Evaluating the presence of response-shift bias in the CLASS with a two-pass survey

W. Brian Lane\textsuperscript{1} and Ramesh Y. Adhikari\textsuperscript{1}

\textsuperscript{1}Department of Physics, Jacksonville University, 2800 University Boulevard North, Jacksonville, FL, 32211 USA

Most pre-to-post-instruction results of the Colorado Learning Attitudes about Science Survey (CLASS) indicate a decline in students’ beliefs and attitudes about learning physics. However, by comparing post-instruction CLASS responses with retrospective responses, we have found that such results can be affected by response-shift bias due to students’ changing self-evaluation metric. Response-shift bias has an overall negative impact on pre-to-post results. We also find that, when asked to predict their end-of-semester responses, students’ predictions of their development were overall positive and were generally met. We discuss the implications of these findings for the interpretation of CLASS results and future explorations of this phenomenon.

I. INTRODUCTION

Since its debut over a decade ago, the Colorado Learning Attitudes about Science Survey (CLASS) has been widely adopted as an assessment tool in physics, chemistry, and biology education\textsuperscript{[1–3]}. When administered at the beginning (PRE) and end (POST) of a physics course, the CLASS yields insight into the change in students’ beliefs and attitudes about the nature of physics, the usefulness of learning physics, and methods of learning physics. This assessment is conducted by comparing students’ agreement and disagreement with 41 statements about physics against expert-like responses, with expert-like and novice-like response rate reported for the survey overall and for statistically validated categories of statements. Unfortunately, when comparing PRE and POST results, the majority of studies involving the CLASS find either no significant change or a deterioration in student attitudes\textsuperscript{[1]}, even in reform-based courses.

However, the self-reported nature of the CLASS leaves its PRE-to-POST results susceptible to response-shift bias\textsuperscript{[4–6]}, a phenomenon in which respondents’ conceptualization of the assessed topic changes over the course of an educational intervention, causing their pre- and post-intervention responses to be based on a different self-evaluation metric. Response-shift bias is known to arise in surveys that ask about attitudinal factors such as self-confidence and interest in the subject, as the CLASS does. For example, CLASS Item 3 asks students to rate their agreement with the statement, “I think about the physics I experience in everyday life.” A student’s response to this item is based on their scope of what topics are included in the subject of physics. Suppose a student has a relatively small scope of what physics includes during the PRE test, and that they respond with agreement to Item 3. Now suppose that, upon reaching the POST test, the student’s scope of what physics includes has grown, and that they again respond with agreement. However, the student likely no longer agrees with their initial self-assessment, believing that this statement was not true about themselves during the PRE test. The lack of change in the student’s PRE and POST response to Item 3 does not detect the change in the student’s self-evaluation metric.

Response-shift bias can be detected by asking participants to answer the same survey retrospectively (RETRO) and comparing the differences between the PRE responses and the RETRO responses\textsuperscript{[4, 7]}. In the example of CLASS Item 3, this would involve asking the student, at the end of the course, whether they thought about the physics they experience in everyday life at the beginning of the course. RETRO-to-POST comparisons typically yield higher gains than traditional PRE-to-POST comparisons, a phenomenon attributed to students’ overestimation of their abilities at the beginning of the assessed intervention\textsuperscript{[6]}. It should be noted, though, that RETRO-to-POST comparisons are not intended to replace PRE-to-POST comparisons, but to complement them with additional insight\textsuperscript{[6]}.

In this article, we evaluate the presence and impact of response-shift bias in CLASS results. In Section II, we describe our implementation of a two-pass survey CLASS administration in which we asked students to complete the survey twice at the beginning of the semester and twice at the end of the semester. In Section III, we establish the presence of response-shift bias in these results and examine changes in student expectations by comparing traditional PRE-to-POST gains with other gains afforded by the two-pass survey administration. In Section IV, we discuss the implications of response-shift bias for CLASS administration and interpretation. Finally, in Section V, we discuss plans to further explore this phenomenon.

II. SURVEY ADMINISTRATION

One established means of evaluating the presence of response-shift bias in self-report assessments such as the CLASS is to administer the survey in the usual PRE and POST manner, and then to conduct a RETRO administration in which, at the end of the intervention, participants report what they currently believe their responses would have been at the beginning of the intervention, based on what they know now\textsuperscript{[4, 7]}. A RETRO administration is visualized in Fig. 1, which depicts survey administration time (beginning of semester or end of semester) in the columns of a table and the survey administration subject (student’s initial state or final state) in the rows, placing the RETRO administration in the upper-right corner (asking students at the end of the semester about their initial state). Being given concurrent
with the POST administration, this RETRO administration reduces the influence of response-shift bias since participants evaluate their initial and final states using the same metric.

In the spring 2017, fall 2017, and spring 2018 semesters, we administered the CLASS in all undergraduate physics sections at a small private liberal arts university. We collected matched PRE and POST responses from a total of $N = 292$ students. The vast majority of these students were enrolled in introductory-level algebra- or calculus-based mechanics, electromagnetism, and astronomy courses, with a few students in upper-level theory and lab courses. During the end-of-semester survey administration, we also solicited RETRO responses from students by asking them to take the survey twice: Once to provide their current responses to each statement, and once to provide what they thought their response would have been at the beginning of the semester. This survey administration explicitly instructed students to use their current (end-of-semester) self-evaluation metric, and having students provide POST and RETRO responses on the same survey allowed students to compare these responses.

In spring 2018, we determined that it would also be helpful to conduct a two-pass survey administration at the beginning of the semester, asking students to provide their current (PRE) responses, and the responses they expected to provide at the end of the semester (EXP). An EXP administration is visualized in the lower-left corner of Fig. 1. This semester saw the collection of matched responses from $N = 88$ students, a subset of the earlier 292 students.

The concurrent structure of these two-pass survey administrations are consistent with other successful experimental CLASS administrations [8] and with other studies of response-shift bias [6]. Almost no students seemed confused by the activity, so there is little concern for overtaxing students’ cognitive loads.

III. RESULTS COMPARISON

In Section III A, we compare the PRE, RETRO, and POST scores and their associated differences for the full set of $N = 292$ students over the spring 2017, fall 2017, and spring 2018 semesters. In Section III B, we compare the PRE, RETRO, EXP, and POST scores and their associated differences for the subset of $N = 88$ students from the spring 2018 semester. These results are reported as percent favorable (average fraction of statements to which students made an expert-like response) and percent unfavorable (average fraction of statements to which students made a novice-like response) for the entire survey, with item categories [1] generally following the same trend as the overall scores. Neutral responses are not recorded as favorable or unfavorable. A difference between PRE, POST, RETRO, or EXP scores is considered statistically significant if its magnitude is greater than twice the standard error in individual student shifts [1]. A decrease in percent favorable score or increase in percent unfavorable score is considered a deterioration, while an increase in percent favorable or decrease in percent unfavorable is considered an improvement. The comparisons we make between the response sets are indicated by arrows in Fig. 1, each of which helps evaluate the presence and impact of response-shift bias.

A. PRE, POST, and RETRO comparison

Figure 2 shows the overall PRE, RETRO, and POST scores for the full set of $N = 292$ students with matched responses. As is typical in the literature, the overall favorable score demonstrates no statistically significant gains, with $\text{PRE} = 51.0 \pm 0.9\%$ and $\text{POST} = 52.8 \pm 1.0\%$. However, comparing this PRE score with the RETRO score of $39.5 \pm 1.1\%$ indicates that, at the end of the semester, students recall their initial state as statistically significantly more novice-like than they reported in their PRE responses. Therefore, we observe a statistically significant improvement when comparing RETRO and POST scores. In other words, at the end of the semester, these students overwhelmingly perceived that their learning attitudes had improved significantly, even though the PRE-to-POST gains indicate otherwise.

This discrepancy occurs consistently in all three semesters, across all four course instructors, and in all CLASS categories. For example, the Personal Interest category (of which the previously mentioned Item 3 is a member) produced favorable scores of $\text{PRE} = 51.5 \pm 1.4\%$, $\text{POST} = 53.4 \pm 1.7\%$, and $\text{RETO} = 39.2 \pm 1.8\%$, demonstrating no statistically significant gains between PRE and POST, but a statistically lower value of RETRO compared with PRE and POST. Therefore, these students recall their initial interest in physics as being lower than they initially reported.

The concept of response-shift bias attributes these discrepancies between the PRE-to-POST gains and the RETRO-to-POST gains to a change in the students’ self-evaluation metric induced by their learning experience. The traditional
FIG. 2. Percent unfavorable and percent favorable scores from the PRE, RETRO, and POST administrations of the CLASS. PRE-to-POST comparison of overall scores shows no statistically significant differences, but RETRO-to-POST comparison shows statistically significant improvement.

PRE-to-POST comparison, while valuable in measuring actual changes in students’ responses, does not seem to detect changes in students’ self-evaluation metric which impacts those responses. Therefore, a PRE-to-POST comparison produces gains that are lower than the students’ end-of-semester perceptions. The addition of a RETRO-to-POST comparison alongside a PRE-to-POST comparison can help assess this additional dimension of student learning.

B. PRE, POST, RETRO, and EXP comparison

Inspired by such trends in the spring 2017 and fall 2017 semesters, in spring 2018 we also requested that students, at the beginning of the semester, report what they expected their responses to be when they completed their physics course. The overall EXP scores, collected from $N = 88$ students, are depicted in Fig. 3, along with the matched PRE, RETRO, and POST overall scores for that semester. The overall favorable scores show a statistically significant difference between PRE $= 51.3 \pm 1.4\%$ and EXP $= 64.7 \pm 1.1\%$, indicating that, at the beginning of the semester, students expected that they would become more expert-like over the course of instruction.

This difference occurs consistently across all four course instructors and in all CLASS categories. For example, the Personal Interest category showed favorable scores of PRE $= 48.9 \pm 2.3\%$ and EXP $= 74.0 \pm 1.9\%$, demonstrating that students expected that their interest in physics would significantly improve by the end of their physics course. This trend might also indirectly verify the claim that students generally know what physicists believe about physics [8], as these students seemed to understand how their views were intended to develop.

The question arises as to whether, at the end of the semester, students perceived that they had progressed toward an expert-like state as much as they expected to at the beginning of the semester. If we compare the gain between EXP and PRE scores (collected at the beginning of the semester) of EXP $- $PRE $= 13.4\%$ with the gain between POST and RETRO scores (collected at the end of the semester) of POST $- $RETRO $= 15.5\%$, we find that these gains are approximately the same. Apparently, these students’ expectations of development were approximately met.

We find similar behavior when comparing EXP $- $PRE gains and POST $- $RETRO gains in all CLASS categories, with only a few exceptions: Personal Interest favorable and Problem Solving Sophistication favorable produced significantly smaller POST $- $RETRO gains than EXP $- $PRE, indicating that students’ expectations of development in these areas were not met. Sense Making/Effort favorable and
Conceptual Understanding favorable produced significantly greater POST – RETRO gains than EXP – PRE, indicating that students’ expectations of development in these areas were exceeded. Such comparisons are useful in helping instructors evaluate which aspects of their course to improve in future semesters.

IV. DISCUSSION

The foremost lesson learned from these results is that PRE-to-POST CLASS gains can be subject to response-shift bias, as evidenced by consistent differences between PRE and RETRO scores and consistent similarities between PRE-to-EXP gains and RETRO-to-POST gains. These similarities in gains indicate that there is a consistency in students’ expectations of their development, but that their self-evaluation metric changes during instruction. Response-shift bias is an important consideration when using the CLASS because it explains how PRE-to-POST gains may suppress important information about the development of students’ self-evaluation metrics. This suppression causes PRE-to-POST CLASS gains to skew in the negative direction compared with RETRO-to-POST gains, which account for students’ changing self-evaluation metrics. One important application of this lesson is that the negative PRE-to-POST shifts observed in many CLASS-based studies may not be as bleak as previously interpreted.

However, response-shift bias does not invalidate the use of the CLASS as an assessment tool or PRE-to-POST gains as a metric of student responses. This standard means of assessment measures actual changes in student attitudes, and not simply students’ perceptions of those changes. Response-shift bias establishes that there is valuable additional information that can be extracted by adding RETRO and EXP administrations to PRE and POST administrations of the CLASS, with little additional time and effort required of instructors.

The two-pass survey method we have conducted also offers important insights into students’ expectations and perceptions of a physics course. Evaluating PRE-to-EXP gains can reveal students’ expectations about how they will develop, which can inform instructors about how to structure their courses and what types of formative assessment to conduct. Evaluating RETRO-to-POST gains can reveal students’ satisfaction with their physics course, especially when compared with their PRE-to-EXP expectations.

V. CONCLUSIONS AND FURTHER QUESTIONS

This work has established the presence of response-shift bias in CLASS responses in undergraduate physics courses by implementing a two-pass survey administration at the beginning and end of various semester-long physics courses. This response-shift bias causes PRE-to-POST gains to suppress information about students’ changing self-evaluation metrics, leading to overly pessimistic PRE-to-POST gains. The addition of EXP and RETRO CLASS administrations highlights these changes in metric and offers more optimistic conclusions about student learning.

Having established the presence of response-shift bias and a means of compensating for it, several avenues of further inquiry arise. First is the question of how a student’s self-evaluation metric shifts over the course of a student’s experience in a physics course. This question is important, as it can inform how instructors attend to metacognitive factors in student learning. We plan to explore this question by interviewing students about the differences between their PRE and RETRO responses and between their EXP and POST responses. Second, one wonders whether response-shift bias has a similar impact when using the chemistry and biology versions of the CLASS. We plan to address this question by extending our two-pass survey administration to include chemistry and biology courses using their respective versions of the CLASS. Lastly, expanded use of the two-pass survey administration prompts the question of how the manifestation of response-shift bias varies with instructor, institution, or course level. Addressing this question will require a much larger data set gathered across multiple institutions.

ACKNOWLEDGMENTS

We are grateful to our departmental colleagues, Terry Ellis and Paul Simony, for administering this survey, and to the students who recorded the survey responses electronically.