A new approach for uncovering student resources with multiple-choice questions

Nolan K. Weinlader, Eric Kuo, Benjamin M. Rottman, and Timothy J. Nokes-Malach
Learning Research & Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA, 15224

The traditional approach to studying student understanding presents a question and uses the student's answer to make inferences about their knowledge. However, this method doesn't capture the range of possible alternative ideas available to students. We use a new approach, asking students to generate a plausible explanation for every choice of a multiple-choice question, to capture a range of explanations that students can generate in answering physics questions. Asking 16 students to provide explanations in this way revealed alternative possibilities for student thinking that would not have been captured if they only provided one solution. The findings show two ways these alternatives can be productive for learning physics: (i) even students who ultimately chose the wrong answer could often generate the correct explanation and (ii) many incorrect explanations contained elements of correct physical reasoning. We discuss the instructional implications of this multiple-choice questioning approach and of students' alternative ideas.

I. INTRODUCTION

For examining students' conceptual knowledge, qualitative multiple choice and free response questions are the typical assessment tools. Such conceptual physics assessments play a major role in PER, uncovering the common conceptual difficulties that exist in the midst of and after instruction [1,2].

On these assessments, what do incorrect answers imply about students' conceptual knowledge? A resources framework (and other knowledge-in-pieces perspectives [3,4], in general) models students as possessing many cognitive resources, reflecting different ideas for learning and doing physics [5]. With these different resources, students can construct multiple explanations or predictions for a physical situation. Therefore, while a student's final, incorrect response to a physics question might be the dominant one they give, alternative explanations may be readily generated from other resources. Supporting these resource-based interpretations, prior research has shown that, even over relatively short episodes, students can shift from an initial explanation to alternative ones [6,7].

Although multiple-choice and free response questions are commonly used to capture and classify student ideas [8–11], these methods commonly focus on capturing students' dominant explanations. In this approach, a large group of students are asked a question, and their answers (indicating their dominant explanations) are used to map out a range of ideas used for understanding a physical topic. However, this approach fails to capture the alternative explanations that any one student might construct, because each student only provides one final answer. Therefore, the range of resources identified in this way represents the range of ideas among a population, but perhaps not the range of ideas easily accessible to any one particular student. Since learning can be viewed as utilizing and building on the appropriate, existing student resources [12], it is instructionally useful to know what alternative resources are accessible to students, even if they are not used in their dominant explanation.

This paper takes on the *methodological* challenge of uncovering the strongest alternatives to students' dominant reasoning. We asked students to generate explanations for *all choices* of a multiple-choice physics question. This contrasts with the typical approach of asking students which choice they think is correct and explaining that *one choice only*. Our approach explicitly prompts students to reveal the alternative explanations that they can construct without any additional instruction. This allowed us to examine whether a student whose dominant answer is incorrect can easily access the resources aligned with the correct answer. We also ask students to rate how likely they believe that each choice is correct. As opposed to simply choosing which answer they believed was correct, these

certainty ratings provided a more fine-grained measurement of the strengths of their conceptual views.
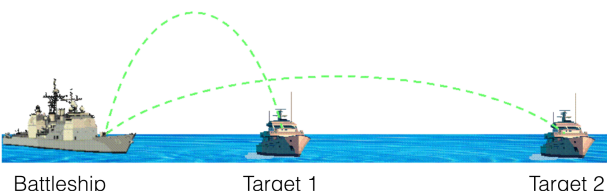
II. METHOD

A. Interview Protocol

Sixteen undergraduates who were enrolled in or had taken college physics were interviewed. The interviews were one-hour long. On a computer, students were led through a sequence of prompts, asking them to make selections on the screen or to provide verbal explanations to a researcher who was observing. Interviews were video recorded to capture these verbal explanations.

Students answered the prompts for 3-6 multiple-choice questions, depending on how quickly they progressed through the interview. In this paper, we discuss results from the first two multiple-choice questions, which all students completed. The Two Boats Q1 and Q2, taken from FlipItPhysics [13], considered a battleship launching shells at two targets (Fig. 1).

These two questions target student understanding of the connection between peak height and time in the air. In Q1, target 2 is hit first, because it has a lower peak height. In Q2, the targets are hit at the same time, because both shells reach the same peak height.

Rather than just having students answer the Two Boats questions, the interview protocol led students through an
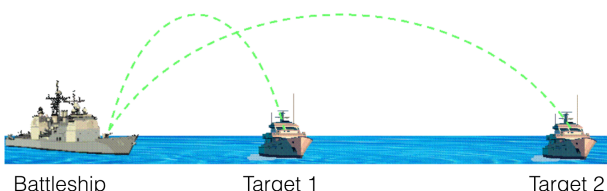
Two Boats Q1 (constant initial speed)

Battleship   Target 1   Target 2

A battleship simultaneously fires two shells with the same initial speed at two different targets. The shells follow the parabolic trajectories shown.

Which target gets hit first?

Two Boats Q2 (constant peak height)

Battleship   Target 1   Target 2

A battleship simultaneously fires two shells with *different* initial speeds at two different enemy ships. The shells follow the parabolic trajectories shown.

Which target gets hit first?

FIG 1. Two qualitative questions asking students to determine the relative flight times of two projectiles.

identical three-part sequence for each question, as follows:

Part 1 (Answer rating #1) – Students started by rating the likelihood of the answers:

Here are three possible answers to this question (A, B, C). Rate the likelihood that you believe each answer is correct. 0 – it's not likely at all. 100 – it's definitely correct.

A) Target 1 is hit first.
B) Target 2 is hit first.
C) The ships are hit at the same time.

Using sliders, students were able to choose whole numbers between 0 (labeled "least likely) and 100 (labeled "most likely") for each multiple-choice option. The total rating for all three had to sum to 100%, as we aimed to capture their belief distributions on a probability scale.

Part 2 (Generate Explanations) – Students were then directed to generate potential explanations for *each* of the three multiple-choice options:

For each choice, provide a reasonable explanation for why someone might choose it. Even if you yourself don't think a choice is correct, try to come up with the most convincing explanation that someone who selects that choice might believe (even if you don't believe that explanation yourself!). Please talk out loud about your explanations.

This prompt is based off of existing instructional approaches that ask students to consider multiple ways of thinking, even if they believe one explanation over another [14]. At the start of the interview, students were told that, when explaining their thinking to the researcher, they might be asked follow-up questions. Here, the follow-up questions aimed to make sure students provided an explanation for all three options and that the researcher understood each explanation.

Part 3 (Answer rating #2) – Students again rated how likely they thought each answer was correct, from 0 to 100. The results reported use this rating as a measure of how much students believed the explanations generated in part 2. The ratings in part 1 are not suitable for this purpose, because it is possible that students have not considered alternative explanations before they explicitly generate them in part 2.

B. Explanation coding

For Two Boats Q1 and Q2, students' verbal explanations for the three choices, A, B, and C, were coded into different categories. The first and second author generated an initial list of codes by listening to an initial subset of students' explanations. Then, the two coders independently coded all students' A, B, and C explanations for the two questions, discussing disagreements until all codes were agreed upon.

III. RESULTS

Broadly, the prompt to provide an explanation for all multiple-choice options was successful at uncovering alternative explanations: only one student failed to provide explanations for all three options on a question.

The number of coded explanations is shown in Table I. These 16 students generated between 3 to 7 different explanations for each multiple-choice option. This showed a wide diversity to students' ideas. Yet, at the same time, students' thinking clustered around a subset of these explanations. We split the most *common explanations* (given by 3 or more students) from the *uncommon explanations* (given by fewer than 3 students). Each common explanation was, on average, given by 7.5 students (SD = 4.3 students), and common explanations accounted for 76% of all explanations generated. All common explanations will be discussed further in the results. However, our primary result depends more on the nature of students' dominant and alternative answers and how closely each aligns with the canonical physical model of projectile motion.

A. Even students who choose the wrong answer can give the correct explanation

The first noteworthy finding is that even when students chose the wrong answer, their alternative explanations still sometimes contained the correct reasoning.

TABLE I. The number of unique explanations coded for each option of Two Boats Q1 and Q2.

	# of explanations (# of common explanations given by 3 or more students)		
	A	B	C
Two Boats Q1	3 (1)	5 (2)	6 (2)
Two Boats Q2	5 (3)	6 (1)	7 (2)

For Two Boats Q1, the correct answer was B (Target 2 is hit first). One valid explanation noted that the vertical peak of shell 2 is lower, using this to infer that the time in the air for shell 2 was shorter. Another answer that we considered valid is that shell 2 had a more direct path and/or lower launch angle. In the case of equal initial speeds, a lower launch angle does correctly imply less time in the air. Therefore, we saw mention of a "more direct path" as indicating a productive resource that could help students understand the angle dependence of time.

Using students' dominant rating on part 3 (answer rating #2) as an indication of which answer they would choose, few students chose the correct answer option for Two Boats Q1. Only 1 student believed choice B was the most likely answer, and 2 students considered it tied with other explanations as the most likely to be correct. Of these

3 students, one gave the "lower peak height" explanation and one gave the "more direct path" explanation.

The majority of students believed C (hit at the same time) to be the most likely answer for Two Boats Q1. 12/16 students believed it was the most likely answer, and 1 student considered it tied with another option. One question we aimed to answer with our new approach was whether these students could generate the correct explanations when explicitly prompted for alternatives. Of the 13 students who believed strongly in choice C, 11 provided a valid explanation for why B might be the answer. 5 students gave the "lower peak height" explanation, 4 students gave the "more direct path" explanation, and 2 students gave both of these explanations. Although standard assessment approaches (which include multiple-choice tests) would capture these students' belief in the wrong answer, it would not capture their ability to generate the valid explanations for the correct answer.

For Two Boats Q2, the majority of students chose incorrect option A: target 1 is hit first (9 students chose it as most likely, 2 tied). Yet, when generating explanations for answer C (the correct answer), 3 of those students gave the correct explanation: because the shells reach the same peak height, they hit at the same time. Though fewer students did so, this again shows that even students who chose the wrong answer can generate the correct explanation when asked for alternatives.

B. Many incorrect explanations contained elements of correct physical reasoning

The second finding was that even invalid explanations could have productive elements of correct physical reasoning. Kinematically, time is related to distance and speed. Many explanations correctly considered how one of these two factors related to time, but ignored the other. For example, some explanations considered distance only:

- Q1/Q2: A – Target 1 is hit first, because it is closer to the battleship (Q1: 14 students, Q2: 12 students).
- Q1: C – Both are hit at the same time, because they travel roughly the same path length. The difference is that shell 1 travels farther vertically and shell 2 travels farther horizontally (8 students).

Although both of these explanations are incorrect, they reflect correct dependences of distance on time if the speeds were equal throughout. However, they are not. For the first explanation, shell 1 has a smaller horizontal component of velocity than shell 2. For the second explanation, the speeds of the two shells do not remain equal across their trajectories. An additional problem with the second explanation here is that one cannot assume the distances traveled are exactly equal. Yet, these incorrect

explanations indicate a valid physical dependence that, used properly, can play a role in learning physics.

Similarly, some explanations considered speed only:

- Q1: C – Both are hit at the same time, because they are launched with the same initial speed (4 students).
- Q2: A – Shell 1 hits first, if it has a greater initial speed (3 students).
- Q2: A – Shell 1 reaches the peak sooner and gravity pulls it down faster, because of the greater angle (4 students).
- Q2: B – Shell 2 hits first, because it's traveling faster to get there (12 students).

These speed only explanations rely on a common *faster means less time* resource, which is consistent with kinematics. What makes these explanations incomplete is that they do not consider the effect of distance on time. For the Q2 explanations, some incorrect assumptions are also made, such as presuming that shell 1 has a greater initial speed or that shell 1 travels faster to the peak.

A complete description of motion here will integrate distance and speed to draw valid conclusions about time (as well as consistently break down the motion into horizontal and vertical components). However, even when both distance and time are included, the conclusion may not be valid. For example, on Q2, 6 students provided this explanation for choice C: shell 2 is traveling further, but it also travels faster, so it hits at the same time as shell 1. This explanation illustrates that, along with consideration of the appropriate quantities, students also need to learn how to draw valid inferences.

Given that this is a kinematics problem, it may seem a trivial conclusion that the incorrect responses commonly include distance, speed, or time, the key physical quantities in this problem. However, more than just naming relevant quantities, students also leverage correct relations between these quantities (e.g., less distance implies less time, more speed implies less time), in albeit incomplete ways. These are notably different from other established mechanics errors that indicate fundamentally incorrect relations between variables (e.g. motion implies a force [15]).

C. Constructing alternatives that aren't believed

Although this methodology can uncover alternative explanations available to students, many of these explanations were associated with low levels of belief. This is a byproduct of the prompt, which asks students to come up with an explanation for each choice, even if they did not believe that explanation themselves.

The benefit of this approach is that students may reveal the correct conceptual thinking in their alternative explanations. Again, in the case of Two Boats Q1, 11 students generated the correct explanation for the correct

answer, B, even though they believed that C was an equally or more likely answer.

At the same time, it may also force students to generate explanations they themselves view as highly unlikely. To illustrate, for Two Boats Q1, none of the 14 students who gave the explanation for A "target 1 is hit first, because it is closer to the battleship" believed A to be the most likely answer. The average part 3 certainty rating of choice A across these students was only 18%. Although many students provided this incorrect explanation, none of them believed it was likely to be correct. Similarly, for Q2, 12 students provided an incorrect explanation for choice B (Shell 2 hits first, because it's traveling faster to get there), but only 3 of these students believed B could be the correct answer. The average part 3 rating of choice B across these students was 19%.

IV. DISCUSSION

While student thinking on a multiple-choice physics question is often categorized as correct or incorrect, asking students to provide plausible explanations for each multiple-choice option reveals that student conceptual knowledge can occupy an intermediate state: a student could believe an incorrect answer, but be able to generate the correct explanation when prompted for alternatives. Current approaches that ask students to give the one answer they think is correct (i.e., their dominant answer) could not detect this intermediate state.

These findings can inform instructional methods that aim to tap into students' resources for learning physics. These existing instructional approaches use particular questions to intuitively activate students' existing resources and then build on these resources to teach physics [16,17]. One research question is whether these instructional approaches are more or less effective depending on how easily those resources can come to mind. By having students give their alternative explanations to a question, we can examine how "close to the surface" those resources are, and future research can investigate whether the availability of these resources (even if they are not believed) predicts the success of resource-focused instruction. A second issue is how easy or hard it is to shift a correct alternative explanation into the dominant one. Given that most students whose dominant explanations to the Two Boats Problems were incorrect could still come up with the correct explanation in their alternatives, the learning process here might be better described as a reweighting of ideas rather than a learning of new ideas. Along with our certainty scale for different explanations, future research can build on the methodology here to explore how successful instruction is at helping students reweight the believability of the different explanations they can construct.

Although one-on-one interviews are a common way to uncover the alternative explanations in students thinking, there are three important ways in which our multiple-choice methodology is different. First, although interviews can capture the spontaneous shifts in student explanation as they think through a problem, many interviewees may stop after giving their dominant explanation. Our approach explicitly cues the generation of alternative explanations. Second, when seeking to explicitly cue alternatives, many of these interview methodologies are designed to induce shifts in students' thinking by injecting new ideas into the discussion [6,18,19] that might refute a student's thinking or teach them something new. Our approach is more neutral, merely presenting the possible outcomes and asking students to present a possible explanation for each one. Third, our approach scales better to larger-N studies that are willing to trade-off rich, narrative detail in how students construct alternative ideas for a broader sense of how common these alternative ideas are. This approach can be easily implemented in existing research contexts where written conceptual questions are given to large groups, though the work required for coding the responses increases as the number of explanations each student generates increases.

A future area for development is on understanding the productive features of students' alternative explanations. While experts can compare students' alternative ideas to canonical physics in an ad-hoc manner based on their expert knowledge, it would be better to have a systematic framework for understanding how student reasoning aligns with (and doesn't align with) a correct physical understanding. One plausible candidate is the formalism of causal Bayesian networks [20], which can represent the complete set of causal relations between factors in a physical situation. Because the reasoning used for conceptual physics questions is often based on cause-and-effect relations (e.g., in the Two Boats Q1, the question asked is isomorphic to asking how changing the launch angle changes the time in the air), causal networks may be well-aligned with the conceptual reasoning used on such problems. We are currently exploring the use of causal networks for describing how students' incorrect explanations can become aligned with correct conceptual reasoning in physics [21].

When it comes to solving physics problems, each student brings multiple resources to the table. The approach in this paper presents an effective way to uncover and explore those resources. It can inform not just what a student is likely to answer, but also reveal the productive alternatives that may easily come to mind.

ACKNOWLEDGEMENTS

[1]   L. C. McDermott and E. F. Redish, *Resource Letter: PER-1: Physics Education Research*, Am. J. Phys., **67**, 755 (1999).

[2]   J. L. Docktor and J. P. Mestre, *Synthesis of Discipline-Based Education Research in Physics*, Phys. Rev. Spec. Top. - Phys. Ed. Res., **10**, 020119 (2014).

[3]   A. A. diSessa, *Toward an Epistemology of Physics*, Cogn. Instr., **10**, 105 (1993).

[4]   A. A. diSessa and B. L. Sherin, *What Changes in Conceptual Change?*, Int. J. Sci. Ed., **20**, 1155 (1998).

[5]   D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, *Resources, Framing, and Transfer* in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 89–120.

[6]   B. L. Sherin, M. Krakowski, and V. R. Lee, *Some Assembly Required: How Scientific Explanations Are Constructed during Clinical Interviews*, J. Res. Sci. Teach., **49**, 166 (2012).

[7]   L. Barth-Cohen, *Threads of Local Continuity between Centralized and Decentralized Causality: Transitional Explanations for the Behavior of a Complex System*, Instr. Sci., **46**, 681 (2018).

[8]   L. M. Goodhew, A. D. Robertson, P. R. L. Heron, and R. E. Scherr, *Student Conceptual Resources for Understanding Mechanical Wave Propagation*, Phys. Rev. Phys. Ed. Res., **15**, 020127 (2019).

[9]   M. E. Loverude, C. H. Kautz, and P. R. Heron, *Helping Students Develop an Understanding of Archimedes' Principle. I. Research on Student Understanding*, Am. J. Phys., **71**, 1178 (2003).

[10] R. E. Scherr, P. S. Shaffer, and S. Vokos, *Student Understanding of Time in Special Relativity: Simultaneity and Reference Frames*, Am. J. Phys., Phys. Ed. Res. Suppl., **69**, S24 (2001).

[11] B. A. Lindsey, P. R. L. Heron, and P. S. Shaffer, *Student Understanding of Energy: Difficulties Related to Systems*, Am. J. Phys., **80**, 154 (2012).

[12] J. P. Smith, A. A. diSessa, and J. Roschelle, *Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition*, J. Learn. Sci., **3**, 115 (1993).

[13] G. Gladding, M. Selen, and T. Stelzer, *FlipItPhysics* (MacMillan Education. https://www.flipitphysics.com, Retrieved on Sept. 30, 2019).

[14] D. Hammer and A. Elby, *Tapping Epistemological Resources for Learning Physics*, J. Learn. Sci., **12**, 53 (2003).

[15] J. Clement, *Students' Preconceptions in Introductory Mechanics*, Am. J. Phys., **50**, 66 (1982).

[16] J. Clement, D. E. Brown, and A. Zietsman, *Not All Preconceptions Are Misconceptions: Finding 'Anchoring Conceptions' for Grounding Instruction on Students' Intuitions*, Int. J. Sci. Ed., **11**, 554 (1989).

[17] A. Elby, *Helping Physics Students Learn How to Learn*, Am. J. Phys., Phys. Ed. Res. Suppl., **69**, S54 (2001).

[18] A. A. diSessa, *An Interactional Analysis of Clinical Interviewing*, Cogn. Instr., **25**, 523 (2007).

[19] A. Gupta and A. Elby, *Beyond Epistemological Deficits: Dynamic Explanations of Engineering Students' Difficulties with Mathematical Sense-Making*, Int. J. Sci. Ed., **11**, 554, **33**, 2463 (2011).

[20] J. Pearl, *Causality: Models, Reasoning and Inference* (Cambridge University Press, Cambridge, U.K., 2000).

[21] E. Kuo, N. K. Weinlader, B. M. Rottman, and T. J. Nokes-Malach, *Using Causal Networks to Represent the Targets of Resource Coordination*, arXiv:1909.01452 (2019).