

Themes in student self-assessments of attitudinal development in the CLASS

William Paulger, Brendan McEnroe, Ivy Shaw, Ian Crawford-Goss, and W. Brian Lane

Department of Physics, University of North Florida, 1 UNF Drive, Jacksonville, FL, 32224

Response-shift bias explains how changes in students' criteria for self-evaluation can suppress the effect of educational interventions in pre-to-post shifts of self-reported metrics, such as the Colorado Learning Attitudes about Science Survey (CLASS). Therefore, the static or declining pre-to-post shifts typically seen in the CLASS might not reflect development in students' reasoning about themselves based on their acquired knowledge of physics. We observe response-shift bias in a downshift between pre-instruction responses obtained at the beginning of the semester and retrospective pre-instruction responses obtained at the end of the semester. We investigated possible causes of response-shift bias in the CLASS by adding a free response question after select CLASS items during a two-pass survey administration, asking students to explain their reasoning behind their expected and perceived shifts. We present the themes that occurred most clearly in students' free responses to one survey item, which highlight the established forms of response-shift bias: recalibration of the students' standards of measurement, reprioritization of the students' values, and reconceptualization of physics content.

I. INTRODUCTION

The Colorado Learning Attitudes about Science Survey (CLASS) quantitatively describes student attitudes and beliefs about physics and learning physics [1]. Students complete the survey by reporting their agreement with a series of statements using a 5-point Likert scale. Scores for the survey are determined by comparing the students' responses with expert-like responses, with scores reported for the survey overall and categorical subsets of statements. Typically, students complete the CLASS before (PRE) and after (POST) a physics course to assess shifts in student perceptions.

Achieving positive gains on the CLASS has proven challenging. Typically, studies using the CLASS show no significant PRE-to-POST change in student attitudes or show that student responses have become significantly less expert-like over instruction. This result occurs most commonly in traditional courses but also appears in engaged-learning courses [2]. While this outcome seems bleak, our previous work has shown that this pessimistic trend may be partially attributable to response-shift bias [3].

Response-shift bias (RSB) occurs when students use different criteria to evaluate themselves during PRE and POST surveys [4–6]. First identified in training and treatment interventions, RSB produces an overestimation of self-assessed metrics in the pre-test, such that traditional PRE-to-POST comparisons can conflate actual changes in metric with changes in subjects' self-evaluation criteria. RSB can be identified and controlled for by administering a *retrospective pre-test* (RETRO), in which post-intervention subjects rate their pre-intervention metrics. Subjects are more likely to complete RETRO and POST self-assessments using the same criteria. As a result, there is a *downshift* between PRE and RETRO scores, such that RETRO-to-POST shifts show larger, more significant, intervention effects [3, 4] and provide a better sense of the subjects' perception of the effectiveness of an intervention [7]. RETRO-to-POST shifts do not replace traditional PRE-to-POST shifts as a measure of growth. Rather, they provide a measure of RSB by accounting for the students' learning during instruction. Despite the ubiquity of RSB and the established validity of the RETRO-POST assessment, this approach remains underutilized today [7].

We note that “bias” here is not an explicitly negative term. In many ways, RSB is an inevitable, even desirable, byproduct of educational interventions, since improving students' capacity to gauge their abilities, progress, and interests in a subject is an explicit goal of education [8, 9]. Indeed, this bias might help explain the discrepancy seen in physics education reform efforts that produce positive conceptual learning gains alongside static or declining CLASS shifts [1, 2].

RSB can manifest in several different forms: recalibration, reprioritization, and reconceptualization [7]. In recalibration, the subject's internal standards of measurement have changed. For example, when responding to CLASS item Q1 (“A significant problem in learning physics is being able to memorize all the information I need to know.”), a student's

criteria for what counts as “significant” or what comprises “information I need to know” might change. In reprioritization, the subject has experienced a change in the values they use to assess themselves. For example, when responding to CLASS item Q34 (“I can usually figure out a way to solve physics problems.”), a student might change their mind about what “figuring out” a physics problem looks like. In reconceptualization, the subject changes their definitions of the concepts being assessed. For example, when responding to CLASS item Q28 (“Learning physics changes my ideas about how the world works.”), a physics class might change a student's scope of “the world.”

In our prior work, we established the presence of RSB in CLASS responses by administering the survey in a two-pass format to undergraduate physics students [3] based on the standard retrospective survey treatment [4, 7]. In this format, students respond to the survey twice at the beginning and twice at the end of the term (see Figure 1).

At the beginning of the term (left column in Figure 1), students provide their pre-instruction responses as usual and the responses they expect (EXP) to give after completing the course. The difference between these PRE and EXP responses (blue and red bars in Figure 2) illustrates the students' initial frame of mind about how they expect their attitudes to develop during the course.

At the end of the term (right column in Figure 1), students provide their post-instruction responses as usual and the responses they retrospectively believe they would have given before starting the course. The difference between these RETRO and POST responses (green and gold bars in Figure 2) illustrates the students' final frame of mind about how they perceive their attitudes to have developed during the course.

Because students tend to give more novice-like responses on the RETRO survey than on the PRE survey (blue and green bars in Figure 2), RETRO and POST scores show significantly more growth than the usual PRE and POST comparison. This downshift between PRE and RETRO is the evidence of RSB. We have observed this downshift in our previous work [3] and replicated these results here in a different institutional context.

The present project explores possible sources of RSB in CLASS responses. We are guided by the research questions, “How does response-shift bias manifest when students explain their answers to the CLASS during a PRE-EXP and RETRO-POST survey administration?” and “What insights does this reasoning offer into how we might better guide our students during the learning process?” This paper focuses primarily on preliminary answers to the first research question based on data obtained from free-response questions we added to the CLASS. In Section II, we outline the context and methods of our study. In Section III, we discuss preliminary themes from a subset of these free responses, illustrated with examples. Finally, we highlight a few reflective thoughts and further questions in Section IV.

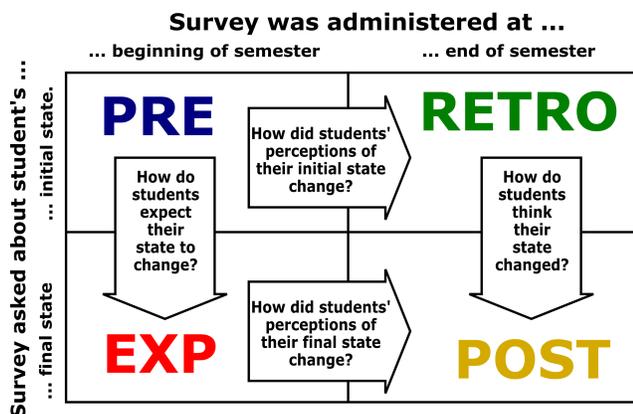


FIG. 1. This diagram illustrates our method of administering the CLASS (originally published in [3], [Creative Commons license 3.0](#)). Students completed the survey at the beginning and end of the semester (represented by the table's columns) with responses indicating their initial and final states (represented by the table's rows).

II. CONTEXT AND METHODOLOGY

We conducted this study during the spring 2021 semester at a mid-sized regional state university that primarily serves undergraduates. We asked the university's 19 physics faculty to deploy our modified survey to their courses; seven faculty members (including two of this paper's authors) agreed to administer the survey across seven different physics and astronomy courses. We collected matched responses from a total of $N = 377$ students. These courses comprise a diverse student population: 54% male, 44% female, 2% non-binary; 59% calculus-based introductory physics, 21% algebra-based introductory physics, 17% introductory astronomy, and 3% upper-division physics courses. We acknowledge that this timeline places these courses and the study during the COVID-19 pandemic, which was unprecedentedly stressful for students and instructors. As such, the seven courses featured split remote/on-site teaching formats, and the surveys were administered on-line.

We administered a two-pass version of the CLASS (PRE-EXP, and RETRO-POST) as described in Section I and [3]. As shown in Figure 2, PRE-to-POST shifts for overall favorable scores show a small significant gain ($+4.8 \pm 0.8\%$). In contrast, PRE-to-EXP shifts ($+9.3 \pm 0.7\%$) and RETRO-to-POST shifts ($+13.8 \pm 1.0\%$) for overall favorable score are notably larger. Essentially, the students overestimate their attitudinal development compared with the traditional survey metrics, the essence of RSB. We see this trend in all statement categories, and within each course.

To explore the reasoning students employed in making this exaggerated self-assessment, we added the following free-response question (FRQ) after four CLASS statements ($S = Q28, Q30, Q35, \text{ and } Q37$ from the Real World Connection category) in the PRE-EXP survey at the beginning of the

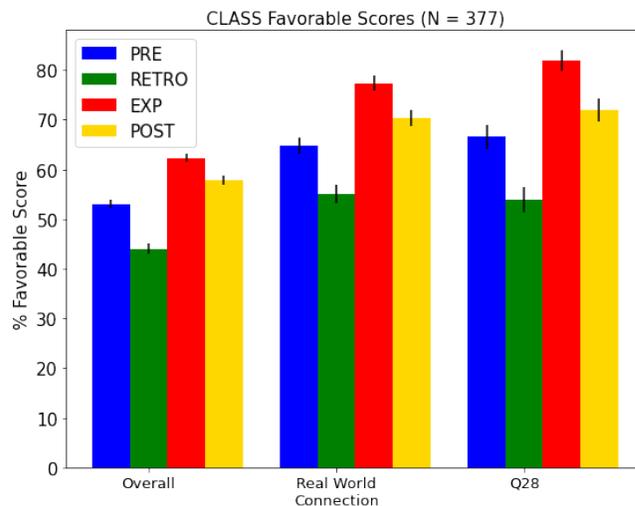


FIG. 2. CLASS results collected from seven undergraduate physics courses for the survey overall, Real World Connection category, and survey item Q28. The PRE-to-POST shifts show small significant gains, while the students' expected shifts (PRE-to-EXP) and perceived shifts (RETRO-to-POST) are notably larger, indicating the influence of response-shift bias. This trend occurs in all survey categories and in all seven courses.

semester:

Answer one of the following questions about Statement S:

- *If you selected the same answer (same number) for both how you feel about this statement now and how you think you'll feel about this statement at the end of the semester, please explain why in 1-2 sentences.*
- *If you selected a different answer (different number) for how you feel about this statement now and how you think you'll feel about this statement at the end of the semester, please explain why in 1-2 sentences.*

We added this same FRQ after these same statements in the RETRO-POST survey at the end of the semester, replacing "how you think you'll feel... at the end of the semester" with "how you think you felt... at the beginning of the semester." We believe the students' answers to these FRQs give insight into *why* there is a downshift between PRE and RETRO responses on these four statements.

For this preliminary report of our work in progress, we discuss responses and trends only for CLASS item Q28: "Learning physics changes my ideas about how the world works." Figure 2 shows that favorable scores for Real World Connection and Q28 follow the trend for the survey overall. We therefore expect these free responses to provide insight into the presence of RSB. Analysis of the remaining items from Real World Connection will follow in future work, as will investigation of RSB in other CLASS categories.

The authors categorized FRQ responses using a team-based, open coding process [10]. To begin this process, each author reviewed a subset of 45 randomly selected FRQ responses (drawn from all Real World Connection statements)

TABLE I. Themes used to categorize student responses to free-response questions for Q28 in the beginning-of-semester (PRE-EXP) survey and end-of-semester (RETRO-POST) survey. Most themes are categorized into an established form of response-shift bias [7].

Theme (RSB form)	Definition	PRE-EXP	RETRO-POST	Sample Responses (PRE-EXP and RETRO-POST)
cause and effect (none)	Response references cause-and-effect relationships.	6.6%	8.5%	“Learning physics in the past <i>has already changed</i> my idea of how the world works before so I would imagine taking this class <i>would have the same effect.</i> ” “Learning physics <i>has made me</i> realize just how much we use it.”
comparison (recalibration)	Response compares two concepts or ideas.	3.7%	13.8%	“I feel as if I do not have a wide enough range of physics knowledge for it to intervene in my everyday life <i>but I could</i> possibly have that vast amount of knowledge and genuine understanding after this semester concludes.” “This course felt very applicable to everyday life, <i>which was different</i> than previous physics courses.”
conceptual knowledge (reconceptualization)	Response references their knowledge of the concepts of physics.	19.6%	19.1%	“Every time I <i>learn something new</i> , I gain <i>new information</i> to inform my view of the world around me.” “I definitely think I have <i>learned more physics</i> this semester, but the scope of my knowledge to what more there is to learn is so great, that it is almost negligible.”
learning motivation (reprioritization)	Response references student’s motivation (or lack thereof) to learn physics.	5.3%	5.8%	“ <i>I have to take this course</i> as a requirement for ROTC. I am so bad at math and science. <i>This has nothing to do</i> with my actual major. This course will hurt my brain more than anything else.” “I am neutral, and still neutral about it. <i>I don’t really like</i> associating school related things with my everyday life.”
real-world (recalibration)	Response references real world applications of physics.	35.2%	34.5%	“I feel that studying physics will give me a better understanding of <i>the world around me</i> . Which in turn will somewhat change my views on how the world works.” “I selected the same answer (4) for statement 28 because I think physics is important in a lot of <i>our daily activities</i> and I have always thought this.”
worldview (reprioritization)	Response references the student’s perception, engagement, or interaction with the world. This includes absolute statements (unqualified by context or conditions).	63.3%	32.7%	“I think that my views would probably stay the same because <i>I think I know everything that I need to know.</i> ” “In the beginning I felt neutral about this statement, I now agree with this statement because <i>physics is involved with everything.</i> ”

to identify themes that occurred frequently in students’ reasoning. We collaboratively defined these themes, merging similar ideas into higher-level concepts that occurred most frequently. These definitions became the basis of our analysis of student responses here. This article discusses six of these themes: cause and effect (a response that references a causal relationship), comparison (a response that compares two concepts or ideas), conceptual knowledge (a response that references the student’s knowledge of physics concepts), learning motivation (a response that references the student’s motivation, or lack thereof, to learn physics), real-world (a response that references real world applications of physics), and worldview (a response that references the student’s perception, engagement, or interaction with the world). This last theme includes responses with absolute statements (unqualified by conditions or context) such as, “Physics is all around us.”

We used these definitions to tag a common sample of 94 randomly selected student responses to assess inter-rater agreement using an intraclass correlation coefficient (ICC) [11]. We chose this metric since ICCs are structured to ac-

count for an unbalanced distribution of positive and negative ratings (i.e., whether a given tag is applied to a response). We arrive at $ICC_{3k} = 0.68$ ($F = 3.16$, 95% confidence interval $0.55 - 0.79$, $p < 0.01$), indicating moderate agreement. Discrepancies in rater agreement tended to manifest from two raters assigning partially matching sets of tags to a given response. For example, one rater might assign “comparison,” “real-world,” and “worldview” to a response, while another might assign “cause and effect,” “real-world,” and “worldview” to the same response, both of which seemed appropriate upon subsequent review. Therefore, the frequency of tags presented in Table I likely represents an underestimate of tag frequency; in the next stage of our study, we plan to refine our tag criteria and improve our inter-rater agreement. With this caveat, we proceeded with each author tagging a unique subset of FRQ responses.

III. THEMES OF STUDENT RESPONSES

Table I presents the themes we identified in student responses to the FRQ, the definitions we used to guide the tagging process, the frequency with which each theme was tagged in Q28 in each survey, and two sample responses (one from PRE-EXP and one from RETRO-POST) for each theme. We have categorized most themes based on the established forms of RSB [7] discussed in Section I that are known to be responsible for the downshift between PRE and RETRO responses. We categorize “comparison” and “real-world” under recalibration, since they represent the students’ standards for making comparisons and identifying connections. We categorize “learning motivation” and “worldview” under reprioritization, since they represent the values that drive students to learn, assign merit to specific capacities, and formulate absolute statements. We categorize “conceptual knowledge” under reconceptualization, since it captures how students expect an increase in physics knowledge to transform their perceptions of physics. The “cause and effect” theme captures a structure frequently found in student responses, rather than content. We expect this theme to be useful when correlated with other themes in further analysis.

Examining the frequency with which each theme occurs in the PRE-EXP and RETRO-POST surveys reveals insight into how students’ reasoning changed between survey administrations. The “comparison” theme occurred much more frequently in RETRO-POST than PRE-EXP. We believe this indicates that students developed more standards for making comparisons during the semester, likely due to the acquisition of additional physics knowledge.

This interpretation is supported by the steady frequency with which the “conceptual knowledge” theme occurs (about 20% in both the PRE-EXP survey and the RETRO-POST survey). These students seemed to expect that acquisition of physics knowledge would directly change their perceptions of the subject.

In contrast, the “worldview” theme occurred much less frequently in the RETRO-POST responses than in the PRE-EXP responses (dropping from approximately two-thirds to less than one-third of student responses). It seems that, after a semester of instruction, the students found it less necessary to reference their view of the world or to make absolute statements, indicating a change in how they derive motivation from their values, assign merit, or formulate absolute statements. Compared with the rise in the frequency with which they make comparisons and the steady role that conceptual knowledge seems to play, we speculate that this shift might be attributable to students’ replacing general absolute statements with experience-based conceptual knowledge.

The “real-world” tag occurred with roughly equal frequency in the PRE-EXP and RETRO-POST responses. This is unsurprising, as Q28 directly solicits students’ thoughts about the relationship between physics knowledge and the

functioning of the (real) world. Students therefore used their standards for evaluating connections between physics and the real world with relatively equal frequency in the two surveys.

Finally, “learning motivation” occurred with similar low frequencies between PRE-EXP and RETRO-POST. We expect this theme to occur more frequently in other statements from the Real World Connection category.

IV. CONCLUSION AND FURTHER QUESTIONS

We have confirmed the presence of response-shift bias in the CLASS across multiple physics courses by observing a downshift between pre-instructional scores and retrospective pre-instructional scores. We then sought to answer the question, “How does response-shift bias manifest when students explain their answers to the CLASS during a PRE-EXP and RETRO-POST survey administration?” We prompted students to explain their reasoning behind giving the same or different answers on different passes of the survey (comparing pre-instructional answer with expected post-instructional answers, and post-instructional answer with retrospective answers). Examining these explanations revealed several trends that might help us explain the manifestation of response-shift bias. The students employed fewer absolute statements about their worldview (reprioritization), developed more standards for making comparisons (recalibration), and expected that learning more physics concepts would lead directly to different perceptions of physics (reconceptualization).

This preliminary discussion of our work in progress is limited to a single item on the CLASS, and we expect analysis of additional items to build on and refine these themes. We also expect improvement in our inter-rater agreement, which we will continue to assess. We plan to examine each of these themes by more finely classifying student responses. For example, student responses related to conceptual knowledge could be further classified in terms of whether students learned physics concepts as well as they expected to, or the degree to which the students seem to value knowing physics concepts. Additionally, we will use these themes as fodder for interviews to solicit further reflection from students. Such interviews will include asking students to elaborate on the differences between their PRE and RETRO responses to probe response-shift bias more directly.

ACKNOWLEDGMENTS

We are grateful to our department colleagues who administered this survey in their courses. The final author thanks Michelle Kuchera for invaluable assistance in setting up his research group. This work was supported by the University of North Florida Office of Undergraduate Research.

-
- [1] W.K. Adams, K.K. Perkins, N.S. Pdolefsky, M. Dubson, N.D. Finkelstein and C.E. Wieman, *Phys. Rev. ST Phys. Educ. Res.* **2**, 1 (2006).
- [2] A. Madsen, S.B. McKagan, and E.C. Sayre, *Phys. Rev. ST Phys. Educ. Res.* **11**, 1 (2015).
- [3] W.B. Lane and R.A. Adhikari, *PER Conference* (2018).
- [4] G.S. Howard and P.R. Dailey, *Journ. of App. Psych.* **64**, 144 (1979).
- [5] R.P. Hechter, *Journ. Sci. Teach. Educ.* **22**, 187 (2011).
- [6] J. Drennan and A. Hyde, *Asst. Eval. Higher Educ.* **33**, 699 (2008).
- [7] T. D. Little, R. Chang, B. K. Gorrall, L. Waggenspack, E. Fukuda, P. J. Allen, and G. G. Noam, *Int. Journ. Beh. Dev.* **44**, 2 (2020).
- [8] H. W. Franz, C. Kaletka, B. Pelka, and R. Sarcina, *Growing Experience: From Unconscious Incompetence to Unconscious Competence*, in *Building Leadership in Project and Network Management* (pages 209–226, Springer, 2018).
- [9] L. Deslauriers, Ls. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, *Proc. of the Natl. Acad. of Sci.* **116**, 39 (2019).
- [10] M. A. Cascio, E. Lee, N. Vaudrin, and D. A. Freedman, *Field Methods* **31**, 2 (2019).
- [11] T. K. Koo and M. Y. Li, *Journ. Chir. Med.* **15**, 2 (2016).