Correlating students' beliefs about experimental physics with lab course success

Bethany R. Wilcox[1] and H. J. Lewandowski[1, 2]

[1]*Department of Physics, University of Colorado, 390 UCB, Boulder, CO 80309*
[2]*JILA, University of Colorado Boulder, Boulder, CO 80309*

Student learning in instructional physics labs is a growing area of research that includes studies exploring students' beliefs and expectations about experimental physics. To directly probe students' epistemologies about experimental physics and support broader lab transformation efforts both at the University of Colorado Boulder (CU) and nationally, we developed the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS). Previous work focused on establishing the accuracy and clarity of the instrument through student interviews and preliminary testing. Ongoing validation efforts include establishing the extent to which student epistemologies as measured by E-CLASS align with other measures of student learning outcomes (e.g., course grades). We find a weak but statistically significant correlation between final course grades and E-CLASS scores from two semesters of upper-division lab courses at CU and no statistically significant correlation for introductory courses. Here, we discuss implications of these findings for the validity of the E-CLASS instrument.

PACS numbers: 01.40.Fk

I. INTRODUCTION

The Physics Education Research (PER) community has a growing body of research dedicated to investigating students' attitudes and epistemologies about what it means to know, learn, and do physics. In addition to having implications for retention and persistence, a student's epistemological stance can potentially impact their content learning in a physics course [1]. Several assessments have been developed to measure students' epistemologies and expectations (E&E) both about physics specifically [2, 3] and the nature of science more generally [4–7].

In developing standardized assessment instruments, there are a number of distinct aspects of validity and reliability that must be established [8]. For example, convergent validity is determined by comparing the scores on the assessment with other, related student outcomes [9]. For conceptual assessments, convergent validity is typically established relative to students' final course grades. However, for E&E assessments, it is reasonable to ask if we expect the same level of correlation with course performance [3, 10], particularly given that promoting expert-like attitudes and beliefs is rarely an *explicit* goal of physics courses. Of the available E&E assessments, only the VASS (Views About Science Survey) has published results reporting a modest but significant correlation ($r \sim 0.3$) between VASS score and final course grades in high school and college physics courses [11]. Weak, but statistically significant, correlations have also been reported between students' scores on either the CLASS (Colorado Learning Attitudes about Science Survey) or MPEX (Maryland Physics Expectations Survey) and their gains on various validated conceptual assessments [12–14]. Literature on other E&E assessments has focused instead on demonstrating that populations who can be expected to have more expert-like epistemologies (e.g., faculty and graduate students) score higher on these instruments than novices [3].

The goal of the work described here is to further explore the concept of convergent validity as it pertains to E&E surveys by examining the relationship between students' overall course grades and their scores on one specific instrument – the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [15]. E-CLASS is a 30 item, Likert-style survey designed to measure students' epistemologies and expectations about experimental physics. The development and initial validation of the E-CLASS is described in Ref. [15]. E-CLASS was developed to support upper-division laboratory course transformation efforts at the University of Colorado Boulder (CU) [16] as well as similar efforts nation-wide. This work is part of ongoing research into the overall validity and reliability of the E-CLASS.

II. METHODS

Data for this study were collected from two semesters of the four core physics laboratory courses at CU. These courses are described in Table I and span both the lower- and upper-division level. E-CLASS was administered online to each of the courses both at the beginning and end of each semester as a normal part of the class. An example item from the survey is given in Fig. 1. Students were awarded only participation credit for completing the survey. Final letter grades were collected for all students who agreed to release their grade data, and students were assigned a standard grade point value for each letter grade (i.e., $A = 4.0$, $A- = 3.7$, $B+ = 3.3$, $B = 3.0$, etc.). Only students with matched post-test E-CLASS scores and final course grades were included in the following analysis. The E-CLASS also includes a filtering question that prompts the students to select 'agree,' not 'strongly agree,' in order to eliminate responses from students who did not read the questions. Students in the final, matched data set had, on average, final course scores

When doing an experiment, I try to understand how the experimental setup works.

What do YOU think when doing experiments for class?
    Strongly disagree 1 2 3 4 5 Strongly agree

What would experimental physicists say about their research?
    Strongly disagree 1 2 3 4 5 Strongly agree

FIG. 1. An example item from the E-CLASS. Students are asked to rate their agreement with the statement from their own perspective and that of an experimental physicist.

TABLE I. Core physics lab courses offered at CU. Valid N represents the number of students for which we have matched post E-CLASS scores and final grades. All courses except [†] were offered twice in the two semesters of data collection.

Course	Year	N	Response Rate
Experimental Physics 1	1st	717	65%
Experimental Modern Phys.	2nd	76	52%
Electronics for Physical Sci.	3rd	64	82%
Advanced Laboratory [†]	4th	16	100%

that were 0.22 grade points higher than the course overall. We anticipated this type of selection effect because survey completion was worth only minimal credit for the students; however, the difference in final grade between the participants and the class overall is small and thus we take our sample to be reasonably representative.

While the survey was administered both pre and post in all courses, the following analysis focuses exclusively on post-test scores rather than pre-post shifts, as our primary goal was to validate students' scores on the instrument rather than to compare the impact of different courses or interventions. We will also limit our analysis to students' personal beliefs, rather than their prediction of what an experimental physicist would say (see Fig. 1). The overall post-test score for each student is given by the fraction of items that they answered favorably (i.e., consistent with established expert responses). For analysis of individual items rather than overall scores, students' responses to each 5-point Likert item are condensed into a standardized, 3-point scale in which the responses '(dis)agree' and 'strongly (dis)agree' are collapsed into a single category. Thus, student responses to individual items are coded simply as favorable, neutral, or unfavorable. The collapsing of the 5-point scale to 3-points is common in analysis of Likert-style items and is motivated by the inherently ordinal, rather than interval, nature of the Likert response scale [17]. Here, we further collapse neutral and unfavorable responses to a single category so that, consistent with the overall E-CLASS score, student responses to each item are classified simply as favorable (+1) or not-favorable (0).

III. RESULTS

Aggregating across all students in all courses (N=873), we found an overall correlation coefficient of $r = 0.05$ between final course grade and E-CLASS score (fraction of items answered favorably). This correlation is neither practically or statistically significant ($p = 0.1$). However, it is also reasonable to expect that this correlation might vary between courses. In particular, the first year lab at CU is a large service course catering primarily to engineering, rather than physics majors. Thus, the learning goals and grading practices of this course are

not necessarily aligned with the learning goals targeted by E-CLASS, which were developed in collaboration with physics faculty to capture their desired learning outcomes for physics students in their upper-division lab courses.

To investigate this potential variability across courses, we divided the students into two subgroups composed of those in the first-year lab (N=717) and those in the second-, third-, and fourth-year labs (N=156), which we will refer to as the beyond-first-year (BFY) labs. Our motivation for dividing the students in this way was threefold. Firstly, it provides a clear distinction between the classes that is applicable beyond CU. Secondly, it preserves sufficient statistical power given the significantly smaller size of the BFY courses. Finally, there is a significant shift in the student population between the first year and BFY labs. The first year lab is taken by most STEM majors including all engineering majors, whereas the BFY courses are taken almost exclusively by physics, engineering physics, and astrophysics majors.

For students in the first-year lab, the correlation between overall E-CLASS score and final grade is small and not statistically significant ($r = 0.01$, $p = 0.7$). However, for the BFY labs, this correlation increases to $r = 0.19$ and is statistically significant ($p = 0.02$). This correlation, while still weak, is similar in magnitude to the correlations reported between CLASS/MPEX scores and conceptual learning gains as measured by conceptual assessments such as the Force Concept Inventory [12–14]. We are not arguing that the relationship between E-CLASS scores and final grades is a causal one; however, these results do suggest that the link between course performance and epistemological stance is stronger in more advanced lab courses than in the first-year lab.

We can also break down student responses to E-CLASS for each of the 30 items individually to examine which of these items best correlates with grades, and investigate how this varies by course level. For example, one item on the E-CLASS asks students to agree or disagree with the statement "Scientific journal articles are helpful for answering my own questions and designing experiments." Most instructors would neither expect their first-year lab students to agree with this statement nor consider this to be one of the goals of their courses. Thus, we would reasonably expect variability in item correlations across both items and courses. A plot of the item-grade correlations for each item on the E-CLASS is given in Fig.
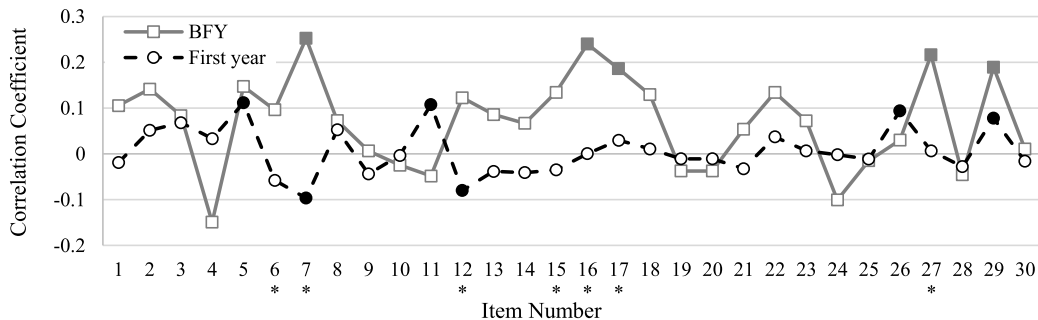
FIG. 2. Item score to final grade correlations for each of the 30 E-CLASS items. Statistically significant correlations (i.e., $r \neq 0$, $p < 0.05$) are indicated by solid markers. Items for which there is a statistically significant difference ($p < 0.05$) between the correlations for the first-year and the BFY courses are indicated with an asterisk. For a list of individual prompts see [15].

2. Fig. 2 also indicates both items for which there is a statistically significant item-grade correlation and items for which there is a statistically significant difference between the correlations for first-year and BFY courses. As our goal is to preliminarily identify questions that might warrant further study, significance tests of these correlations have not been corrected for multiple-tests.

There are multiple reasons why an item might show high or low correlation with overall grade; thus, on its own, the correlation coefficient between item score and final grade (Fig. 2) provides limited insight into the nature of the relationship between these two measures. To understand these correlations and the differences between first-year and BFY courses, we must examine the distribution of students' final grades relative to their responses to each item. As an example of this, we focus on Q16, which asks students to rate their agreement with the statement "*The primary purpose of doing physics experiments is to confirm previously known results.*" Fig. 3 provides the fraction of students who answered favorably and not-favorably for each final letter grade in both the first-year and BFY courses.

From Fig. 3 we can see that a significant fraction of the students in both the first-year and BFY labs do not respond favorably to this question. However, for the first-year lab, the fraction of favorable responses only fluctuates by roughly 10% based on course performance, whereas there is a general downward trend to the fraction of favorable responses as final grade decreases for the BFY labs. This result is consistent with the types of experiments done in these courses. The first-year lab at CU involves almost exclusively confirmatory labs in which the students measure the value of a previously known parameter. Alternatively, the labs in the BFY courses are more exploratory and design-based. These courses often also include a final project that is selected and designed by the students.

Another question that stands out as a potentially interesting case study is Q7, which asks students to rate their agreement with the statement "*I don't enjoy doing physics experiments.*" A plot like that in Fig. 3 shows that in the first-year lab, high performing students (final grades between A and B-) were evenly split between favorable and not-favorable responses, while the lower performing students (final grades between C+ and F) had a slightly higher fraction of favorable responses. This result may be partially explained by the student population of this course. A large fraction of the students were STEM (not physics) majors fulfilling a required natural science credit. These students are often motivated not by an enthusiasm for physics, but rather the desire to get a good grade and continue onto their major courses or complete their degree. Alternatively, in the BFY labs, the fraction of students with favorable responses was highest for the A students and had a general downward trend as final grades decreased. As students in these courses were primarily in physics or physics-related majors, they are more likely to be intrinsically motivated by an enthusiasm for physics and/or discouraged by poor performance.

IV. DISCUSSION

We examined the correlation between students' overall E-CLASS score and their final grade for two semesters of four lab courses at CU. We found that while there was no statistically significant correlation between grades and E-CLASS score for the student overall, there was a small but significant correlation of $r = 0.19$ for the students in the beyond-first-year labs. The lower correlation in the first-year course is likely related to both the nature of those courses and the fact that the E-CLASS includes some items that are specifically designed to target upper-division rather than introductory learning goals. However, $r = 0.19$ still represents a fairly weak correlations, suggesting that E-CLASS scores are not good predictors of students' course performance at either level (or vice versa). To interpret these findings, we need to take a closer look at the way final grades are determined for these courses. In the first-year lab course at CU, 60% of students' grades are based on their lab reports, while the remainder is a combination of participation, traditional homework, and pre-lab assignments. Lab reports are graded on clarity of both presentation and physics
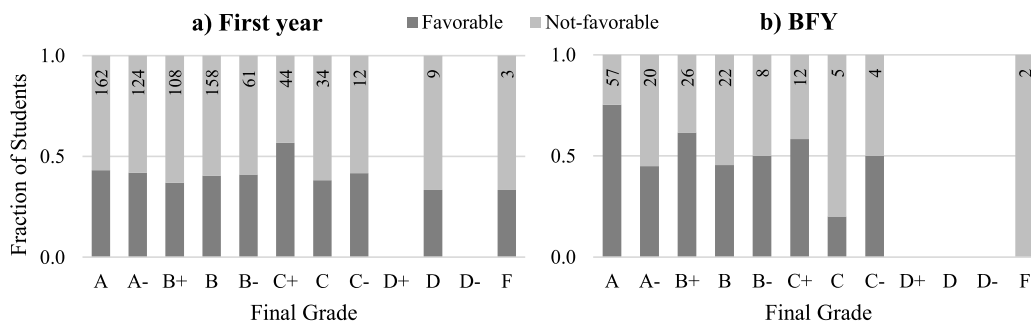
FIG. 3. Fraction of students who answered favorably vs. not-favorably to Q16 for each letter grade in: a) the first-year lab and b) the beyond-first-year labs. Number of students who received each final letter grade is given at the top of each bar.

content, including data presentation, data analysis, and discussion of uncertainty.

Grading schemes in the BFY courses, while varied, are generally characterized by roughly 50-80% of a students' final grade stemming from performance on lab notebooks, lab reports, and, for the 3rd and 4th year courses, the final project. The final project is typically a multi-week group project in which the students propose their own project topics rather than being assigned one or choosing from a predetermined list. The open-ended nature of these projects make them one place where we might expect to see clearer connections between students' epistemologies and their successes and failures when completing their project. Indeed, we might hypothesize that the inclusion of the final project may account for the marginal increase in the correlation between E-CLASS score and final grade for the BFY courses. However, our results still suggest that students' epistemological knowledge and beliefs are not being effectively captured by their final course score. If we, as laboratory instructors, value the epistemological development of our students, we need to begin measuring that development explicitly through our grading practices.

The validation of the E-CLASS is ongoing. Current work includes utilizing the growing, national data set of student responses from multiple courses and levels across many institutions to establish the statistical validity and reliability of the instrument for a broad student population. This national data set may also help us to determine which grading practices best align with the goal of promoting students' epistemological development. Future work could include longitudinal studies of the change in students epistemologies as they advance through the physics curriculum, and further investigation of the link between students' epistemology and their success on project-based assignments by correlating student's E-CLASS scores with their scores on the final project in the BFY labs.

ACKNOWLEDGMENTS

[1] L. Lising and A. Elby, Am. J. Phys. **73**, 372 (2005).
[2] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, Phys. Rev. ST-Phys. Educ. Res. **2**, 010101 (2006).
[3] E. F. Redish, J. M. Saul, and R. N. Steinberg, Am. J. Phys. **66**, 212 (1998).
[4] I. Halloun, in *AIP Conf. Proc.*, Vol. 399 (1996) p. 605.
[5] F. Abd-El-Khalick, N. G. Lederman, R. L. Bell, and R. S. Schwartz, JRST **39**, 497 (2001).
[6] S. Chen *et al.*, Sci. Educ. **90**, 803 (2006).
[7] A. Elby, J. Frederiksen, C. Schwarz, and B. White, "Epistemological beliefs assessment for physical science (ebaps)," (2001).
[8] P. Engelhardt, in *Getting Started in PER*, Vol. 2 (2009).
[9] W. K. Adams and C. E. Wieman, Int. J. Sci. Educ. **33**, 1289 (2011).
[10] D. Hammer, Phys. Teach. **27**, 664 (1989).
[11] I. Halloun, Educational Research Center, Lebanese University, Beirut, Lebanon **5**, 3 (2001).
[12] K. Perkins, W. Adams, S. Pollock, N. Finkelstein, and C. Wieman, in *AIP Conf. Proc.*, Vol. 790 (2005) p. 61.
[13] S. J. Pollock, in *AIP Conf. Proc.*, Vol. 790 (2005) p. 137.
[14] G. Kortemeyer, Phys. Rev. ST-Phys. Educ. Res. **3**, 010101 (2007).
[15] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. Lewandowski, Phys. Rev. ST-Phys. Educ. Res. **10**, 010120 (2014).
[16] B. M. Zwickl, N. Finkelstein, and H. Lewandowski, Am. J. Phys. **81**, 63 (2013).
[17] M. Lovelace and P. Brickman, CBE-Life Sciences Education **12**, 606 (2013).