

# Impact of perceived grading practices on students' beliefs about experimental physics

Bethany R. Wilcox<sup>1</sup> and H. J. Lewandowski<sup>2</sup>

<sup>1</sup>*Department of Physics, Colorado School of Mines, Golden, CO 80401*

<sup>2</sup>*JILA, National Institute of Standards and Technology and University of Colorado, Boulder, CO 80309*

Student learning in undergraduate physics laboratories has been a growing area of focus within the PER community. Lab courses have been called out as critical elements of the undergraduate curriculum, particularly with respect to improving students' attitudes and beliefs about experimental physics. Previous work within lab learning environments has focused on the effectiveness of curricular innovations or changes to pedagogy; however, one aspect of the learning environment that has not been investigated is the impact of grading practices on students' beliefs and practices. We explore the possible link between students' perceptions of what is valued and rewarded by course grades and their beliefs about the nature and importance of experimental physics as measured by the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS). We find that there is a significant correlation between students' perceptions of the value of certain activities and their personal epistemologies with respect to those activities.

PACS numbers: 01.40.Fk

## I. INTRODUCTION

Student learning in physics laboratory courses has been a growing area of focus within the PER community. Lab courses have been repeatedly cited not only as critical elements of the undergraduate curriculum, but also as places where there is a significant need for improvement [1–3]. In particular, the undergraduate physics laboratory is an important environment with respect to improving students' attitudes and beliefs about the nature and process of experimental physics, as well as its place within the discipline [1, 4, 5]. Previous research on student learning within the lab environment has included developing consensus learning goals [4], investigating the effectiveness of changes in classroom practice [6, 7], or complete overhauls of the laboratory environment to improve student outcomes [8, 9].

One aspect of lab courses that has not been investigated is how grading practices may impact what aspects of experimental physics are perceived by students as valued and valuable within their course. This, in turn, may impact their beliefs about the nature and process of experimental physics more generally. Previous literature on grading practices in the context of lecture courses indicates that what is rewarded in the classroom often has a significant impact on students' behavior, study patterns, and learning [10–12], and that impact is often larger than when the instructor only talks about it being important [13]. For example, traditional grading practices often encourage rote learning and may result in negative learning outcomes particularly with respect to attitudes and epistemologies [10, 13]. In lab courses specifically, it may be that certain grading practices encourage students to adopt novice-like ideas about what experimental physics is and how it is done.

Here, we explore the possible link between students' perceptions of what is valued and rewarded by course grades and their beliefs about the nature and importance of experimental physics as measured by the Col-

orado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [14]. E-CLASS is a 30 item, Likert-style survey in which students are presented with a statement (e.g., “Designing and building things is an important part of doing physics experiments.”) and asked to rate their level of agreement both from their personal perspective when doing experiments in class and that of a hypothetical experimental physicist. The E-CLASS was developed in conjunction with laboratory course transformation efforts at the University of Colorado Boulder (CU) [4]. The instrument was validated through student interviews and expert review [14], and was tested for statistical validity and reliability using responses from students at multiple institutions and at multiple course levels [15]. This work is part of ongoing analysis of a growing, national data set of student responses to the E-CLASS.

## II. DATA SOURCES & METHODS

Data for this study were drawn from an existing data set consisting of seven semesters of students' responses to the E-CLASS collected between 01/2013 and 05/2016. These data were collected from multiple institutions across the United States through the E-CLASS centralized, online administration system [16]. Students completed the E-CLASS both pre- and postinstruction, typically in the first and last week of the course respectively. Only students for whom we had matched pre- and postinstruction responses were included in the analysis. The final matched data set included  $N = 7167$  matched responses from 130 distinct courses at 75 institutions. The institutions in the data set spanned a range of institution types including 2-year ( $N_{inst} = 3$ ) and 4-year colleges ( $N_{inst} = 35$ ), as well as masters ( $N_{inst} = 8$ ) and Ph.D. granting universities ( $N_{inst} = 21$ ). Several of the courses used the E-CLASS during multiple semesters, thus the full data set includes student

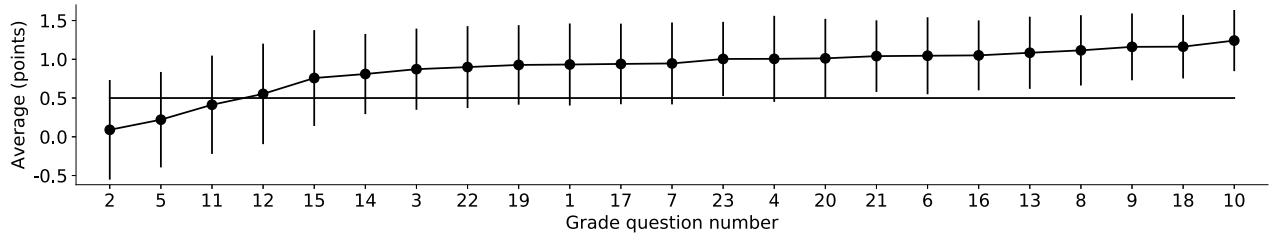


FIG. 1. Sorted plot of the average student rating for each of the 23 grades questions. See Table I for the prompts for each grade question. Vertical bars characterizing the spread are one standard deviation (not one standard error of the mean).

responses from 206 separate instances of the E-CLASS. These courses also span multiple levels including first-year (FY,  $N_{courses} = 102$ ) introductory courses and beyond-first-year (BFY,  $N_{courses} = 104$ ) courses.

Response options for E-CLASS items are given on a 5-point Likert scale (‘strongly agree’ to ‘strongly disagree’). For scoring purposes, students’ responses to each 5-point Likert item were condensed into a standardized, 3-point scale in which the responses ‘(dis)agree’ and ‘strongly (dis)agree’ were collapsed into a single category [17]. Students’ responses to individual items were given a numerical score based on consistency with the accepted, expert response: favorable (+1), neutral (+0), or unfavorable (-1). A student’s overall score on the assessment is given by the sum of their scores on each of the 30 E-CLASS items resulting in a possible range of scores of  $[-30, 30]$ . For more information on the scoring of the E-CLASS see Ref. [15]. Because the E-CLASS data are fundamentally ordinal, rather than interval, in nature, we utilize the non-parametric Mann-Whitney U test [18] to determine statistical significance of differences in raw scores. When calculating correlations between items, we use the non-parametric Spearman correlation coefficient [19].

In addition to the E-CLASS item prompts, the post-instruction E-CLASS also includes 23 items asking students to rate how important – on a 5 point Likert scale from ‘not at all important’ to ‘very important’ – particular elements of experimental physics were for earning a good grade in their course (see Table I) [14]. For the purposes of quantifying students responses, the grades questions were assigned a numerical score from  $[-2, 2]$  points with -2 points for ‘not at all important’ up to 2 points for ‘very important.’ We did not further collapse the numerical scores on the grades questions to a 3-point scale primarily because of the non-trivial distinction between “not at all important” and “somewhat unimportant.” Each grade question was designed to pair with one of the E-CLASS prompts. For example, the E-CLASS item, “Whenever I use a new measurement tool, I try to understand its performance limitations,” has the matched grade question, “How important for earning a good grade was understanding the performance limitations of the measurement tools.” Seven of the 30 E-CLASS prompts did not have a matched grade question because these items targeted student affect or confidence in ways which were unlikely

to be directly incorporated into the overall course grading structure (e.g., “I don’t enjoy doing physics experiments.”). While students’ perceptions of how important certain elements of the course were for their grade represents, at best, a proxy for the instructor’s actual grading practices, we argue that students’ perceptions may in fact be more important for this investigation. We posit that, in order for a particular grading practice to impact the students’ ideas, the students must both be aware of it and believe that it has a meaningful impact on their grade.

### III. RESULTS

To begin, we examine how important students perceive the elements of experimental physics targeted by E-CLASS were for earning a good grade by averaging students’ ratings for each of the 23 grades questions (Fig. 1). Only two items – “reading scientific journal articles” and “randomly changing things to fix a problem with the experiment” – had an average rating in the range of ‘somewhat unimportant’ or ‘not at all important’. An additional two questions were rated on average as ‘neither unimportant nor important’ – “thinking up my own questions to investigate” and “designing and building things.” All remaining questions were rated in the range of ‘somewhat important’ to ‘very important.’ The averages for FY and BFY students separately showed small but statistically significant (Mann-Whitney U and Holm-Bonferroni [20] corrected  $p < 0.05$ ) differences for 11 of the 23 questions with BFY rating the activity as more important in all cases; however, the general trends described above were consistent across course levels.

To determine the extent to which students’ responses to the grades questions relate to their personal E-CLASS responses, we calculate the Spearman correlation coefficient between each grade question and its matched personal question. However, due to the large size of our data set, even very small correlations ( $r < 0.1$ ) were statistically significant. In fact, all of the correlations presented here were statistically significant. For this reason, the concept of practical significance is more relevant in this case. One common threshold used in the literature for determining the practical significance of correlation coefficients is  $r > 0.3$ . To determine if this thresh-

TABLE I. Grade question prompts and the associated correlations for both FY and BFY students. All prompts complete the question, “How important for earning a good grade was...,” and students selected from five options – not at all important, somewhat unimportant, neither unimportant nor important, somewhat important, or very important. Bolded correlations are practically significant and statistically significant difference between the FY and BFY are marked with asterisks.

Q#	Prompt	Correlation	
		BFY	FY
5	...randomly changing things to fix a problem with the experiment?	<b>-0.46</b>	<b>-0.44</b>
11	...thinking up my own questions to investigate?	<b>0.40</b>	<b>0.40</b>
2	...reading scientific journal articles?	<b>0.39</b>	0.24*
4	...working in groups?	<b>0.37</b>	<b>0.38</b>
19	...understanding the performance limitations of the measurement tools?	<b>0.35</b>	<b>0.32</b>
6	...calculating uncertainties to better understand my results?	<b>0.35</b>	<b>0.33</b>
12	...designing and building things?	<b>0.33</b>	<b>0.31</b>
22	...making predictions to see if my results are reasonable?	<b>0.31</b>	<b>0.34</b>
1	...understanding how the measurement tools and sensors work?	<b>0.31</b>	0.29
16	...thinking about sources of systematic errors?	0.28	0.25
13	...understanding the equations and physics ideas that describe the system I am investigating?	0.28	0.29
18	...making conclusions based on data using scientific reasoning?	0.27	0.23
8	...understanding how the experimental setup works?	0.24	0.27
17	...thinking about the purpose of the instructions in the lab guide?	0.23	0.24
9	...understanding the relevant equations?	0.22	0.27*
15	...communicating scientific results to peers?	0.22	0.29*
20	...communicating results with the correct sections and formatting?	-0.19	-0.20
23	...confirming previously known results?	-0.18	-0.17
7	...choosing an appropriate method for analyzing data (without explicit directions)?	0.17	0.11*
10	...using a computer for plotting and analyzing data?	0.15	0.20
21	...learning to use a new piece of lab equipment?	0.13	0.13
14	...overcoming difficulties without the instructor’s help?	0.13	0.10
3	...understanding the approximations and simplifications that are included in theoretical predictions?	0.13	0.17

old is appropriate for our purposes, we calculated correlations between students’ preinstruction responses to the E-CLASS personal questions and their postinstruction responses to the grades questions. As students are unlikely to have gained significant experience with the course grading structure within the first week of class, these two responses should not, in theory, be causally linked. This implies that any observed correlation does not represent the impact of grading practices on students’ ideas. None of the correlations of the grades questions with preinstruction responses exceeded the  $r > 0.3$  threshold, indicating that this is likely an appropriate cutoff for determining practical significance.

Correlations between each grade question and its matched postinstruction personal question are given in Table I. Of the grades questions, only three target activities for which we would anticipate that a person with expert-like personal views would select ‘not at all important’ or ‘somewhat unimportant’ with respect to their importance for earning a good grade – “...randomly changing things to fix a problem with the experiment,” “...communicating results with the correct sections and formatting,” and “...confirming previously known results.”

Consistent with this, these three grades questions were the only questions for which the correlation between students’ personal views and their responses to the grades question was negative (see Table I). The correlation was practically significant for only one of these three questions – “How important for earning a good grade was randomly changing things to fix a problem with the experiment.” Of the 20 positive correlations, six were practically significant for both FY and BFY students, with an additional two that were practically significant for BFY only (see Table I). The differences between the correlations for FY and BFY students were statistically significant for only four questions (see Table I).

With respect to questions whose correlations were not practically significant for either FY or BFY, there are several possible explanations, which may impact the interpretation of the low correlation values. For example, if all students, regardless of their personal views rated the activity as important for their grade, we would see no correlation. Alternatively, if students had significant variation in their responses to the grade question independent of their personal views, we would also expect to see little to no correlation. For 7 of the 14 questions that saw no

practically significant correlation (see Table I), the majority of students with favorable or unfavorable personal views rated the activity as important, suggesting that the lack of correlation is because all students see these activities as important for their grade. For the remaining questions, the lower correlation was typically a result of significant spread in how important students with unfavorable personal views perceived the activity to be, while students with favorable personal views more consistently rated the activity as important for their grade.

#### IV. CONCLUSIONS & LIMITATIONS

Here, we investigated the possible link between students' perceptions of what is valued and rewarded by course grades and their beliefs about the nature and importance of experimental physics. Using a large-scale, national data set of students responses to the E-CLASS, we compared students personal views with respect to particular elements of experimental physics to how important they believed that element was for earning a good grade in their lab course. We found that there was a statistically and practically significant correlation between students' perceptions of the value of certain activities and their epistemologies with respect to that activity in their lab course. While the purely quantitative analysis provided here cannot establish causality, there are at least two possible causal mechanisms that are consistent with our findings. One possibility is that what students' perceive as being explicitly valued and rewarded with respect to grades has a direct impact on their attitudes and beliefs about the nature and process of experimental physics. This hypothesis is also consistent with previous literature that demonstrates a link between assessment

strategies and student outcomes [10–13]. An alternative explanation for our findings is that students with expert-like attitudes and beliefs are better able to recognize ways in which particular elements of experimental physics may impact their grade. However, for either of these two causal mechanisms, our findings suggest that laboratory course instructors should take care not only to intentionally design their grading practices to target the aspects of experimental physics they value and want their students to value, but also to make those grading practices transparent to their students.

In addition to not establishing a causal mechanism, there are several other limitations to these findings. Despite spanning a large number of institutions, courses, and student populations, our data set is not comprehensive nor randomly selected. For example, instructors in our data set represent a self-selected group that generally chose to use E-CLASS without pressure from their department or colleagues. Additionally, we do not have data on instructors' actual grading practices that could be used to determine the accuracy of students' perceptions of the grading structure. In the future, qualitative analysis of graded course artifacts and instructors' grading schemes could help to determine if students were correct in their perceptions of what elements of experimental physics contributed significantly to their course grade. Such analysis could also help to identify grading practices that are most effective at encouraging expert-like epistemologies about experimental physics.

#### ACKNOWLEDGMENTS

This work was funded by the NSF-IUSE Grant DUE-1432204 and NSF Grant PHY-1125844. Special thank you to the members of PER@C for all their feedback.

- 
- [1] AAPT Committee on Laboratories, “AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum,” (2015).
  - [2] S. Olson and D. G. Riordan, Executive Office of the President (2012).
  - [3] NRC Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, *BIO2010: Transforming undergraduate education for future research biologists* (NAP (US), 2003).
  - [4] B. M. Zwickl, N. Finkelstein, and H. Lewandowski, *Am. J. Phys.* **81**, 63 (2013).
  - [5] R. Trumper, *Science & Education* **12**, 645 (2003).
  - [6] M. Wells, D. Hestenes, and G. Swackhamer, *Am. J. Phys.* **63**, 606 (1995).
  - [7] N. G. Holmes, C. E. Wieman, and D. A. Bonn, *PNAS* **112**, 11199 (2015).
  - [8] E. Etkina and A. V. Heuvelen, in *PERC 2001*, PER Conference (Rochester, New York, 2001).
  - [9] R. J. Beichner *et al.*, *Research-based reform of university physics* **1**, 2 (2007).
  - [10] A. Elby, *Am. J. Phys.* **67**, S52 (1999).
  - [11] P. Black and D. Wiliam, *Assessment in Education: principles, policy & practice* **5**, 7 (1998).
  - [12] C. Henderson, E. Yerushalmi, V. H. Kuo, P. Heller, and K. Heller, *Am. J. Phys.* **72** (2004).
  - [13] A. H. Schoenfeld, *Educ. psychologist* **23**, 145 (1988).
  - [14] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. Lewandowski, *PRST-PER* **10**, 010120 (2014).
  - [15] B. R. Wilcox and H. J. Lewandowski, *PRST-PER* **12**, 010123 (2016).
  - [16] B. R. Wilcox, *et al.*, *PRST-PER* **12**, 010139 (2016).
  - [17] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, *PRST-PER* **2**, 010101 (2006).
  - [18] H. B. Mann and D. R. Whitney, *The annals of mathematical statistics* , 50 (1947).
  - [19] C. Spearman, *The American Journal of Psychology* **15**, 72 (1904).
  - [20] S. Holm, *Scandinavian journal of statistics* , 65 (1979).