

Developing coupled, multiple-response assessment items addressing scientific practices

Katherine D. Rainey,¹ Amali Priyanka Jambuge,² James T. Lavery,² and Bethany R. Wilcox¹

¹*Department of Physics, University of Colorado Boulder, 390 UCB, Boulder, CO 80309*

²*Department of Physics, Kansas State University, Manhattan, Kansas 66506*

Science education literature has called for blending of scientific practices with conceptual knowledge in higher education. With this move, there must also come a shift in the ways we assess students. To date, most research-based assessments place their main focus on conceptual knowledge as opposed to scientific practices, which may in part be due to the difficult nature of assessing scientific practices. Additionally, most assessment items addressing scientific practices are in free-response formats, which require in-class administration and scoring by hand, which can be onerous. Coupled, multiple-response (CMR) items pose a unique opportunity for assessing scientific practices because they elicit student reasoning while also allowing for streamlined, automated scoring. Grounded in Evidence-Centered Design, in this paper, we present the first stages in developing a generalizable process for creating CMR items that address scientific practices. We illustrate this process through an example from upper-division thermal physics.

I. INTRODUCTION

Standardized assessments allow an instructor a glimpse into what their students do and do not understand about course content and can also be used to inform curricular and instructional changes. Research-based conceptual assessments provide a standardized measure that allows instructors to compare their students' performance with students from similar courses. Though some research-based assessments in physics focus on scientific reasoning and lab skills, most focus on specific conceptual knowledge [1, 2]. However, researchers and practitioners have begun to move away from the idea of assessing content knowledge (or skills) in isolation to more authentically reflect physics as a practice [3–5]. Such a shift also requires a shift in how research-based assessments are developed and structured.

In this paper, we ground our work in Evidence-Centered Design (ECD) to present a generalizable process for developing coupled, multiple-response (CMR) assessment items that address scientific practices (SPs). SPs refer to the ways in which scientists use their knowledge to engage in science. We hypothesize CMR formats provide a promising way to assess SPs for standardized assessments that need to be distributed broadly because they provide insight into student reasoning while allowing for streamlined scoring [6]. However, there are no CMR items that explicitly address SPs in their design. Thus, there exists a need for a theory-grounded approach to develop such tasks; this paper aims to address this gap. We begin by presenting background grounding our work, including the Three-Dimensional Learning Assessment Protocol (3D-LAP) and ECD (Sec. II). We then present the process of developing CMR items addressing SPs and illustrate this process using an example from upper-division thermal physics (Sec. III). We conclude with a discussion of considerations for others when applying this process (which is in a theoretical stage) and future work (Sec. IV).

II. BACKGROUND

Presently, most conceptual assessments in physics are composed of free-response (FR) and/or multiple-choice (MC) items [1]. MC (single correct response per item) and multiple-response (MR, potential for multiple correct or partially correct responses per item) formats allow for more streamlined scoring of student responses, giving them some significant advantages with respect to broad adoption. Additionally, there is potential for these assessments to be administered in an online format [7] outside of class time, making them advantageous over purely FR formats by reducing the need for class time and allowing for automated scoring. However, traditional MC and MR tasks typically give little to no insight into the thought processes or reasoning invoked by students in order to come to their answer. This is a problem particularly for assessing SPs, as the process by which students come to their conclusion is key.

Coupled, multiple-response (CMR) items can bridge the

gap between MC and FR items [6], providing more insight into student reasoning while maintaining the more advantageous scoring mechanism of MC. This is of particular importance for broad, large-scale implementations of an assessment, in which a FR item would be cumbersome to analyze and score quickly and consistently. CMR item structure typically involves two or more parts. The first part of the item poses a MC question. The following part(s) of the item pose questions invoking student reasoning—students are asked what approach they used or to select one or more provided reasoning elements that mirror the reasoning they used to answer the MC prompt. This format presents an opportunity for assessing SPs because one can ask about students' conclusions *and* the reasoning they used to reach it, aligning much more with SP engagement than traditional MC allows.

Researchers and practitioners are increasingly interested in what students know and how they can use that knowledge [3, 8]. In addition to this college-level focus, the Next Generation Science Standards (NGSS) for K-12 education adopts a holistic view of science understanding and engagement that integrates disciplinary core ideas (e.g., forces and motion), crosscutting concepts (e.g., patterns), and science and engineering practices (e.g., developing and using models) [4].

NGSS recommends these “three dimensions” of learning (core ideas, crosscutting concepts, and science and engineering practices) be integrated in all aspects of science learning, from instruction and curriculum to assessment. One approach to developing assessments such as these is to utilize the 3D-LAP, which is meant to help assessment tasks align with these new efforts towards integrating three-dimensional learning [5]. In particular, it was designed to support and characterize development of three-dimensional *college-level* assessment tasks in physics, chemistry, and biology. Here, we focus mainly on scientific practices (SPs), which refer to the ways in which scientists engage in their disciplines, such as developing and using models or using mathematical thinking. We note that the criteria articulated by the 3D-LAP are individualized for the particular SP being targeted with the assessment task and vary based on the goals of the item. Thus, the criteria we use for the item we present in this paper would differ for tasks addressing different SPs.

Another approach to developing assessment items integrating SPs is to utilize ECD, which emphasizes the relationship between learning goals (claims of what students know and can do), evidence such as observations or performances, and task features. Following Harris *et al.* [9], an initial step for ECD is to develop a knowledge-in-use (KiU) table (not shown due to space limitations), which is composed of five rows. The first row describes a learning performance (LP) that describes what students should be able to do with content. The second row describes focal knowledge, skills, and abilities (KSAs). The third row defines evidence required in student responses to demonstrate proficiency with the learning performance (from row 1), which we refer to as evidence statements (ESs). The last two rows of the KiU table define task features characteristic to all other tasks addressing the same

learning performance within the assessment and variable task features unique to the individual task being created. In this work, we focus mainly on the first three rows of the KiU table to utilize ECD, informed by the 3D-LAP, to create rich CMR assessment tasks addressing SPs.

III. DEVELOPING CMR ITEMS ADDRESSING SCIENTIFIC PRACTICES

Here, we outline a generalizable process for developing CMR items that explicitly address SPs blended with scientific concepts. First, we discuss important considerations for developing, and eventually implementing, CMR items addressing SPs. This includes addressing issues related to scaffolding, formatting, and the necessity of explicit directions.

Due to the nature of SPs, students should be allowed to engage with the question and targeted SP independently. If one wants to measure how and if students are truly engaging in scientific practices when solving a problem, the item should allow them to consider the prompt without any assistance or scaffolding [4, 10]. A side effect of scaffolding is the possibility of students looking ahead and seeing possible approaches before attempting their own solution, therefore resulting in responses that may not reflect the actual process students would invoke if they were asked to solve the problem independently. This is an important consideration because SP-engagement can involve several different paths to come to the same conclusion, and assessment developers would want to avoid forcing a particular type of response or approach.

We conclude it is best to *avoid* scaffolding within these types of assessment items, at least initially. Fortunately, administering these CMR items online provides a mechanism to avoid such scaffolding. By utilizing this format, different parts of the CMR item can be separated into multiple pages to prevent students from seeing possible options before trying the problem on their own. This requires removal of the “back” button, to prevent students from changing their initial answer after seeing possible approaches.

The process presented here includes unpacking content; writing and piloting a FR item; and writing and refining a CMR item. We note the process presented has been modified and is not exactly the process we followed; however, changes in the process have been made to better align with that of Harris *et al.* [9]. We also present an example CMR item developed for inclusion in an upper-division thermal physics assessment (see Fig. 1). The item addresses concepts of energy flow, entropy, and the SP of “using mathematics.” We focused on this content because energy flow and entropy are core concepts in upper-division thermal physics and “using mathematics” was identified by physics instructors as a valued practice in these courses [11].

1. Unpack the content. The process begins by identifying and unpacking target content for a particular item or task. We worked with concept-based objectives (developed prior to this work) and SPs. From this, one should construct a KiU table (not shown due to space limitations), as described in depth by

Harris *et al.* [9].

Our thermal physics item development began with the LP: “Students will be able to use math to investigate the energy flow direction of two substances in thermal contact by maximizing the entropy of the system.” KSAs for our example included identifying the second law of thermodynamics as applicable to the phenomenon, identifying spontaneous energy flow direction, and specific mathematical knowledge. All authors then worked together to articulate four ESs:

- ES1 A statement that identifies energy flows spontaneously such that the entropy of the system is maximized and energy flow into a system increases its temperature and energy flow out from the system reduces its temperature (in the absence of work).
- ES2 Use of given entropy-temperature relation to see how entropy changes with temperature.
- ES3 Simplified mathematical expressions.
- ES4 A statement of the energy flow direction.

Articulating these ESs is key for item development as these statements serve as a baseline for constructing provided CMR selection options (i.e., CMR selection options need to span the space of all ESs). These ESs are shown and identified within the item in Fig. 1. Characteristic task features across SP-targeted assessment items stem from the 3D-LAP criteria for the SP of “using mathematics” [5]. In particular, our table and resulting item align with the criteria of (1) giving an event or phenomena (objects in thermal contact); (2) asking students to use a mathematical representation (given mathematical expressions relating entropy and temperature); and (3) asking students to select a consequence of their results (direction of spontaneous energy flow). Variable task features include specification of the amount of scaffolding, but this component of the KiU table only makes sense when multiple tasks are being constructed, which we have not done yet.

2. Write the free-response item. The next step in the process is to write a FR item from the KiU table. One must ensure the prompt addresses KSAs and would evoke the ESs necessary to demonstrate student proficiency with the articulated LP. For a more in-depth synopsis of this process, see Harris *et al.* [9]. In the FR prompt, it is important to include explicit expectations for students (e.g., “*Explain your reasoning*” or “*articulate and justify any assumptions you make*”); this is helpful when developing the CMR item, as students’ responses to the FR prompt are used to develop the CMR item.

For our FR item, we used the phenomena of spontaneous energy flow between items in thermal contact. The prompt itself gives mathematical relationships between entropy and temperature for the two objects. Then, it requires students to identify that energy flows spontaneously such that the system entropy is maximized, *and* utilize mathematical methods to determine what direction of energy flow would accomplish that. We note that the idea of maximizing entropy is not directly elicited in the prompt, and the requirement of invoking mathematics is only implied via the prompt providing mathematical expressions of $S(T)$. The FR item matched closely to

Page 1: Experimental measurements of the heat capacity at constant volume of an unknown substance (Substance A) were used to find its functional dependence of entropy with respect to temperature:

$$S_A(T) = aT^2,$$

where a is a positive constant coefficient and T is the temperature. A different substance (Substance B) is found to have the entropy-temperature relationship:

$$S_B(T) = bT^{-2},$$

where b is a positive constant coefficient. If Substances A and B are close to (but not exactly) the same temperature and put in thermal contact, with neither allowed to expand or contract, which way would energy spontaneously flow?

a) from A to B
b) from B to A
c) energy would not flow
d) depends on the temperature of the two substances
e) impossible to determine

ES4

ES1: A statement that identifies energy flows such that the entropy of the system is maximized and energy flow into a system increases its temperature and energy flow out from the system reduces its temperature.

ES2: Use of given $S(T)$ relations to see how entropy changes with temperature.

ES3: Simplified mathematical expressions.

ES4: A statement of the energy flow direction.

Page 2: Which of the following ideas *did you use* to approach this problem? (*select all that apply. If you did not use an option below to answer the previous problem, do not select that option, even if it holds true otherwise.*)

ES1

☐ energy flows from high to low temperatures
☐ energy flow from one substance (1) to a second substance (2) would increase the temperature of (1) and decrease the temperature of (2)
☐ energy flow from one substance (1) to a second substance (2) would decrease the temperature of (1) and increase the temperature of (2)
☐ expressions for entropy can be used to *directly* find direction of energy flow
☐ entropy and energy flow are unrelated
☐ total entropy should decrease after spontaneous processes
☐ total entropy should increase after spontaneous processes
☐ total entropy should not change after spontaneous processes
☐ no energy is transferred through work because no work is done
☐ information about work is not provided but is needed to determine energy flow
☐ information about the temperatures of the substances is needed
☐ other: _____

Page 3: Which of the following relations *did you use* to answer this problem? (*select all that you used*)

ES2 **ES3**

☐ $dS_A > 0$ when $dT_A > 0$	☐ $dS_B > 0$ when $dT_B > 0$	☐ $dU = TdS - PdV + \mu dN$	☐ dS_{universe} is dependent on dS_A and dS_B
☐ $dS_A > 0$ when $dT_A < 0$	☐ $dS_B > 0$ when $dT_B < 0$	☐ $W = 0$	☐ dU depends on work (W) and heat (Q)
☐ $dS_A < 0$ when $dT_A > 0$	☐ $dS_B < 0$ when $dT_B > 0$	☐ $\frac{1}{T} = \frac{dS}{dU}$	☐ other: _____
☐ $dS_A < 0$ when $dT_A < 0$	☐ $dS_B < 0$ when $dT_B < 0$	☐ $Q \propto TdS$	
☐ $dS_A = 0$ when $dT_A > 0$	☐ $dS_B = 0$ when $dT_B > 0$	☐ $dU \propto dT$	
☐ $dS_A = 0$ when $dT_A < 0$	☐ $dS_B = 0$ when $dT_B < 0$		

FIG. 1. An example CMR item developed to address SPs in upper-division thermal physics, with annotations of where ESs are addressed. Each box represents a portion of the question that would be displayed on a separate page. This item would need to be administered via an online platform with no “back” button, and each page would include a note for students stating they will not be able to return to the page once their answer is submitted. We note the prompt on Page 1 would be reiterated on each subsequent page; we exclude them due to space limitations. Additionally, we note the formatting of answers on Page 3 has been adjusted due to space limitations.

Page 1 in Fig. 1, though it did not include the MC options.

3. Pilot the free-response item. Once the FR task is finalized, it must then be piloted to a group of students in a pencil-and-paper format. This format is preferable because some tasks may elicit students (explicitly or otherwise) to create figures, diagrams, graphs, or mathematical representations, all of which can be considered important evidence of SP engagement. A goal of this FR pilot administration is to see what students do to answer the prompt—to see how they apply what they know—to inform development of the CMR item.

After our FR item was finalized, it was piloted in an upper-division thermal physics course on the last day of class; we re-

ceived 32 student responses. We recommend getting enough responses to ensure you capture the majority of common ideas/response patterns for that population, as they will be used for CMR item development. After administration, responses to the FR item are reviewed and coded [12]. Common answers and approaches are then summarized based on these codes and used to inform the CMR item development.

4. Write the coupled, multiple-response item. Using the original FR prompt, the KiU table, and the common approaches used by students, a CMR item is developed. As mentioned in Sec. II, a CMR item is typically composed of a MC question followed by one or more MR questions. The

number of MR questions within the item depends on the ESs. For example, our thermal physics item contained three key components in the required evidence: an answer (ES4), physical reasoning (ES1), and use of mathematics (ES2 and ES3). Thus the questions within the item are composed of (i) a MC question asking for the answer, (ii) a MR question asking for physical reasoning invoked, and (iii) a MR question asking about mathematical approaches utilized.

Each MR question has possible selections, or *reasoning elements* [6], that are informed by common approaches used by students and the ESs. For example, many students knew temperature differences cause energy flow and attempted to find an expression for $T(S)$ to determine which object had a higher temperature; thus, a MR option on Page 3 of the item (see Fig. 1) included an option to select the mathematical definition of temperature ($1/T = dS/dU$). Additionally, to address ES1, Page 2 included an option to select “total entropy should increase after spontaneous processes.” We note that, though the list of MR options within the item are long, it has been shown that long lists of options in CMR items do not significantly change response patterns compared to FR items [6].

From the analysis of student responses to the FR item, three common approaches to determine the direction of energy flow emerged: (1) maximizing system entropy by varying T for each object (via partial derivatives, varying T directly, etc.); (2) setting work equal to zero and using the thermodynamic identity to determine the sign of dU for each object; and (3) using the mathematical definition of temperature and invoking the equipartition theorem to write an expression for $T(U)$ to determine which object has a higher temperature. Approach (1) utilizes appropriate reasoning strategies and results in the correct direction of energy flow. Approach (2) does not make physical sense (e.g., utilizes mathematical relationships inconsistent with the physical situation) but results in the correct direction of energy flow. Approach (3) results in a non-physical (i.e., imaginary) expression for temperature and no answer for the direction of energy flow.

When writing the CMR item, we ensured that all three of these possible approaches were reflected in possible combinations of MC and MR options by taking all of the elements from student responses and making them possible options to our questions within the item. As Fig. 1 shows, each page of the full item encompasses each part of the possible approaches: an answer for energy flow direction in MC form (Page 1, ES4), an articulation of physical reasoning utilized to determine their answer in MR form (Page 2, ES1), and an articulation of mathematical approaches utilized in MR form (Page 3, ES2 and ES3). We note that often, many MR options within a CMR item may hold true, but only *some* are utilized by students when solving the problem. To address this when applicable, we suggest an explicit prompt after the question directing students to only select responses corresponding with their method used to answer the MC question, such as “if you did **not** use an option below to answer the previous problem, **do not select that option**, even if it holds true otherwise.”

After an initial CMR version has been prepared, the item should be piloted via student interviews. This is done to ensure the task is interpreted the intended way and that lines of reasoning used by students are accurately captured by the provided CMR options. Using results of these interviews, the CMR item can be adjusted and modified. After this refinement process, the item can be administered more broadly; this falls in the realm of future work for our presented example.

Comments on the process. Note that the process described above, at various points, is iterative. Throughout the process all components, such as content, formatting, and item refinement, should be discussed frequently and revisited. One component of the process worth particular, iterative attention is the KiU table. With the table being constructed prior to a pilot administration of a FR version, the required evidence may change based on what students actually say. For example, students may have approaches not considered by item developers during initial development of the table that should be included as possible ESs prior to constructing the CMR item.

IV. CONCLUSION & FUTURE WORK

Built on a foundation of Evidence-Centered Design, we outlined a generalizable process for developing assessment items that address scientific practices in a coupled, multiple-response format. Though the full extent of the process’s generalizability has yet to be explored in depth, the process of unpacking content, writing and administering a FR item, and turning that item into a CMR item presented here outlines a process that can act as a foundational reference for others.

We have yet to fully apply the presented process to create items that target other SPs, though initial efforts are underway. One example of a challenge faced in applying the described process is the variable nature of the tasks, which is in part determined by the specific SP being targeted (e.g., constructing explanations). Additionally, the approach for constructing the CMR item, as well as the eventual online structure, will vary based on the types of responses provided by students on the FR version and the nature of the prompt, presenting another challenge.

This work is ongoing and the example presented has not been finalized. To finalize the item in Fig. 1, we will conduct student interviews with the item and use those results to refine it before implementing it more broadly in an online format. After the online administration, student responses will be analyzed and used to inform further revisions of the item.

ACKNOWLEDGMENTS

We thank M. Vignal for his input in item revisions and the students who participated in the FR pilot. This work was funded by the Center for STEM Learning at the University of Colorado Boulder and the Department of Physics at Kansas State University. This material is based upon work supported by the National Science Foundation under Grant No. 1726360.

-
- [1] Physport: Supporting physics teaching with research-based resources, <https://www.physport.org/> (2020).
 - [2] J. T. Lavery and M. D. Caballero, Analysis of the most common concept inventories in physics: What are we assessing?, *Physical Review Physics Education Research* **14**, 010123 (2018).
 - [3] J. Kozminski, H. Lewandowski, N. Beverly, S. Lindaas, D. Deardorff, A. Reagan, R. Dietz, R. Tagg, J. Williams, R. Hobbs, *et al.*, Aapt recommendations for the undergraduate physics laboratory curriculum, *American Association of Physics Teachers*, 29 (2014).
 - [4] National Research Council *et al.*, *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (National Academies Press, 2012).
 - [5] J. T. Lavery, S. M. Underwood, R. L. Matz, L. A. Posey, J. H. Carmel, M. D. Caballero, C. L. Fata-Hartley, D. Ebert-May, S. E. Jardeleza, and M. M. Cooper, Characterizing college science assessments: the three-dimensional learning assessment protocol, *PloS one* **11** (2016).
 - [6] B. R. Wilcox and S. J. Pollock, Coupled multiple-response versus free-response conceptual assessment: An example from upper-division physics, *Physical Review Special Topics-Physics Education Research* **10**, 020124 (2014).
 - [7] B. R. Wilcox and S. J. Pollock, Student behavior and test security in online conceptual assessment, in *Physics Education Research Conference Proceedings*, PER Conference (Provo, Utah, 2019).
 - [8] S. Singer and K. A. Smith, Discipline-based education research: Understanding and improving learning in undergraduate science and engineering, *Journal of Engineering Education* **102**, 468 (2013).
 - [9] C. J. Harris, J. S. Krajcik, J. W. Pellegrino, and A. H. DeBarger, Designing knowledge-in-use assessments to promote deeper learning, *Educational Measurement: Issues and Practice* **38**, 53 (2019).
 - [10] J. W. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing what students know: The science and design of educational assessment*. (ERIC, 2001).
 - [11] K. D. Rainey and B. R. Wilcox, Faculty survey on upper-division thermal physics content coverage, in *Physics Education Research Conference Proceedings*, PER Conference (Provo, Utah, 2019).
 - [12] V. Braun and V. Clarke, Using thematic analysis in psychology, *Qualitative research in psychology* **3**, 77 (2006).