).

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

[4] S. Freeman, S. Eddy, M. McDonough, M. Smith, N. Okoroafor, H. Jordt, and M. Wenderoth, Active Learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U. S. A. **111**, 8410 (2014).

[5] J. Michael, Where's the evidence that active learning works?, Adv. Physiol. Educ. **30**, 159 (2007).

[6] J. Von Korff, B. Archibeque, K. Gomez, T. Heckendorf, S. Mckagan, E. Sayre, E. Schenk, C. Shepherd, and L. Sorell,

Secondary analysis of teaching methods in introductory physics: A 50k-student study, Am. J. Phys. **84**, 969 (2016).

[7] M. Chi, N. Leeuw, M.-H. Chiu, and C. Lavancher, Eliciting self-explanations improves understanding, Cogn. Sci. **18**, 439 (1994).

[8] M. Chi, M. Bassok, M. Lewis, P. Reimann, and R. Glaser, Self-explanations: How students study and use examples in learning to solve problems, Cogn. Sci. **13**, 145 (1989).

[9] E. Furtak, R. Bakeman, and J. Buell, Developing knowledge-in-action with a learning progression: Sequential analysis of teachers' questions and responses to student ideas, Teach. Teach. Educ. **76**, 267 (2018).

[10] A. Gotwals and D. Birmingham, Eliciting, identifying, interpreting, and responding to students' ideas: Teacher candidates' growth in formative assessment practices, Res. Sci. Educ. **46**, 365 (2015).

[11] B. Kristinsdóttir, F. Hreinsdóttir, Z. Lavicza, and C. Wolff, Teachers' noticing and interpretations of students' responses to silent video tasks, Res. Math. Educ. **22**, 135 (2020).

[12] M. Luna and S. Selmer, Examining the responding component of teacher noticing: A case of one teacher's pedagogical responses to students' thinking in classroom artifacts, J. Teach. Educ. **72**, 579 (2021).

[13] B. Tataroğlu Tasdan and M. G. Didiş Kabar, Pre-service mathematics teachers' responding to student thinking in their teaching experiences, J. Pedagog. Res. **6**, 87 (2022).

[14] C. Ebby, J. Remillard, and J. D'Olier, Pathways for analyzing and responding to student work for formative assessment: The role of teachers' goals for student learning, Working Paper, CPRE Working Papers, n.d, https://repository.upenn.edu/cpre_workingpapers/22.

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

[16] H. Kang and C. Anderson, Supporting preservice science teachers' ability to attend and respond to student thinking by design, Sci. Educ. **99**, 863 (2015).

[17] V. Dini, H. Sevian, K. Caushi, and R. Picón, Characterizing the formative assessment enactment of experienced science teachers, Sci. Educ. **104**, 290 (2020).

[18] E. Etkina, D. Gitomer, C. Iaconangelo, G. Phelps, L. Seeley, and S. Vokos, Design of an assessment to probe teachers' content knowledge for teaching: An example from energy in high school physics, Phys. Rev. Phys. Educ. Res. **14**, 010127 (2018).

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

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

[21] O. de Jong, J. Van Driel, and N. Verloop, Preservice teachers' pedagogical content knowledge of using particle models in teaching chemistry, J. Res. Sci. Teach. **42**, 947 (2005).

[22] J. Gess-Newsome, Pedagogical content knowledge, in *International Handbook of Student Achievement* (Routledge, New York, 2011), pp. 257–259.

[23] *Understanding and Developing Science Teachers' Pedagogical Content Knowledge*, edited by J. Loughran, A. Berry, and P. Mulhall (Sense Publishers, Rotterdam, 2012).

[24] *Re-Examining Pedagogical Content Knowledge in Science Education*, edited by A. Berry, P. Friedrichsen, and J. Loughran (Routledge, New York, 2015).

[25] H. Jing-Jing, A critical review of pedagogical content knowledge' components: Nature, principle and trend, Int. J. Educ. Res. **2**, 411 (2014), https://www.ijern.com/April-2014.php.

[26] 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* (Kluwer Academic Publishers, Dordrecht, 1999), pp. 95–132.

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

[28] J. Gess-Newsome, A model of teacher professional knowledge and skill including PCK: Results of the thinking from PCK summit, in *Re-Examining Pedagogical Content Knowledge in Science Education* (Routledge, New York, 2015), pp. 28–42.

[29] *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, edited by A. Hume, R. Cooper, and A. Borowski (Springer, Singapore, 2019).

[30] J. Driel, D. Beijaard, and N. Verloop, Professional development and reform in science education: The role of teachers' practical knowledge, J. Res. Sci. Teach. **38**, 137 (2001).

[31] L. Seeley, S. Vokos, and E. Etkina, Examining physics teacher understanding of systems and the role it plays in supporting student energy reasoning, Am. J. Phys. **87**, 510 (2019).

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

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

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

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

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

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

[38] J. P. Smith, A. A. diSessa, and J. Roschelle, Misconceptions reconceived: A constructivist analysis of knowledge in transition, J. Learn. Sci. **3**, 115 (1993).

[39] J. Roschelle, Learning in interactive environments: Prior knowledge and new experience, in *Public Institutions for Personal Learning: Establishing a Research Agenda*

(American Association of Museums, Washington, DC, 1997), pp. 37–51.

[40] A. Disessa, Unlearning Aristotelian physics: A study of knowledge-based learning, Cogn. Sci. **6**, 37 (1982).

[41] L. Resnick, Mathematics and science learning: A new conception, Science **220**, 477 (1983).

[42] A. A. diSessa, Knowledge in pieces, in *Constructivism in the Computer Age*, edited by G. Forman and P. Pufall (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1988), pp. 49–70.

[43] D. B. Harlow and J. A. Bianchini, Knowledge-in-pieces—Andrea A. diSessa, David Hammer, in *Science Education in Theory and Practice, An Introductory Guide to Learning Theory*, Springer Texts in Education (Springer, Cham, 2020), pp. 389–401.

[44] D. Hammer, Student resources for learning introductory physics, Am. J. Phys. **68**, S52 (2000).

[45] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, J. Learn. Sci. **12**, 53 (2003).

[46] D. Hammer, A. Elby, R. Scherr, and E. Redish, Resources, framing, and transfer, in *Transfer of Learning from a Modern Multidisciplinary Perspective* (Information Age Publishing, Greenwich, CT, 2005), pp. 89–119.

[47] A. Richards, D. Jones, and E. Etkina, How students combine resources to make conceptual breakthroughs, Res. Sci. Educ. **50**, 1119 (2020).

[48] D. Ball, With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics, Elem. Sch. J. **93**, 373 (1993).

[49] T. Carpenter, E. Fennema, and M. Franke, Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction, Elem. Sch. J. **97**, 3 (1996).

[50] D. Hammer, F. Goldberg, and S. Fargason, Responsive teaching and the beginnings of energy in a third grade classroom, Rev. Sci. Math. ICT Educ. **6**, 51 (2012), https://pasithee.library.upatras.gr/review/article/view/1694/1626.

[51] D. Hammer and E. H. van Zee, *Seeing the Science in Children's Thinking: Case Studies of Student Inquiry and Physical Science* (Heinemann, Portsmouth, 2006).

[52] V. Jacobs, L. Lamb, and R. Philipp, Professional noticing of children's mathematical thinking, J. Res. Math. Educ. **41**, 169 (2010).

[53] M. Sherin and E. van Es, Using video to support teachers' ability to notice classroom interactions, J. Technol. Teach. Educ. **13**, 475 (2005), https://www.academia.edu/2490187/Using_Video_to_Support_Teachers_Ability_to_Notice_Classroom_Interactions.

[54] B. Harrer, V. Flood, and M. Wittmann, Productive resources in students' ideas about energy: An alternative analysis of Watts' original interview transcripts, Phys. Rev. ST Phys. Educ. Res. **9**, 023101 (2013).

[55] *Responsive Teaching in Science and Mathematics*, edited by A. Robertson, R. Scherr, and D. Hammer (Routledge, London, 2016).

[56] M. Windschitl, J. Thompson, and M. Braaten, *Ambitious Science Teaching* (Harvard Education Press, Boston, 2018).

[57] M. Windschitl, J. Thompson, M. Braaten, and D. Stroupe, Proposing a core set of instructional practices and tools for teachers of science, Sci. Educ. **96**, 878 (2012).

[58] L. Shepard, The role of assessment in a learning culture, Educ. Res. **29**, 4 (2000).

[59] N. R. C. NRC, *Classroom Assessment and the National Science Education Standards* (National Academy Press, Washington, DC, 2001).

[60] E. Furtak, *'Flying Blind': An Exploration of Beginning Science Teachers' Enactment of Formative Assessment Practices* (New Orleans, LA, 2011).

[61] S. Brookhart, *How to Give Effective Feedback to Your Students* (Association for Supervision and Curriculum Development, Alexandria, VA, 2008).

[62] J. Hattie and H. Timperley, The power of feedback, Rev. Educ. Res. **77**, 81 (2007).

[63] D. Buggé, Improving scientific abilities through lab report revision in a high school investigative science learning environment classroom, Phys. Rev. Phys. Educ. Res. **19**, 020166 (2023).

[64] P. Peterson, E. Fennema, T. Carpenter, and M. Loef, Teacher's pedagogical content beliefs in mathematics, Cognit. Instr. **6**, 1 (1989).

[65] B. Kramarski and Z. Mevarech, Enhancing mathematical reasoning in the classroom: Effects of cooperative learning and metacognitive training, Am. Educ. Res. J. **40**, 281 (2003).

[66] M. Veenman and J. Beishuizen, Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint, Learn. Instr. **14**, 621 (2004).

[67] M. Zion, T. Michalsky, and Z. Mevarech, The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills, Int. J. Sci. Educ. **27**, 957 (2005).

[68] E. Ader, A framework for understanding teachers' promotion of students' metacognition, Int. J. Math. Teach. Learn. (2013), https://www.cimt.org.uk/journal/.

[69] J. Flavell, Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry, Am. Psychol. **34**, 906 (1979).

[70] E. Etkina, When learning physics mirrors doing physics, Phys. Today **76**, No. 10, 26 (2023).

[71] D. Hestenes, Toward a modeling theory of physics instruction, Am. J. Phys. **55**, 440 (1987).

[72] E. Etkina, D. T. Brookes, and G. Planinsic, *Investigative Science Learning Environment When Learning Physics Mirrors Doing Physics* (Morgan & Claypool Publishers, San Rafael, CA, 2019), https://doi.org/10.1088/2053-2571/ab3ebd.

[73] E. Etkina, Using early teaching experiences, and a professional community to prepare pre-service teachers for every-day classroom challenges, to create habits of student-centered instruction, and to prevent attrition, in *Recruiting, and Educating Future Physics Teachers: Case Studies, and Effective Practices*, edited by C. Sandifer and E. Brewe (American Physical Society, College Park, MD, 2015), pp. 257–274.

[74] E. Etkina, D. Brookes, G. Planinsic, and A. Van Heuvelen, *Instructor Guide for College Physics: Explore and Apply*, 2nd ed. (Pearson, New York, 2019).

[75] E. Simonović, Master thesis—in Progress, University of Ljubljana, Faculty of Mathematics and Physics.

[76] E. Etkina, G. Planinsic, and A. Van Heuvelen, *College Physics: Explore and Apply*, 2nd ed. (Pearson, New York, 2019).

[77] A. R. Warren, Impact of teaching students to use evaluation strategies, Phys. Rev. ST Phys. Educ. Res. **6**, 020103 (2010).

[78] E. Burkholder, L. Blackmon, and C. Wieman, Characterizing the mathematical problem-solving strategies of transitioning novice physics students, Phys. Rev. Phys. Educ. Res. **16**, 020134 (2020).

[79] E. Etkina, B. Gregorcic, and S. Vokos, Organizing physics teacher professional education around productive habit development: A way to meet reform challenges, Phys. Rev. Phys. Educ. Res. **13**, 010107 (2017).

[80] T. Carpenter, E. Fennema, P. Peterson, C.-P. Chiang, and M. Loef, Using knowledge of children's mathematics thinking in classroom teaching: An experimental study, Am. Educ. Res. J. **26**, 499 (1989).