Student performance and perception on a lab practicum assessment in a reformed introductory physics lab

Julian D. Gifford (he, him, his)
Physics, University of Colorado - Denver, 1200 Larimer St. NC3123, Denver, CO, 80204

A growing body of research on laboratory classes has encouraged a shift away from traditional, cook-book, content-validation activities and towards more authentic, inquiry-based explorations. Over the last several years we have reformed our Introductory Experimental Physics course based on this call. Despite positive responses from students and the instructional team, a consistent issue was the authentic, individual assessment of experimental content knowledge and practical laboratory skills. This paper discusses the goals and structure of a new laboratory practicum assessment, and reports on student performance and reception of the assessment over two semesters of data collection. Results show that student performance is stable over two semesters, and that students broadly feel that the practicum is a fair and accurate assessment of their performance in lab. Differences between the two courses (Lab 1 and Lab 2) and between the two semesters of data collection are explored in detail.

2025 PERC Proceedings edited by Pawl, Zwolak, and Leak; Peer-reviewed, doi.org/10.1119/perc.2025.pr.Gifford

I. INTRODUCTION

There is a broad research vein in the Discipline Based Education Research (DBER) community focused on laboratory courses and student reasoning in experimental science [1–6]. This work has led to a push for reformed laboratory courses that move away from "cookbook" validation activities and focus on the development of experimental skills including experimental design, set-like reasoning, and assessment of model validity [7–11]. While there has been a recent increase, substantially less attention has been given to forms of *assessment* in reformed laboratory courses [12, 13] and specifically to student *perceptions* about these assessments.

Over several years, my institution has reformed our introductory physics lab courses in line with the suggestions of the DBER community. Since they are run as stand-alone courses, divorced from the introductory lecture classes, it is critical to have a course-based assessment of these skills, as they are not specifically addressed in any lecture assessments. Previously, we used online quizzes to assess students' understanding of basic concepts and data analysis skills, but had no measure of individual students' experimental measurement skills. Anecdotally, these online assessments included high levels of cheating, and so did not adequately measure students' knowledge or highlight gaps in understanding. To address these issues, we have developed and implemented a laboratory practicum exam. In tandem with these changes, we developed a practicum survey to measure students' reaction to this new assessment, focused on their views about the content tested on the practicum and their experience taking it. This survey is *not* a laboratory concept inventory [14] or validated affective survey [15], it is a course-level survey meant to capture student perspectives on a changing curriculum.

Section II discusses the practicum broadly, including its motivations, goals, structure, and implementation. The bulk of this paper is an analysis of students' response to the practicum, looking both at student performance in section III B and student perception in section III C.

II. THE PRACTICUM

The reformed lab curriculum is a two-semester sequence (Lab 1 and Lab 2) divorced from lecture content and focused on experimental design, data analysis, linearization techniques, and assessing and developing models. Each course is based around 11 core labs; the first seven lab activities establish the primary concepts and skills, and the remaining four provide experience applying and extending these techniques. The capstone of each course is a two week "freeform lab" where students are responsible for planning, designing, and executing an experimental exploration from scratch.

While in-lab time is entirely focused on collaborative group work, students individually submit a prelab and postlab quiz and their own lab notebook. The notebooks and postlab quizzes are the primary formative assessments, and the freeform lab is the primary summative assessment of many of the course learning goals [7]. However, there is also a need to formally assess students' individual grasp of the basic experimental concepts, as well as their ability to engage in simple data collection and analysis. This is the goal of the practicum.

The practicum assessment comes after the seven core labs. This placement provides both students and their TAs a check on their performance, and ensures that students have established the foundational skills required for the second half of the semester. Thus, while it is in many ways a summative assessment, the practicum also serves a formative role.

The practicum is a hands-on, in-person assessment that occurs during the normally scheduled 110 minute lab time. There are three assessment goals for the practicum: foundational concepts, experimental techniques, and data analysis. These goals are assessed via four multiple choice questions, two experiment tasks, and one Excel task, respectively. The experiment tasks are based directly on the prior labs, *e.g. build a pendulum with a period of 1.2 seconds. Clearly show any calculations and lay out your pendulum for your TA.* To complete this task, students must calculate the appropriate length and then construct the pendulum using the same materials used in the prior pendulum lab activities. TAs graded this task based on a correct calculation and a consistent pendulum length accounting for the center of mass of the pendulum bob. The Excel task requires students to apply the data analysis skills they have developed to address a familiar, but novel, question, *e.g. You conduct an experiment, changing the quantity "Q" and measuring the response "R," these data are presented in the table below. Using the linearization techniques we've developed in class, determine the type of relationship (linear, power, exponential, etc.) between Q and R.* For this task, students must plot the initial (non-linear) data and then make a log-log and two semi-log plots. The type of relationship is based on the most-linear of these plots.

During the practicum, the lab room is broken into three stations (experiment 1, experiment 2, and Excel). Students are randomly assigned to one of these stations to start, and have 30 minutes to complete the task. Once students have finished they call over their TA to be graded. After all students have been graded, the entire class rotates to the next station. Students can work on the multiple choice questions at any point, and can use the remaining time following the third rotation.

Each task is designed so that our TAs can complete them in around 5 minutes, our Learning Assistants can complete them in around 10 minutes, and prepared students can complete them within 15 minutes. This means that the allotted time is approximately *double* the expected time. This additional time is meant to allow the TAs to grade students and to eliminate the need to schedule separate exams for students with time accommodations. In the two semesters that we have run the practicum, the only station that has shown any indication of being time sensitive is the Excel station. While multiple students have been concerned about the timing before taking the practicum, no students have indicated that they had insufficient time, and most report having substantial down time.

III. PRACTICUM RECEPTION

Two primary sources of data are used to assess the impact of the lab practicum: student *performance* as measured by average practicum scores and student *perception* measured by responses to the (optional) practicum survey. Performance scores come from all enrolled students, while survey responses represent a majority (80%) of all students. See Table I for a breakdown of N values and response rates across the two courses (Lab 1 and Lab 2) over two semesters.

TABLE I. N values and average scores. N_c is the total number of students in the *course*, while N_s is the number of *survey* responses.

Course	N_c	N_s	RR	Score
Lab 1 Sem 1	221	182	82.4%	$70.3 \pm 1.0\%$
Lab 1 Sem 2	135	106	78.5%	$76.5 \pm 1.3\%$
Lab 2 Sem 1	105	80	76.2%	$84.6 \pm 1.3\%$
Lab 2 Sem 2	162	137	84.6%	$89.8 \pm 0.9\%$
Combined	623	505	81.1%	79.2%

A. Data collection

Student performance data come directly from practicum scores. Table I shows these scores and standard errors, separated by course and semester but with the multiple sections of each course combined.

Student perception data was collected through responses to the practicum survey. Students were given extra credit towards their course score, and provided informed consent. Extra credit was awarded for any students who submitted a survey, even if it was blank. In total, there were four fully blank surveys and six partially complete surveys. Fully blank surveys were omitted, but partial survey responses were retained, percentages below are reported based on N for each question. As shown in Table I, the response rate across both courses and semesters was approximately 80%, so the survey response data represents the majority of all enrolled students.

The survey has two sets of 5-point Likert-style questions (strongly disagree to strongly agree). The first set is focused on student perceptions about the *content* on the lab practicum, while the second set is focused on student perceptions about their *experience* taking the practicum. The exact language for these questions is provided in Table II. Following these question blocks, students were asked to provide any general feedback or comments in a free response format. Demographic information was collected at the end of the survey.

TABLE II. Content (C) and experience (E) question sets from the practicum survey.

C1	The practicum tested conceptual knowledge based on prior labs
C2	The practicum tested experimental skills based on prior labs
C3	The practicum tested content/skills NOT covered in prior labs
C4	The lab activities prepared me for the practicum
C5	The practicum was a fair assessment
E1	I felt I was well prepared for the practicum BEFORE taking it
E2	I felt I was well prepared for the practicum AFTER taking it
E3	The instructions on the practicum were clear
E4	The practicum was harder than I expected it to be
E5	The practicum was a stressful experience
E6	I'm happy with how I did on the lab practicum
E7	My score on the practicum is a fair reflection of my performance in the lab

B. Student *performance*: practicum scores

There appear to be two clear trends in student scores. First, that for both courses scores seem to improve by about 5% from semester 1 to semester 2; second, that the scores for Lab 2 are notably higher than for Lab 1 (roughly 15% in both semesters, well outside the standard errors). The first of these trends is from a change in the grading structure, described below, while the second highlights a distinction between the two lab courses that is also evident in the survey responses.

In semester 1 the three stations of the practicum were graded out of 4 points each, for a total of 16 points including the multiple choice questions. In semester 2, the three stations were graded out of 5 points, for a total of 19 points. This change means that the score awarded for a single, notable mistake was a 4/5 (80%) rather than a 3/4 (75%). Thus, identical performance in semester 2 received a score about 5% higher, which is almost exactly the increase in practicum scores. Given this, the average scores show that student performance on the practicum is stable. While there is a decrease in the total number of students in semester 2, consistent with the "on-semester" sequencing of introductory physics courses, there were no notable changes to the makeup of the student body. This suggests that the overall difficulty of the practicum was similar between semesters, even though the specific questions changed.

While the minor shift in performance between semesters is primarily a result of the change in grading structure, the significant difference in average scores between courses highlights a fundamental characteristic of our lab courses. While a proper analysis is beyond the scope of this work, the instructional team believes a primary factor is increased student familiarity in Lab 2 with the structure of inquiry courses, rather than the difficulty or student mastery of specific physics or laboratory content. While the *causes* of the differences between our lab students is a topic of ongoing study, for the present work it is sufficient to note that differences exist and that students in Lab 2 generally perform better overall.

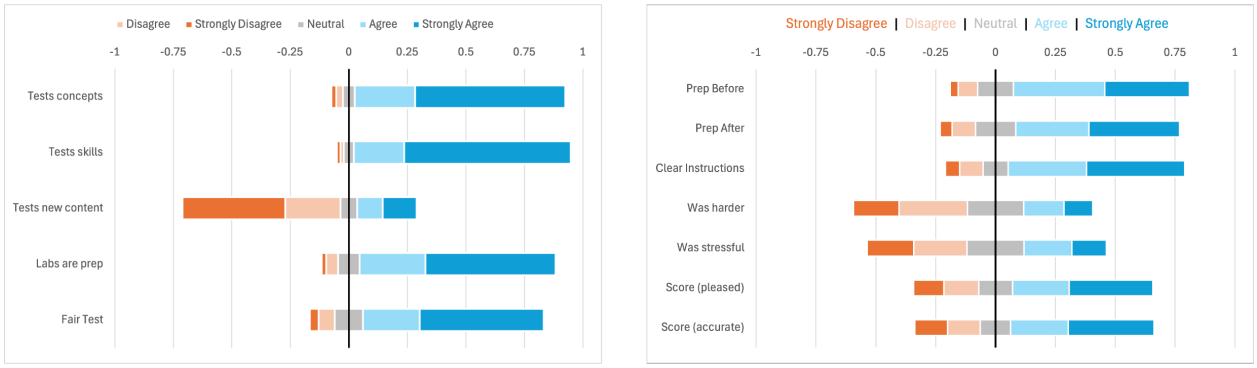

FIG. 1. Overall survey results expressed as percentages for the content [left] and experience [right] question sets. Data is combined over both courses (Lab 1 and Lab 2) and both semesters (S1 and S2) with $N_{tot} > 503$ on each question.

C. Student *perception*: survey results

Figure 1 shows diverging bar plots of aggregate responses across both courses and semesters for the content [left] and experience [right] question sets. While most questions represent the full survey population ($N = 505$), due to partial survey responses some questions have fewer responses, though all have $N > 503$. A "truly neutral" response would be centered at 0 and extend from -0.5 to $+0.5$, thus bars that fall between ± 0.5 and ± 0.75 indicate *general* agreement or disagreement, while bars that fall between ± 0.75 and ± 1 indicate *strong* overall agreement or disagreement.

Student responses to the content set are encouraging, aligning with our "preferred" answer in every case. Results show that students strongly agree (bar edge above 0.75) that the practicum is a fair assessment, focused on both concepts and skills developed throughout the labs, and that engaging in the lab activities prepared them for the practicum. While the disagreement is weaker (edge between -0.5 and -0.75) students also disagree that the practicum tested content that was *not* supported by prior labs.

Results for the experience set are also encouraging, though with overall more neutral views. Students felt prepared for the practicum both before and after they took it. They also strongly agree that the instructions were clear. This result is particularly relevant given that a common response in the written feedback was that the instructions were vague and did not tell students exactly what to do. While this is clearly a critical point for affected students, survey results indicate that this is not a common belief across all students.

Students slightly disagree that the practicum was either harder than expected or particularly stressful, though these results are both close to neutral. This is ideal, as we do not want the difficulty to deviate in either direction from course norms and assessments are inherently stressful. From the written feedback, the most mentioned point of stress was the in-the-moment grading, though some responses indicated they liked knowing their score immediately. Another potential point of stress was the unequal difficulty of the three stations.

Interestingly, while students strongly agreed that the practicum was a fair assessment (C5), they show weaker agreement that they are pleased with their scores (E6) or that their scores are an accurate reflection of their overall performance in lab (E7). Responses to these two questions are effectively identical, suggesting that students' sense of how accurately their score reflects their performance is more related to their happiness with that score than to their perceived fairness of the test.

1. *Course comparison and semester shifts*

While the overall results are encouraging, there are substantial differences between the two courses in student performance, so it is enlightening to look at these results split by lab course. Figure 2 shows the mean student responses separated by course but averaged across both semesters. In these plots Likert responses have been assigned integer values from -2 to $+2$. Values near zero indicate a neutral (or balanced) response, while a value near or beyond ± 1 indicates strong agreement/disagreement.

Though not identical, the overall trends for the content questions are qualitatively similar between the two semesters and compared to the combined data in Fig.1. The primary difference is that students in Lab 1 were more neutral in every response, displaying less "favorable" beliefs towards the practicum than their peers in Lab 2. While this trend continues for the first three experience questions, the responses deviate substantially for the final four experience questions.

The Lab 1 responses to questions E4-E7 all show slightly positive averages indicating weak overall agreement. However, since the averages are all below 0.1, these responses are best categorized as *neutral*. While it is not "bad" for students to have a neutral view about the difficulty, stress, or accuracy of their scores, the splits between the lab courses are notable. Overall, it is clear that Lab 1 students have a less positive response to the practicum than Lab 2 students, but neither course holds negative views about the practicum.

Semester 1 was our initial implementation of the lab

FIG. 2. Average responses to content [top] and experience [bottom] question sets split by course (Lab 1 and Lab 2). Values near zero are neutral, while values near or beyond ± 1 indicate strong (dis)agreement.

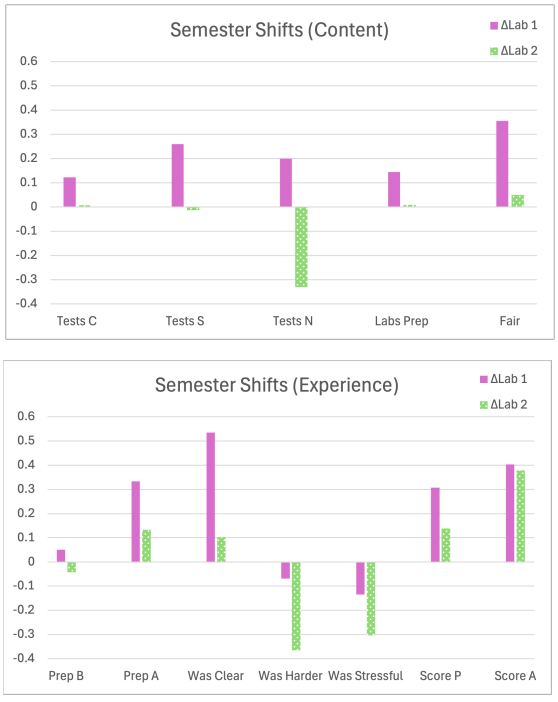

FIG. 3. Average *shifts* in responses to content [top] and experience [bottom] question sets split by course. These shifts are approximately equal to Cohen's d, so shifts below 0.1 are negligible, shifts of 0.1-0.3 are small, and shifts above 0.3 are medium.

practicum, and based on feedback from students and our instructional team we made several updates between the two semesters. As discussed above, one of these updates was to grade the stations out of 5 points rather than 4. Additionally, our TAs had a better sense of what to expect, how students would respond, and their own role in the practicum environment and so were better able to prepare students for the *experience* of the practicum. While a deeper analysis of these changes is a topic of future work, preliminary results of *shifts* between semesters are shown in Fig 3. These plots show average raw shifts on the 5-point Likert scale. The average standard deviation for all questions is 1.1, which means that the shift is approximately equal to the effect size measured by Cohen's d [16]. Thus, a shift below 0.1 is negligible, a shift of $0.1 - 0.3$ is small, and a shift of $0.3 - 0.6$ is medium.

With the exception of the small positive shift for Lab 1 on question C3 (the practicum tests new content), all shifts on both content and experience are in the "preferred" direction. Many shifts for Lab 2 are negligible, but the larger shifts (C3, E4, E5, E7) further solidify students' belief that the practicum is a fair and accurate measure of their performance, that is neither more difficult than expected or particularly stressful. Since the majority of students taking Lab 2 in semester 2 had taken the practicum in semester 1, the negative shifts for questions E4 and E5 for Lab 2 indicate growing comfort with the structure and style of the practicum. Compared to Lab 2, Lab 1 shows a greater number of shifts and larger effect shifts. In particular, Lab 1 shifts on questions C5, E2, E3, and E7 show substantially improved views about the fairness, clarity, and accuracy of the practicum.

IV. CONCLUSIONS

The structure and goals of a laboratory practicum assessment are presented. While a unique and potentially uncomfortable assessment, student responses indicate that it was well-received. On the instructional side, the practicum holds students accountable for developing content knowledge, data analysis and Excel fluency, and basic experimental techniques. Students largely view it as a fair assessment that attends to the goals of the course. This work contributes to the growing body of work on alternative forms of assessment in laboratory courses, offering an example of a specific type of alternative assessment and providing preliminary evidence of its efficacy and students positive perceptions. While future work will code free response results and further investigate the differences between the two lab courses, these initial results provide strong support for the use of a lab practicum to measure both practical and conceptual skills in a reformed introductory experimental physics lab.

Acknowledgments: This work was not supported by a specific grant from any funding agency. Thanks to the UCD physics department for travel and conference funds.

[1] J. Walker, V. Sampson, S. Southerland, and P. Enderle, "Using the laboratory to engage all students in science practices," *Chemistry Education Research and Practice* **17**, 1098 (2016).

[2] B.M. Zwickl, D. Hu, N. Finkelstein, and H. J. Lewandowski, "Model-based reasoning in the physics laboratory: Framework and initial results" *Phys. Rev. ST Phys. Educ. Res.* **11**, 020113 (2015).

[3] A. Hofstein and V. N. Lunetta, "The laboratory in science education: Foundations for the twenty-first century", *Science Education* **88**, 1 (2003).

[4] E. Etkina, A. Murthy, and X. Zou, "Using introductory labs to engage students in experimental design", *Am. J. Phys.* **74** (11), 979-986 (2006).

[5] N.G. Holmes and C.E. Wieman, "Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses", *Phys. Rev. Phys. Educ. Res.* **12**, 020103 (2016).

[6] N.G. Holmes *et. al*, "Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content", *Phys. Rev. Phys. Educ. Res.* **13**, 010129 (2017).

[7] E.F. Redish, *Teaching Physics with the Physics Suite* (John Wiley & Sons 2003).

[8] N.G. Holmes and C.E. Wieman, "Introductory physics labs: We can do better", *Physics Today* **71**, 38-45 (2018).

[9] 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).

[10] B.R. Wilcox and H.J. Lewandowski, "Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics", *Phys. Rev. Phys. Educ. Res.* **12**, 020132 (2016).

[11] B.R. Wilcox and H.J. Lewandowski, "Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics", *Phys. Rev. Phys. Educ. Res.* **13**, 010108 (2017).

[12] S.F. Wolf, M.W. Sprague, F. Li, A. Smith-Joyner, and J.P. Walker, "Introductory physics laboratory practical exam development: Investigation design, explanation, and argument", *Physics Education Research Conference Proceedings* (2019).

[13] L. Ketonen, A. Lehtinen, and P. Koskinen, "Assessment designs of instructional labs: A literature review and a design model," *Phys. Rev. Phys. Educ. Res.* **19**, 020601 (2023).

[14] C. Walsh, K. N. Quinn, C. Wieman, and N.G. Holmes. "Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking" *Phys. Rev. Phys. Educ. Res.* **15**(1), 010135 (2019).

[15] Zwickl, B., Finkelstein, N., and Lewandowski, H. "Development and Validation of the Colorado Learning Attitudes about Science Survey for Experimental Physics," *Physics Education Research Conference Proceedings* (2012).

[16] T. Baguley, "Standardized or simple effect size: what should be reported?", *Br. J. Psychol.* (2009).