Idea use curves

Alex Axthelm,¹ Michael Wittmann,^{1,2} Carolina Alvarado² and Laura Millay²

¹*University of Maine, Department of Physics and Astronomy
5709 Bennett Hall, Orono, ME, 04469-5709*

²*University of Maine, Center for Research in STEM Education
5727 Estabrooke Hall, Orono, ME, 04469-5727*

Abstract. A variety of tools have been created to understand student performance on multiple-choice tests, including analysis of normalized gain, item response curves, and more. These methods typically focus on correct answers. Many incorrect responses contain value and can be used as building blocks for instruction, but present tools do not account for productive reasoning leading to an incorrect response. Inspired by Item Response Curves, we introduce Idea Use Curves, which relate frequency with which an idea is used to student performance. We use this tool to consider ideas which may be present in both correct responses and distractors, letting us attend more to students' conceptual understanding. This tool is made with the goal of identifying ideas that are consistently used by students who perform well or poorly, allowing researchers and instructors to look beyond the "correct/incorrect" paradigm. We explore student reasoning about energy as a proof of concept for this method.

PACS: 01.40.ek, 01.40.Fk, 01.40.gf

I. INTRODUCTION

In Physics Education Research, we have several Multiple Choice Assessments (MCAs) designed around distractors indicative of common misconceptions. These MCAs are capable of being administrated to many students and evaluated efficiently. However, because the number of options that a student can select is finite, traditional analysis of these assessments provides limited information about students' thought processes. The aim of this paper is to introduce a method that allows researchers and instructors to use existing MCAs as tools to gain further insight into student's conceptual knowledge.

This method is situated in a Knowledge in Pieces framework [1], meaning that an individual student's knowledge is multifaceted. A student may provide the wrong answer, but have partially correct reasoning. Such a student may be calling upon knowledge pieces in an inappropriate context, or making inappropriate connections between them [2].

The method is based on Item Response Theory (IRT) [3]. IRT provides information to psychometricians about the quality of a given question, by comparing the frequency of response choice with latent ability (approximated here by the total score on the assessment). This information allows the items creator to determine the efficacy of the distractors for the question. These distractors are often based on student ideas identified in prior research [3,4].

IRT parameters served as an inspiration to relate the frequency with which a student selects a given distractor with their latent ability. This relation allows researchers to identify which responses are consistently used by high scoring students, and which ones are used by students who perform poorly on the assessment overall. Responses typically in-

clude more than a single idea. The method discussed in this paper focuses on an idea, rather than a question response, as the unit of analysis of an assessment.

Other approaches, such as Model Analysis [5], offer a way to analyze MCAs, as well. The method we introduce here approaches MCAs similarly (looking for student consistency in responses) but with a different goal. While Model Analysis allows investigators to see a change in coherent models that students use, the goal of this paper is to develop a tool which allows investigators to see how ideas affect student understanding.

By exploring the ideas students use, rather than their correct or incorrect answers, a new focus is placed on their reasoning, and the process by which the students arrive at their answers. There is currently no method to explore the frequency with which students use the ideas contained in responses to MCAs. This led us to introduce a new method for exploring the ideas used by students while using MCAs.

We present a new methodology to analyze MCAs: Idea Use Curves. We introduce how the new method works and present the analysis of an item to illustrate the new method in the context of energy in middle school students. This method is presented as a proof of concept. The data and consequently the topics for analysis were chosen because the authors were already exploring student thinking on energy, however this method is generalizable to other topics.

II. ANALYZING THE USE OF ENERGY IDEAS

There exists a body of prior research on student thinking in energy, but little on the smaller-grain pieces of knowledge that students use when thinking about energy. The knowledge pieces that are analyzed in this paper are based on the existing literature, and are distinctly related to energy, rather than more generic p-prims. [6] This was a

deliberate decision, based on the data available. The tool relies on data from multiple questions with responses that can arise from the use of a given idea. We chose our ideas to analyze based on prevalence in the survey and clarity of idea in the responses. As more questions include an idea in their responses, we can be more confident that students are actually choosing to use that idea.

A. Population

We analyzed existing data from a survey of energy topics administered to middle school students. Students ranging from 7th to 8th grade were given a computer-based online survey of questions on topics in mechanical and thermal energy. The survey was administered across 3 school years (Fall 2012 through Spring 2015), as part of a larger collaborative study of middle school students understanding of energy (NSF MSP-0962805). Students were given the survey both prior to and post-instruction ($N_{\text{pre}} = 1915$, $N_{\text{post}} = 1670$, $N_{\text{total}} = 3585$). Because this analysis is not concerned with evidence of student growth and instead looks at idea use alone, we made no distinction between pre-tests and post-tests.

B. Assessment

The survey administered to students underwent revisions from year to year as goals in the overarching project were refined. In each year however, the survey included items that were drawn from the AAAS 2061 Item Bank [4] as well as some that were of original design. The AAAS items were chosen for the survey because AAAS provides the misconceptions used to design each response, which helped us identify the ideas present in each.

Due to the changes in the survey across years, not every question asked was eligible for analysis, due to incomplete data. There were six questions that were asked to all students, across all years. Four questions asked about mechanical systems, and two questions asked about thermal systems. All six questions were drawn from the AAAS item bank.

C. Ideas for Analysis

We selected two ideas to analyze: “Energy can be used up” and “Energy can be transferred”. The idea that energy can be used up is included by AAAS in the distractors of several items in the Project 2061 item bank [4]. It is common for students to employ this idea at the core of their reasoning, as a way of explaining the real-world effects of air-resistance, and other phenomena which must be “overcome.” [7] The idea that energy can be transferred was chosen in contrast to the idea that it is used up. “Energy can be transferred” is in line with expert thinking, as part of the model that energy is a substance like quantity that flows [8]. Use of this idea allows for a different type of reasoning about energy.

Our choice of these particular ideas for analysis emerged from the limitations of the assessment. These ideas were chosen because each appeared in at least two of the six questions analyzed, allowing for a more reliable analysis than ideas present in only one of the questions. Out of the 6 questions analyzed, we identified 3 questions with a total of 4 (1 + 1 + 2) responses which contain the idea that energy is used up. We identified 2 questions with a total of 5 responses (2 + 3) that contain the idea that energy can be transferred.

Note that the word “idea” in this context is deliberately left open to interpretation. This type of analysis could be applied to any idea that is present in the responses in the MCA. Caution must be exercised when choosing ideas to look for, however. Ideas that are present in too few questions limit the opportunities for students to select that response, resulting in much lower confidence that students are, (or are not) consistently using that idea.

Questions for which all responses contain a given idea were omitted from the analysis for that idea, as they do not provide information as to whether the student chooses to use that idea. For example, one item included “energy can be transferred” in each response, and so was excluded from analysis for that idea, although it was still considered for overall score.

Despite its presence in every item, we explicitly excluded the idea “Energy is conserved” from our analysis, as this idea was present only in correct responses (although it was present in every correct response). Consequently, the analysis revealed nothing new about the students, as this idea effectively was a proxy for the correct/incorrect paradigm.

A student is playing with a pendulum (a ball attached to the end of a string). He gives the ball a push and watches the ball as it swings from side to side. After a while, the ball stops swinging. Why does the ball stop swinging?

- A. The motion energy (kinetic energy) of the ball is transferred somewhere else, like the air, as the ball swings from side to side.
- B. A little bit of the ball's motion energy (kinetic energy) is used up each time it swings from side to side, but the motion energy is not transferred anywhere else.
- C. Some motion energy (kinetic energy) is transferred somewhere else, like the air, and some motion energy is used up.
- D. An object only has motion energy (kinetic energy) when a person makes it move, and the student is no longer pushing the ball.

FIG. 1: “The Pendulum Question.” Item NG065004: A pendulum stops swinging because the motion energy of the ball is transferred somewhere else, like the air, as the ball swings from side to side. Retrieved from AAAS Project 2061 Assessment Website and formatted for this paper.

An example of an item from our assessment is Item NG065004 from the AAAS item bank; hereafter referred to as “the pendulum question,” shown in Figure 1. Both responses “B” and “C” contain the idea that energy is used up. Similarly, responses “A” and “C” contain the idea that energy can be transferred. For questions where the idea is not explicitly stated in the response, we determined that an idea is present by considering if a student could reasonably construct a line of reasoning which includes the idea, and leads to the response.

It is important to note that because a response will contain multiple ideas, it is not possible to list all possible ideas that a student may call upon when providing a particular response. It is more productive to determine if a response can call upon a particular idea than attempt to list all the ideas that a response can call upon. Additionally, it is important to note that a student will call upon multiple ideas simultaneously when answering a question, and students will consistently use multiple different ideas depending on context [1,2,5,6].

III. IDEA USE CURVES

Idea Use Curves are tools that present the relation between students' overall performance (represented by overall score) and the frequency with which they select a response that contains a given idea. This relation allows for a simple visual interpretation of an idea's productivity, e.g. it represents the likelihood that a student who consistently makes use of this idea will perform well overall.

The idea use score for a given idea is calculated by dividing the number of times students selected a response that contains that idea by the number of questions that offer responses that include that idea. While comparing idea use to overall score may seem contrary to the goal of valuing student reasoning outside of a correct/incorrect paradigm, calculating overall score using the total number of questions, rather than only the questions which include the idea at hand allows us to determine how the idea contributes to a student's overall understanding of the larger topic (energy). Two examples are shown in Figure 2.

The confidence intervals for the Idea Use Curves are calculated using the Clopper and Pearson (exact) method for binomial proportion confidence, where each question which includes an idea is a trial, and each time that idea is used a success [9]. Data were analyzed using R version 3.1.2.7. The authors wrote the script that was used to perform the analysis.¹

¹ The script is available on Github at (<https://github.com/AlexAxthelm/StudentEnergySurveypublic/tree/feature/idea-curves/data-import>).

IV. RESULTS

The Idea Use Plots for the two ideas analyzed are visible in Fig. 2. Students in the top row of clusters (centered around idea use = 1.0) answered consistently using that idea, whenever it was an option. This indicates that this is likely an idea that the student pays attention to, and will continue to use.

There were no correct responses that include the idea

(a) Energy is Transferred

(b) Energy is Used Up

FIG. 2 Idea Use Curves. Area of circle indicates number of students. Red dot indicates idea use and average score for all students with given overall score. Grey band represents 95% confidence interval. Dashed lines indicate expected idea use and overall scores from random guessing. For both ideas, actual use differed from expected use ($p < 0.0001$)

that energy is used up. “Energy can be transferred” was present in the correct answers to both questions that included that idea. Therefore, there must be some correlation between total score and idea use, as it is impossible to answer every question correctly and not use the idea “Energy can be transferred.” There is no such relation for low scores, however; students can consistently use the same idea but answer every question incorrectly. This shows that these students are calling upon and connecting their ideas inappropriately. Students who perform well overall are likely to consistently use the idea that energy can be transferred, and unlikely to use the idea that the energy can be used up. This can be seen in students who did not receive a perfect score, but still scored well overall (>50%).

The idea that energy is used up was used discernably ($p < 0.01$) more than would be expected due to chance by students who scored less than 25% overall, and less than expected by students who scored more than 50%. This indicates that this idea is a very effective distractor for students who have a poor overall understanding of energy, but an ineffective one for students who do have a good understanding of energy.

Idea Use Curves can show which ideas are important for students who perform well. For example, the idea that energy can be transferred was used less than would be expected due to chance by students who scored less than 50% overall. Thus, this is a ripe idea for intervention, as students who more consistently use this idea demonstrate a better understanding of energy overall, even on questions that are not about energy transfer.

V. DISCUSSION AND CONCLUSIONS

The limited number of items in which ideas are present reduces the granularity with which we can determine the relationship between score and idea use. However, as a proof of concept, we are able to show that students who

have a good understanding of energy (as evidenced by overall score) use ideas differently than those who do not.

Idea Use Curves provide a way to look across several multiple choice questions at themes in students’ idea use and the influence of idea use on performance. This method may be useful from a pedagogical standpoint, allowing teachers to identify which ideas they should focus on with their classes. A productive idea (that perhaps is not obviously part of answering a given question) might be emphasized, while use of unproductive ideas could be addressed through careful interventions. The Idea Use Curves could help teachers quickly see the effect of idea use across a full test. In research, this method could allow question designers to see if their distractors have a common idea that is influencing the assessment as a whole. It also provides an opportunity to characterize patterns and changes in students’ idea use before and after instruction, providing far more information about student thinking than is available from using only correct/incorrect measures such as the normalized gain.

This new method shows the ideas that students are using when responding to multiple choice questions. This tool allows instructors to see which ideas are most productive for their class. By valuing the productive reasoning that students do, even if it does not lead to a correct response, teachers are better able to see their students on the pathway to thinking like a physicist. By exploring the ideas contained in responses to existing questions we can better understand our students’ reasoning, without needing to give them new instruments.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation, DRL-0962805. We thank the teachers of the Maine Physical Sciences Partnership who participated in the activity. We would also like to thank our reviewers for their valuable feedback.

[10] C. F. Herrmann-Abell and G. E. DeBoer, in *Teach.*

-
- [1] J. P. Smith III, A. a. diSessa, and J. Roschelle, *J. Learn. Sci.* **3**, 115 (1994).
 - [2] M. C. Wittmann, *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2**, 020105 (2006).
 - [3] F. B. Baker, *The Basics of Item Response Theory* (2001).
 - [4] A. A. for the A. of Science, (2015).
 - [5] L. Bao and E. F. Redish, *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **2**, 010103 (2006).
 - [6] A. a. diSessa, *Cogn. Instr.* **10**, 105 (1993).
 - [7] S. Kesidou and R. Duit, *J. Res. Sci. Teach.* **30**, 85 (1993).
 - [8] E. Brewe, *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **7**, 020106 (2011).
 - [9] R Development Core Team, (2008).

Learn. Energy K--12 Educ. (Springer, 2014), pp. 103–133.