Assessment feedback: A tool to promote scientific practices in upper-division

Amali Priyanka Jambuge,[1] Katherine D. Rainey,[2] Bethany R. Wilcox,[2] and James T. Laverty[1]

[1]*Department of Physics, Kansas State University, Manhattan, Kansas 66506*
[2]*Department of Physics, University of Colorado Boulder, 390 UCB, Boulder, CO 80309*

Students' scores on assessments play a vital role in course modifications, though their effectiveness relies on the quality of the interpretation of these scores. We adapt the notion of assessments as a change agent so that a well-developed rubric accompanied by intentionally designed instructor feedback can act as a tool to inform course improvement. In conjunction with work developing a standardized upper-division thermal physics assessment, this pilot work articulates a methodology to determine feedback for instructors to inform how well their courses support students in meeting learning goals. In this paper, we present an example using a task targeting the scientific practice of "using mathematics" to explicate this methodology. This work highlights the importance of assessment feedback to inform explicit course modifications in physics.

I. INTRODUCTION

Assessment results can provide evidence of what students have learned from a particular course. Yet assessment results should not stand alone; they should be interpreted and used to improve curriculum and instruction [1]. We often tend to summarize our students' achievement in scores, and we interpret higher scores as indications of successful instruction. Unless we deliberately pay attention to how scores relate to student learning, it is obscure what aspects of learning are reflected in these scores. Therefore, reflecting on what these scores tell us about student learning and using them to improve our courses is an essential aspect of the instructional practice.

The tendency to emphasize scores is encouraged by the current standardized assessments widely being used in physics. Hake [2], Freeman *et al.* [3], and Von Korff *et al.* [4] used concept inventories as probes to determine the effectiveness of instructional methods, i.e. interactive engagement over traditional. They found that normalized gain calculated based on the average of the students' pre- and post-test scores in interactive engagement classrooms is larger than in traditional classrooms, concluding that instruction related to interactive engagement is better at facilitating students' learning over traditional, lecture-based methods.

However, increases in scores do not necessarily communicate what works or what does not work, particularly for an individual faculty member, i.e. how scores tell instructors what they should change about their instruction (providing *actionable feedback* to instructors; henceforth referred to as simply *feedback*). Although standardized assessments like the FCI [5] are feasible to use in large classroom settings, they are not designed to provide descriptive, actionable feedback to improve the course. According to the work by Madsen *et al.* [6], physics faculty want a deeper understanding of what the numerical scores of research-based assessments convey, and how they can interpret the scores to make concrete changes to their teaching to better assist student learning.

In this paper, we present pilot work to provide instructors feedback around their students' performance conducted as part of a larger project to develop a summative standardized assessment for upper-division thermal physics. This researcher-generated feedback specifically aligns with how well the course supports students in meeting learning goals. Interpreting students' performances to determine the extent to which they meet the learning goals could be challenging for instructors, especially when it comes to making sense of a new aspect of students' learning (such as scientific practices [7]). We explore how we can synthesize students' written work to give instructors feedback about students' performance and how they can modify their course to better assist students in achieving particular learning goals. We build on the work conducted by Harris *et al.* [8] and Stephenson *et al.* [9] to articulate a methodology to determine feedback that goes beyond reporting numerical scores to the instructors of the course.

This work is an initial effort to articulate a theory-based methodology for developing and delivering instructor feedback. To illustrate this, we give an example using upper-division student responses to free-response tasks developed during the initial phase of task development. This methodology will be further refined in the future for the process of developing feedback for coupled, multiple-response assessment tasks [10] to efficiently support the use of future assessments across institutions nationwide with streamlined scoring. In the remainder of the paper, we explain our theoretical approach leveraging Evidence-Centered Design [11] to design assessment tasks that address scientific practices, discuss our data collection and analysis, and articulate the methodology for developing the rubric and the feedback to instructors.

II. THEORETICAL APPROACH

There are recent calls to address scientific practices in college-level courses [12–15]. Scientific Practices (SPs) are the practices scientists engage in to explore real-world phenomena. Including them in college courses emphasizes bringing knowledge closer to its usage in a way that scientists do. Blending SPs with concepts into curriculum, instruction, and assessment provides students avenues for deep learning[7].

Matz *et al.* [16] hypothesized that if a course is transformed to assimilate this new view of bringing knowledge closer to its usage, the concomitant assessments should also change. Developing assessment tasks to incorporate SPs is challenging [12], however. Here, we argue the other way around, i.e. we can transform the nature of the course based on assessment results. In other words, the assessment itself can be used as a driver to promote, and help faculty incorporate SPs into their courses.

Evidence-Centered Design has been recommended as a theoretical underpinning for assessment task design that has the potential to elicit students' abilities to engage in SPs [17]. Evidence-Centered Design consists of several steps to ensure coherent assessment task design. We hereby articulate those steps adapting the work of Harris *et al.* [8] to develop assessment tasks addressing SPs using Evidence-Centered Design. In step 1, we need to define the assessable statements of what students should know and be able to do with that knowledge, i.e. *learning performance (LP)*. Step 2 involves articulating a set of proficiencies, i.e. *Knowledge-Skills-Abilities (KSAs)*, to demonstrate the LP. In step 3, we need to state the observable features of the student responses to argue that they meet the LP in the form of *Evidence Statements (ESs)*, aligning with the KSAs. In the final step, the *task features* should facilitate students to elicit the target KSAs.

The Three-dimensional Learning Assessment Protocol (3D-LAP) [13] contains criteria to determine the task features to elicit evidence of students' abilities to engage in SPs. In the 3D-LAP, each SP consists of criteria that should be collectively justified for an assessment task to have the potential to elicit that SP. For example, to elicit students' abilities to en-

TABLE I. Here we define and list the Learning Performance (LP), Knowledge, Skills, and Abilities (KSAs), and Evidence Statements (ESs) that align with the task in Fig. 1. The * represents modifications made after going through student responses to the task.

Step	Description	Examples for the Task Shown in Fig. 1	
LP	Assessable statements of what students should know and be able to do with that knowledge.	Students will be able to use math to investigate the heat flow direction of two substances in thermal contact by maximizing the entropy of the system.	
KSAs	The proficiencies to be targeted by the assessment task.	KSA1	Identify the second law of thermodynamics as an appropriate concept in understanding the direction of energy flow between two systems in thermal contact and identify heat flows into a system, increases its temperature and heat flows out from the system reduces its temperature.
		KSA2	Identify the relationship between the change in temperature, and the change in entropy.
		KSA3	*Accurate mathematical manipulations.
		KSA4	Identify the heat flow direction.
ESs	Observable features of the students' performances.	ES1	A statement that identifies heat flows such that the entropy of the system is maximized and heat flow into a system increases its temperature and heat flow out from the system reduces its temperature.
		ES2	Use of given entropy-temperature relation to see how entropy changes with temperature.
		ES3	*Simplified mathematical expressions.
		ES4	A statement of the heat flow direction in the given problem.

gage in the SP of Using Mathematics, the task should 1) give a phenomenon, 2) ask students to demonstrate a relationship between parameters, and 3) ask students to give an interpretation of their answer. The phenomenon can be associated with a target concept (e.g., entropy), thereby gaining the form of task features that has the potential to elicit student abilities to blend SP with concepts.

Moreover, the task design is accompanied by the rubric development. The ESs lay out the foundation for the rubric development so that they collectively describe the extent to which students meet the LP [8, 9]. However, tasks and the rubric should be validated with student responses to explore if they are working as intended. This validation process typically leads to iterative modifications of the task and the rubric.

In the following section, we closely follow the theoretical approach we presented here with the goal of developing feedback to instructors about their course improvement with regards to both concepts and scientific practices.

Experimental measurements of the heat capacity at constant volume of an unknown substance (Substance A) at low temperatures 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 put in thermal contact and neither is allowed to expand or contract, which way would heat flow? Show your work and/or explain your reasoning.

If needed, the thermodynamic identity is $dU = TdS - PdV + \mu dN$.

FIG. 1. Assessment task designed to elicit evidence that students are achieving the LP articulated in Table I.

III. THE FEEDBACK DEVELOPMENT PROCESS

Before turning to our feedback mechanism, we explicate our task design process addressing SPs, the data collection approach, and the design process of the rubric that the feedback is rooted in.

The task development team used a survey to identify three SPs – Using Mathematics, Using Models, and Constructing Explanations – as practices valued by instructors who teach upper-division thermal physics nationwide [18]. We also identified entropy as a primary content focus among those instructors. Thus, the initial phase of task development included tasks that have the potential to elicit students' abilities to engage in those SPs with the concept of entropy. However, in this paper, we use the task addressing Using Mathematics

with the concept of entropy to demonstrate the methodology for developing instructor feedback. The presented methodology does not reflect the sequence we exactly followed during our research process; we have modified, and optimized the process based on our research experience. We, the authors, collectively discussed and refined the rubric and feedback.

Task Design: Figure 1 shows the assessment task aligned with the following LP: "students will be able to use math to investigate the heat flow direction of two substances in thermal contact by maximizing the entropy of the system." The target KSAs required to elicit this LP that we set forth prior to developing the task are shown in Table I. The corresponding ESs required to demonstrate those KSAs are also shown in Table I. To elicit the KSAs to meet the LP, we used the task features from the 3D-LAP as explained in Sec. II, along
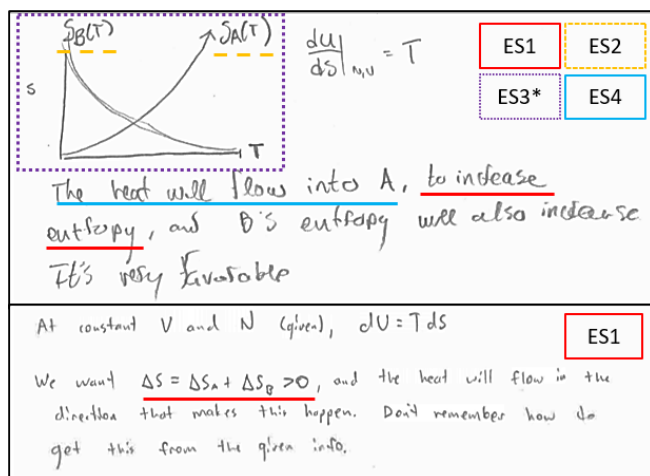
FIG. 2. Two sample student solutions, one with a rating of 4 (top) and another with a rating of 2 (bottom). Notes in the figure indicate which aspect of the student's solution counts as evidence (e.g., ESs).

with the phenomenon related to the "heat flow between two systems in thermal contact." For example, the "use of given entropy-temperature relation" (ES2) provides evidence that students have "identified the relationship between the change in temperature to change in entropy" (KSA2).

Piloting the task with students required us to modify, and add, another KSA and ES (specifically, ES3* in Table I) to accommodate unanticipated patterns in students' work. This iterative refinement of the KSAs and ESs based on student work is necessary to ensure that the rubric and the feedback capture all relevant aspects of students' work. Additionally, we ordered the ESs such that ES2 is built on ES1, ES3* is built on ES2, and so on. We followed this in light of capturing students' coherent problem-solving patterns.

Data Collection: The data for this pilot work were drawn from 32 students in an upper-division thermal physics course at a public research university in the US. We collected, and scanned their written solutions to the task in Fig. 1 as part of a pilot administration of an upper-division assessment. The students were given credit for their participation only.

Rubric Design: Each rubric component shown in Table II measures a specific ES, which builds from its corresponding KSA. We applied each ES to the students' solutions to examine the features of their KSAs. For example, the graphical representation of entropy vs temperature to visualize the change in entropy with temperature emerged as evidence to support KSA2. Further, taking derivatives, partial derivatives, variable manipulations, and substitution of expressions emerged as evidence to support KSA3*. We then modified the KSAs accordingly and assigned ratings from 0-4 that maximally capture the student problem-solving patterns.

These ratings highlight the levels of sophistication needed to meet the LP based on students' KSAs and their connection with coherent problem-solving. Thus, ratings give rise to the extent students meet the LP. A rating of 4 on the rubric indicates that students met the LP as their responses captured all the required ESs. On the other hand, a rating of 0-3 indicates that students' responses did not include one or more ESs in Table I to fulfill the LP. Figure 2 showcases coded sample students' solutions attributed to ratings 4 and 2 respectively. According to the rubric in Table II, a student solution with a rating of 4 meets all the required ESs to meet the LP and a student solution with a rating of 2 does not display evidence for ES2, ES3*, and ES4 to meet the LP.

Instructor Feedback: Each rubric rating has a piece of feedback associated with it to help instructors modify their instruction. These suggested modifications are attributed to the specific recommendations that facilitate students towards meeting the LP. The way we determine which feedback to give to an instructor is to identify the most common rubric ratings. Then, we can locate the ESs yet required to fulfill the LP. As each ES is built on the previous ESs, we prioritize the lowest order ESs (e.g., if students needed support in both ES2, and ES3 to meet the LP, we focus on ES2 first as ES3 should be built around ES2 for coherent problem-solving), and thereby the corresponding KSAs that the students should be given more opportunities to engage in.

For example, if the most common rubric rating is 2, giving students more opportunities to intertwine the appropriate concept they unpacked–heat flows such that the entropy of the system is maximized–with mathematics will help them achieve the LP. On the other hand, if the most common rubric rating is 4, the instruction supported students to meet the LP and, thus, the existing instruction works well for most students to achieve the LP, suggesting no course modifications pertaining to the LP are needed (see rating 2, and 4 in the rubric, provided in Table II to trace the KSAs students demonstrated).

To operationalize the feedback for an actual classroom, we showcase a prototype of the feedback that would be given to the instructor of the course from which the the data for this study was drawn. The portion of the student population that fell into each rating from 0-4 are 25%, 38%, 3%, 6%, and 28% respectively; 28% of the students meet the LP while the rest (72%) require one or more ESs to fulfill the LP. Out of the 72%, the majority (38%) of the students received a rating of 1, requiring evidence corresponding to one or more KSAs to meet the LP as explained with more details below.

This 38% of students demonstrates their ability to identify the relationship between change in temperature to change in entropy (KSA2), thereby showcasing ES2. However, the relations they developed, is not built around the targeted concept–heat flows such that the entropy of the system is maximized, i.e. KSA1, and thus requiring ES1 to satisfy coherent problem-solving. As stated earlier, we prioritize the lowest order ESs that give rise to coherent problem-solving, and thereby the corresponding KSAs. Thus, this suggests this population of students would benefit from the specific changes to curriculum and instruction that intertwine KSA1, and KSA2.

TABLE II. The rubric to inform course improvement (We only include elements corresponding to scores of 2 and 4 due to space limitations).

Rating	Rubric Components			
	ES1	ES2	ES3*	ES4
4	Students unpack that heat flows such that the entropy of the system is maximizing, AND Heat flows into a system, increases its temperature and heat flows out from the system reduces its temperature	Build up accurate mathematical relations for how entropy is changing with the change in temperature, OR Build up accurate graphical representations that visualize the change in entropy with the temperature	Accurate mathematical manipulations such as taking derivatives, OR Partial derivatives, OR Variable manipulations, OR Substitution of mathematical expressions, OR Not required (if students used the graphical representation approach)	Heat flows from B to A as it's entropically favorable for the entropy of substance B to increase when the temperature is decreasing and entropy of substance A to increase when the temperature is increasing
⋮				
2	Students unpack that, heat flows such that the entropy of the system is maximizing	No mathematical relation built up for how entropy is changing with the change in temperature	No mathematical manipulations	Heat flow direction not expressed
⋮				

Giving this population of students more opportunities to explore the applicability of the second law of thermodynamics to determine the direction of heat flow of two systems in thermal contact would help them reach the trajectory towards meeting the LP. Further, emphasizing that the mathematical relations connecting change in temperature to change in entropy should be developed in light of appropriate concepts needed to determine the heat flow direction–heat flows such that the entropy of the system is maximized–will be beneficial for this population of students.

Additionally, these recommendations would be beneficial for the portion of the student population that fell into the next most common rating (25%), who needed help indeed to begin with the task itself. Though we lay out the recommendations for the instructor of this course, we leave it to the instructor to determine how to operationalize these suggested modifications in their own classroom as per their teaching philosophy.

IV.   CONCLUSION AND FUTURE WORK

As initial efforts to give instructor feedback about how well their instruction supports students to meet learning goals, we articulate a methodology to develop a rubric with accompanying feedback. The rating produced by the rubric represents a measure of the extent to which students meet the LP. We argue that this rubric has the potential to interpret students' abilities to meet the LP and that, if the instructor values that LP, the course should also be modified along the line of facilitating students toward meeting that LP. We have also extended this methodology to articulate instructor feedback to better aid students' abilities to meet the LP aligned with the SP of "Using Models", and the methodology shows promising flexibility to transfer across different SPs. However, more work is needed to check its flexibility across other SPs.

The current format of the suggested feedback is based on the most common rating students hold. More work is needed to articulate the ways feedback can be given based on different distributions of student ratings.

While this feedback mechanism helps us to characterize student responses in our dataset, the generalizability of this in characterizing student responses in a different data set also requires more research. One potential for future work is to expand the student population of our data set and collect responses from several different institutions. We can also use student interviews to further investigate their abilities to meet the LPs. Think-Aloud interviews [19], in particular, would be beneficial to elicit student reasoning while they go through tasks, and thus, modifying our methodology by data triangulation.

While we used student responses to a free-response task to develop this rubric, we are modifying it to align with the structure of coupled, multiple-response tasks to accommodate the different nature of student responses inherent in that format [20]. Thereby, we can modify our methodology to determine instructor feedback for tasks in a coupled, multiple-response format. We will also integrate the concepts from available feedback models to identify best practices for instructional feedback. We plan to interview instructors who teach upper-division thermal physics courses, who will be the end-users of the rubric along with its feedback, for their perceptions of the utility of the rubric along with the feedback.

ACKNOWLEDGMENTS

[1] J. W. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing what students know: The science and design of educational assessment.* (ERIC, 2001).

[2] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, American journal of Physics **66**, 64 (1998).

[3] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Oko-roafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proceedings of the National Academy of Sciences **111**, 8410 (2014).

[4] J. Von Korff, B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell, Secondary analysis of teaching methods in introductory physics: A 50 k-student study, American Journal of physics **84**, 969 (2016).

[5] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, The physics teacher **30**, 141 (1992).

[6] A. Madsen, S. B. McKagan, M. S. Martinuk, A. Bell, and E. C. Sayre, based assessment affordances and constraints: Perceptions of physics faculty, Physical Review Physics Education Research **12**, 010115 (2016).

[7] N. R. Council *et al.*, *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (National Academies Press, 2012).

[8] C. J. Harris, J. S. Krajcik, J. W. Pellegrino, and A. H. De-Barger, Designing knowledge-in-use assessments to promote deeper learning, Educational Measurement: Issues and Practice **38**, 53 (2019).

[9] N. S. Stephenson, E. M. Duffy, E. L. Day, K. Padilla, D. G. Herrington, M. M. Cooper, and J. H. Carmel, Development and validation of scientific practices assessment tasks for the general chemistry laboratory, Journal of Chemical Education **97**, 884 (2020).

[10] 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).

[11] R. Mislevy and M. Riconscente, Evidence-centered assessment design: Layers, structures, and terminology (padi technical report 9) (2005).

[12] M. M. Cooper, M. D. Caballero, D. Ebert-May, C. L. Fata-Hartley, S. E. Jardeleza, J. S. Krajcik, J. T. Laverty, R. L. Matz, L. A. Posey, and S. M. Underwood, Challenge faculty to transform stem learning, Science **350**, 281 (2015).

[13] J. T. Laverty, 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).

[14] 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).

[15] J. Kozminski, H. Lewandowski, N. Beverly, S. Lindaas, D. Deardorff, A. Reagan, R. Dietz, R. Tagg, M. EblenZayas, J. Williams, *et al.*, Aapt recommendations for the undergraduate physics laboratory curriculum, AAPT: College Park, MD, USA (2014).

[16] R. L. Matz, C. L. Fata-Hartley, L. A. Posey, J. T. Laverty, S. M. Underwood, J. H. Carmel, D. G. Herrington, R. L. Stowe, M. D. Caballero, D. Ebert-May, *et al.*, Evaluating the extent of a large-scale transformation in gateway science courses, Science advances **4**, eaau0554 (2018).

[17] N. R. Council *et al.*, *Developing assessments for the next generation science standards* (National Academies Press, 2014).

[18] 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).

[19] K. A. Ericsson and H. A. Simon, Verbal reports as data., Psychological review **87**, 215 (1980).

[20] K. D. Rainey, A. P. Jambuge, J. T. Laverty, and B. R. Wilcox, Developing coupled, multiple-response assessment items addressing scientific practices, in *Physics Education Research Conference Proceedings*, PER Conference (Online, Virtual, 2020).