

# Pilot Testing Dichotomous Classification Questions for Assessing Student Reasoning

Christopher M. Nakamura<sup>1</sup>, Meagan A. Donnelly<sup>1</sup>, Rajani Muraleedharan<sup>2</sup>, and Marie Cassar<sup>3</sup>

<sup>1</sup>*Saginaw Valley State University Physics Department,  
7400 Bay Road, University Center, MI, 48710*

<sup>2</sup>*Saginaw Valley State University Department of Electrical and Computer Engineering  
7400 Bay Road, University Center, MI, 48710*

<sup>3</sup>*Saginaw Valley State University Psychology Department  
7400 Bay Road, University Center, MI, 48710*

**Abstract.** Student reasoning remains an important topic in physics education. To begin studying reasoning in a simple context we pilot tested short, dichotomous classification questions. These questions present students with a physical situation and ask for classification as possible or impossible, followed by a logical justification and several Likert scale self-assessments. The questions require reasoning, but can be figured out with one key idea, and few logical steps. We developed 16 questions and tested them in clinical interviews with 22 undergraduates who had studied algebra-based or calculus-based physics through electricity and magnetism. Using qualitative and quantitative analysis we find evidence that these questions can help identify differences in students' verbalized explanations that evidence their reasoning. Using these differences we posit testable hypotheses about how subtle differences in the questions can impact students' reasoning. This pilot-testing suggests this type of question may be a useful tool for studying reasoning in physics education.

**PACS:** 01.40.Fk

## I. INTRODUCTION

Understanding student reasoning has been and continues to be a key area of interest in physics education research (PER). [1-4] Conceptual questions and problem-solving are clearly natural contexts for studying reasoning. Our current understanding of cognition suggests that how students respond to questions, and thus what we learn about their reasoning from the responses, depends on how the questions are asked and how the students are asked to respond [4]. Investigating different kinds of questions and the differences in how students respond helps us understand the interactions between student and question that shapes the responses we use to gauge understanding and reasoning.

Here we report early work investigating how students respond to three-stage dichotomous classification questions. These questions ask students to evaluate the possibility of a simple physics scenario (possible/impossible), explain how the classification was reasoned out, and then self-evaluate their responses on several metrics via 5-point Likert scales. Our motivation for studying students' responses to dichotomous classification questions comes from the notion that generating very simple reasoning tasks could provide a clearer picture of student reasoning behaviors that might then elucidate student reasoning in more complex situations, like problem solving.

Specifically we will address results associated with a subset of data from four questions that demonstrate how this type of question may be effective or ineffective for

looking at reasoning and generating testable hypotheses about student reasoning. Our primary goal in this pilot test is to provide initial insight into whether these types of questions might be useful for studying reasoning in physics and if so how to best proceed to develop and use them.

## II. RESEARCH DESIGN

### A. Question Design

Our dichotomous classification questions consist of three parts. The questions present student participants with simple scenarios relevant to their introductory physics classes. Participants are first asked to figure out whether the scenario is physically possible as described and provide that dichotomous classification. They are then asked to explain the reasoning that led to that classification. Finally, they provide a rating via 5-point Likert scales on five metrics:

1. Confidence in classification
2. Confidence in explanation
3. Interest in the question
4. Frustration in answering the question
5. Anxiety in answering the question

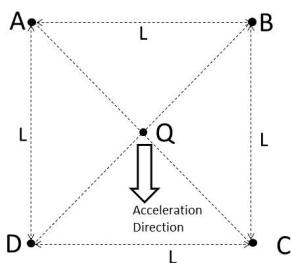
Here 5 corresponds to high levels for the metric. The metrics were chosen because of longer-term interests in assessing affect along with reasoning. Likert scales have significant limitations in terms of analysis, but represent a useful starting point for obtaining information about affect.

The last metrics may also serve as a test for unfit questions. Participants should not be more bored, frustrated or anxious than in normal academic work. Questions that elicit high negative values should be eliminated.

Our questions are conceptual and appropriate for students in calculus-based or algebra-based physics courses. We have far developed 16 questions. Eight of them focus on mechanics ideas and 8 focus on circuits or electricity and magnetism ideas. We order them so they alternate between first and second semester ideas. An example question prompt is shown in Fig. 1.

### Scenario 10

A charged particle Q is placed at rest in the middle of four charged particles A,B,C,D. A, B,C, and D are fixed in place; Q is free to move. All five charges are of the same sign, with  $A=B=2C$  and  $C=D$ . The particle Q is observed to accelerate in a downward direction. Is this possible? Why or why not?



**FIG 1.** This figure shows the manner of presentation of our questions. This scenario focuses on electrostatic forces.

There are several potential advantages offered by this question design. First, in physics, framing a question around discerning whether a situation is possible or not is very much in line with the interface of theoretical and experimental science. One is, in essence, asking a student to put together her or his knowledge of physical theory and her or his expectations of how the world works (and we know that those two things are not always the same) to ascertain whether a situation could happen as described. This is essentially a thought experiment. As a result it is relatively easy to turn most physical situations into this kind of question. The questions can be built around expectation violation, and in fact what we see in this pilot study suggests that there could be interesting opportunities to look at expectation violation at a fairly precise level.

### B. Interview Protocol

Our protocol centered on presenting the scenarios to the students via computer. After a discussion of IRB informed consent and IRB approved consent form signature, participants were presented the questions via a Microsoft Powerpoint presentation. Instructions and an example scenario were also presented in this manner, as were questions related to the participants' prior studies in physics and engineering, their primary language, and college major. This design allowed us to construct a more controlled,

uniform protocol. While a member of the research team was present for the entirety of the interview, the finalized protocol involved very little interaction with that staff member. In the earliest interviews, the staff member would ask the student for their responses verbally, ask for their reasoning explanation and confirm that the participant had provided all the information s/he wanted to include in the explanation. We found that if we used the Powerpoint presentation it could provide the questions to prompt them for their responses and that the staff member need not ask each time. We believe this results in a more uniform interview because the questions are presented each time without auditory cues or miscues.

Interviews were audio and video recorded using two software tools. Camtasia [5] was used to capture the on-screen question prompts. Simultaneously a webcam and lapel microphone were used to audio and video record the participant (audio was captured with redundancy via Camtasia and an audio recording tool called Audacity [6]).

### C. Sample Selection

Our sample consists of 22 students enrolled in algebra-based physics, calculus-based physics and upper-division science and engineering courses at Saginaw Valley State University between the summer 2014 and the winter 2015. Since our goal was to begin testing the questions we did not place emphasis on getting equal numbers of students from these groups. Volunteers were recruited through classroom visitations with instructor permission. Four algebra-based physics students, three calculus-based physics students and three upper-level science students volunteered and were compensated with \$10 giftcards. Additionally, in winter of 2015 two instructors (one of whom, Nakamura, is an author) of concurrent second semester calculus-based physics classes offered a small extra-credit incentive to participate. Another 12 students volunteered from those classes, bringing the total to 22. Interviews for those students were conducted by staff uninvolved in teaching the courses (Cassar). Alternative extra-credit assignments were offered to those uninterested in participating. Recruiting was timed to ensure that students had studied the relevant material. We collected minimal demographic information. Five of the volunteers were female; English was not the first language for four volunteers. Students were told to expect interviews to take up to one hour. Thirty-five minutes was typical. One student did not answer all questions, so in the Data Analysis section 21 responses are recorded for the last three questions.

## III. DATA ANALYSIS

A few obvious facets of the data set admit quantitative analysis. These are students' classifications of the scenarios and their Likert ratings. That said, our primary mode of analysis was qualitative. Students' explanations were analyzed thematically. The interview audio and video files

were reviewed and coded at the sentence level or response level depending on the amount of detail in the response. Commonalities and distinctions in student responses were identified that could point towards interesting aspects of how students approached the questions. These commonalities and distinctions form the basis highlighting interesting facets of student reasoning. In the results section below we describe the qualitative distinctions. The short nature of the responses coupled with the tightly focused question contexts resulted in focused responses that yielded relatively clear groups.

## IV. RESULTS

The results of our analysis are presented in this section. Table I summarizes students' classification of the scenarios. Results of the qualitative analysis of each question are presented and then results relevant to all 4 questions.

**TABLE I.** Classification results for the four questions

	Possible	Impossible	% Correct
Track	6	16	27%
Square	15	6	71%
Atwood	16	5	76%
Conductor	15	21	71%

### A. Track & Distance Runner Question

Students, presented with a closed sprinting track and an open cross-country course, were asked about the possibility of a distance runner having greater average velocity than a sprinter running one lap. Numeric values showed the sprinter's higher speed. The key observation is that the sprinter runs a closed track and has zero average velocity.

Few students correctly classified the scenario. There are two common approaches to this question: a conceptual vector approach and a scalar calculation approach. Most students, 17, approached this question from the computation perspective, calculating the speeds. One of these 17 argued that velocity was not a useful concept here and maintained a speed argument despite recognizing the point of the question. Three of the 17 used the scalar calculation perspective but miscalculated and marked the scenario possible. Three students approached the question from a vector sum perspective and concluded that the track runner's average velocity was zero. One student argued that the track shape should make no difference. One student had a conceptual approach, but not based in vector ideas.

### B. Square Electrostatics Question

In this question students were asked about the possibility of a positive charge placed in the center of a square of positive charge accelerating in a certain direction. While the center charge would be repelled by all the other charges, the resulting acceleration should be as indicated.

Thus, this scenario is possible. The exact presentation of the question is shown as an example in Fig. 1.

All students focused on the relative signs and magnitudes of the charges, though none completely stated Coulomb's law. Seven students indicated that the repulsive force from charges C and D (the smaller charges) was greater than the repulsive force from A and B. Four students mentioned an equilibrium position that the charge would move towards. One student mentioned the equidistance of the corners was important, no one mentioned symmetry as being important.

### C. Modified Atwood's Machine Question

Here students are presented with a mass on a level frictionless surface connected to a string that goes over a low-mass, low-friction pulley to a second mass that hangs vertically. They are told that the hanging mass is observed to accelerate at a greater rate than the other mass, which is not possible, assuming the string is nearly inextensible.

The scenario is impossible. Eleven students identified the connection between the masses as the reason they must accelerate together. Of these, 5 mention the string's inextensibility. The only other student who mentioned the stretching of the string did so as an argument for why it was possible for the objects to have different accelerations. Five students primarily focused on which mass was larger as a means of predicting whether one would accelerate faster. Two students focused on the tension force acting equally on each mass. Two students focused on friction determining whether the masses would have the same acceleration.

### D. Electric Field in a Conductor Question

Here students were asked about the possibility of a certain electrodynamic configuration for a conductor wired into a circuit with a battery such that current flows clockwise and the conductor has a non-zero electric field inside, pointing in the direction of the current flow.

This scenario is possible from a classical E&M perspective. In answering this question virtually all (19 of 21) students made an appeal based on how the electric field should point with respect to the current. Most said the current and field should point in the same direction a few said they should oppose each other. Two students discussed magnetic fields, suggesting they misread or misinterpreted the question. One student argued, using Gauss's law (mistaking current for flux), the electric field in the conductor should be zero.

### E. Additional Findings

Students expressed high confidence in classifications and explanations. A summary of the ratings is given in Table II. While we do not advocate averaging Likert data to make absolute statements about attitudes, it is useful for assessing people's willingness to rate in particular ways.

**TABLE II.** Mean Likert scale ratings with standard deviations for confidence in classification and explanation.

	Track	Square	Atwood	Conductor
Classification	3.7±1.2	4.2±0.9	4.0±1.3	3.1±1.2
Explanation	3.8±1.2	4.1±0.9	4.0±1.1	2.8±1.3

In general interest was above 3. Anxiety and frustration were below 2. We see no evidence that these questions are particularly onerous or tedious. Anecdotal evidence, in the form of student comments, suggests that a significant portion of students found the questions interesting.

## V. DISCUSSION

The first concern we had developing these questions was that they would be “merely recall” or fail to elicit student reasoning. Our observations in the interviews relieved this concern. It was clear that students were not behaving as if they felt that they could answer the questions by just remembering facts. Students took time to figure out the answers and vocalized their thoughts willingly.

The sprinter-distance runner question prompts an interesting question: Do the numbers in the questions prime students for a wrong approach? If so, can we change the question to prime students for the vector approach? A natural way would be to provide a purely symbolic prompt. This hypothesis can be tested empirically.

The most interesting facet of the square electrostatic question is the number of students who indicated that charges C and D were greater in magnitude. Two explanations, one mundane, come to mind. The mundane one is that they misread the question. The other is that the way the question provides the information ( $A=B=2C$ ) primes them to think of C as larger- it has the 2. This is a testable hypothesis. We can ask the question giving the information as text and look for the error. It is also interesting that Coulomb’s law was implicit or perhaps absent. The same answers could have been obtained via the primitive idea that positive charges repel, and bigger charges repel more. We are looking at ways to change the question to discern whether Coulomb’s law is being used.

The Atwood’s machine question shows that while a majority of students can recognize that the accelerations must be equal, few cite inextensibility of the string as the reason. It is an example of a question that requires physical reasoning with an idea that is not a big idea in physics. It

may be interesting for comparisons with questions that focus on big ideas. We are exploring that comparison.

The electric field in the conductor question is an example of a question that may not work well. It was written to violate the expectation that the electric field inside a conductor (importantly, but often forgotten, at equilibrium) is zero. One student did reason in that direction; most students made associations between the direction of current flow and electric field direction that led them to correct or incorrect answers without consideration of whether in fact there should be a field. One could argue that they just know when the fields should and should not be zero. We see no evidence for this. What seems more likely is that we are attempting to test an idea associated with a certain context outside that context and doing so fails to prompt students to consider the ideas we are looking for. Context matters so this question may not be very effective.

A question related to the high confidence that students report via the Likert scales is whether feedback would be useful in this type of research design. The Kruger-Dunning effect is a well-known effect in which novices rate themselves higher in competency than more-competent individuals [7]. The high confidence ratings may be an instance of this effect. Feedback might be useful in that case. The level of detail that feedback should take is not obvious and represents an additional parameter to vary.

## VI. CONCLUSIONS & FUTURE WORK

We have pilot tested questions that combine dichotomous classification, conceptual explanation and self-assessment. We have observed evidence that the questions can elicit student reasoning. We have observed clear, but unsurprising challenges that students encountered while answering the questions. We have identified ways in which the questions could be modified to test hypotheses about how question presentation may impact reasoning. Future work will focus on analysis of more data, revision of the questions and development of experimental procedures that compare response sets obtained with differing questions and that investigate the impact of feedback.

## ACKNOWLEDGEMENTS

We acknowledge A. Rupp and J. Knight for early work on question design and S. Nuss-Warren for reviewing the manuscript. C. Nakamura acknowledges support from Saginaw Valley State University via faculty research grants.

- 
- [1] B. W. Dreyfus, V. Sawtelle, C. Turpin, J. Gouvea and E. F. Redish, *Phys. Rev. ST-PER* **10**, 812 (2014).
  - [2] W. Hung, and D. H. Jonassen, *International J. of Sci. Ed.* **28**, 1601 (2006).
  - [3] W. J. Leonard, R. J. Dufresne, and J. P. Mestre *Am. J. Phys.* **64**, 1495 (1996).

- [4] E. F. Redish and E. Kuo, *Sci. & Educ.* **24**, 561 (2015).
- [5] <https://www.techsmith.com/camtasia.html>
- [6] <http://audacityteam.org/>
- [7] J. Kruger and D. Dunning, *J. Personality & Soc. Psych.* **77**, 83 (1999).