Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom

Louis Deslauriersa,1, Logan S. McCartya,b, Kelly Miller, Kristina Callaghanb, and Greg Kestina

Students learn more when they are actively engaged in the classroom than they do in a passive lecture environment. Extensive research supports this observation, especially in college-level science courses (1–6). Research also shows that active teaching strategies increase lecture attendance, engagement, and students’ acquisition of expert attitudes toward the discipline (3, 7–9). Despite this overwhelming evidence, most instructors still use traditional methods, at least in large-enrollment college courses (10–12).

Why do these inferior methods of instruction persist? Instructors cite many obstacles preventing them from adopting active teaching strategies, such as insufficient time, limited resources, a lack of departmental support, concerns about content coverage, and concerns about evaluations of their teaching (13–18). They also perceive that students resist active teaching strategies and prefer traditional methods (10, 14, 17, 19–22). Indeed, one-third of instructors who try active teaching eventually revert to passive lectures, many citing student complaints as the reason (23). Instructors report that students dislike being actively engaged during a semester-long course, their learning may be impaired during the initial part of the course. We discuss strategies that instructors can use, early in the semester, to improve students’ response to being actively engaged in the classroom.

Materials and Methods

Our study sought to measure students’ perception of learning when active learning alone is toggled on and off. This contrasts with typical educational interventions that include active engagement as one component of many changes to a course. We compared actual learning to students’ feeling of learning. There is nothing known about how students naturally react to active learning without any promotion from the instructor. In addition, previous studies used different course materials for active versus passive instruction, potentially confounding the effect of pedagogy with that of course materials.

In this report, we identify an inherent student bias against active learning that can limit its effectiveness and may hinder the wide adoption of these methods. Compared with students in traditional lectures, students in active classes perceived that they learned less, while in reality they learned more. Students rated the quality of instruction in passive lectures more highly, and they expressed a preference to have “all of their physics classes taught this way,” even though their scores on independent tests of learning were lower than those in actively taught classrooms. These findings are consistent with the observations that novices in a subject are poor judges of their own competence (27–29), and the cognitive fluency of lectures can be misleading (30, 31). Our findings also suggest that novice students may not accurately assess the changes in their own learning that follow from their experience in a class. These misperceptions must be understood and addressed in order for research-based active instructional strategies to be more effective and to become widespread.

Significance

Despite active learning being recognized as a superior method of instruction in the classroom, a major recent survey found that most college STEM instructors still choose traditional teaching methods. This article addresses the long-standing question of why students and faculty remain resistant to active learning. Comparing passive lectures with active learning using a randomized experimental approach and identical course materials, we find that students in the active classroom learn more, but they feel like they learn less. We show that this negative correlation is caused in part by the increased cognitive effort required during active learning. Faculty who adopt active learning are encouraged to intervene and address this misperception, and we describe a successful example of such an intervention.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: louis@physics.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821936116/-/DCSupplemental.

First published September 4, 2019.
learning (FOL) following each of 2 contrasting instructional methods: active learning (experimental treatment) and passive lecture (control). The entire protocol was repeated twice in physics courses taught during fall and spring at Harvard University. These calculus-based introductory courses cover topics in mechanics at a level appropriate for physics majors. Classes meet for 90 min twice each week, during a semester lasting 15 wk. The regular instructor for these courses had used a consistent syllabus and instructional approach for a number of years prior to the study and continued the same approach in both courses described here. Typical class meetings consisted of chalkboard lectures enhanced with frequent physics demonstrations, along with occasional interactive quizzes or conceptual questions. In the instructional taxonomy of Stains (12) this approach would likely be classified as interactive lecture, with lecturing as the primary modality, supplemented by student in-class activities. Consequently, while active learning was already a part of the instructional style during the semester, students in the experimental group had to adjust to an increase in the amount of active learning, while those in the control group had to adjust to a complete elimination of any active engagement.

Although most of the students in these courses were considering majoring in physics, fewer than one-third actually did so; the others majored in life sciences, math, engineering, computer science, economics, or other fields. Harvard also offers an alternative introductory mechanics course that includes advanced topics like Lagrangian mechanics, and this honors-level course tends to attract the most well-prepared physics majors. However, there is a more diverse range of students in the courses studied here. Indeed, although the students in the more advanced course are often quite exceptional, the students in this study have backgrounds comparable to those of physics majors at other major research universities. For instance, the students who took part in this study completed the Force Concept Inventory (FCI), which measures basic conceptual knowledge about mechanics (32), and the Colorado Learning Attitudes towards Science Survey (CLASS), which measures the extent to which students’ perceptions about physics are similar to those of experts (7, 8). The pretest FCI scores in this study (Table 1) are similar to those clustered near the high end of the distribution of university scores in the metaanalysis published by Hake (1), which confirms that the students in our study have high school preparation comparable to that at other top universities. The CLASS survey is perhaps more relevant as it measures expert thinking in physics instead of only background physics knowledge. The pretest CLASS scores in this study (Table 1) are comparable to those of first-year physics majors (or intended physics majors) at the University of Colorado (33), the University of California San Diego (34), or the University of Edinburgh (35).

The experimental intervention took place during 2 consecutive class meetings in week 12 of each course. Students were randomly assigned to 2 groups and told to report to 2 different classrooms: room A with instructor A and room B with instructor B. For the first class meeting, on the topic of statics, students learned about these concepts and equations along with example problems targeting specific concepts and equations along with example problems targeting specific concepts and equations. During each class meeting, the instructor presented slides based on the handouts and the instructor, and only active engagement with the material was toggled on and off. Previous well-controlled studies that compared active versus passive learning, such as the studies included in ref. 4, used distinctly different class materials with each group, potentially confounding active engagement with changes in class content (3). Likewise, studies that compared students’ responses to active versus passive learning typically did not use precisely the same class content. Students who claimed to prefer one mode of instruction over the other might have been responding to differences in content or class materials in addition to differences in the amount of active engagement.

### Results and Discussion

At the end of each class period, students completed a brief survey to measure their FOL followed by a multiple-choice TOL.

<table>
<thead>
<tr>
<th>Measure of background and proficiency</th>
<th>Group A</th>
<th>Group B</th>
<th>t</th>
<th>df*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring semester (enrollment: 65 students)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCI pretest score (0–30)</td>
<td>23.3</td>
<td>6.57</td>
<td>24.8</td>
<td>4.35</td>
<td>−0.96</td>
</tr>
<tr>
<td>CLASS pretest score (%)</td>
<td>75.7</td>
<td>10.04</td>
<td>73.4</td>
<td>11.93</td>
<td>0.82</td>
</tr>
<tr>
<td>Average of first 2 midterms (%)</td>
<td>77.9</td>
<td>13.4</td>
<td>78.1</td>
<td>14.4</td>
<td>−0.07</td>
</tr>
<tr>
<td>Fall semester (enrollment: 92 students)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCI pretest score (0–30)</td>
<td>23.6</td>
<td>4.19</td>
<td>24.6</td>
<td>3.40</td>
<td>−1.05</td>
</tr>
<tr>
<td>CLASS pretest score (%)</td>
<td>77.7</td>
<td>9.84</td>
<td>75.5</td>
<td>14.04</td>
<td>0.75</td>
</tr>
<tr>
<td>Average of first 2 midterms (%)</td>
<td>70.5</td>
<td>13.0</td>
<td>72.02</td>
<td>13.9</td>
<td>−0.54</td>
</tr>
</tbody>
</table>

*Some study participants did not have pretest data; all 149 participants had midterm scores.
Students rated their level of agreement on a 5-point Likert scale, with 1 representing strongly disagree and 5 representing strongly agree. Students first evaluated the statement “This class mostly involved me as a listener while the instructor presented information.” As expected, the students in the passive lecture agreed more strongly (mean = 3.9) than those in the active classroom (mean = 2.9, \( P < 0.001 \)). Note that even in the experimental group, about 50% of the class time featured the instructor giving concise, targeted feedback as minilectures following each group activity (3, 6, 36). The students then assessed their own FOL by rating their level of agreement with 4 additional statements, each of which probed some aspect of their perceived learning from the class. The primary FOL item asked students to evaluate the statement “I feel like I learned a great deal from this class.” The remaining FOL questions were highly correlated with this primary question, so we could use either this question alone or a composite of all 4 survey items to measure students' overall FOL. Fig. 1 lists the 4 FOL questions asked in the survey.

The subsequent tests of learning (1 on statics and 1 on fluids) each consisted of 12 multiple-choice questions. The students were encouraged to try their best on each TOL and were told that they would be good practice for the final examination but were reminded that their score on the TOL would not directly affect their course grade. Students were also told that they would receive participation points toward their final grade for completing the TOL and the FOL surveys. (The TOL and FOL questions are provided in SI Appendix.)

The bar graphs shown in Figs. 1 and 2 highlight several aspects of these FOL and TOL results. We note, in particular, the following observations (all of which are confirmed by a more detailed statistical analysis): 1) All of the FOL responses show a consistent student preference for the passive lecture environment. 2) Scores on the TOL, by contrast, are significantly higher in the active classroom. 3) These trends are similar for both the statics and fluids topics. Given the crossover study design (Table 2), it appears that the shift in TOL and FOL scores between passive and active learning was not strongly affected by the choice of topic, instructor, or classroom.

We constructed linear regression models (fixed-effects models) to identify the factors contributing to these observed differences in TOL and FOL scores. To control for student-level variation, we included 3 measures of students’ individual background and proficiency in physics: the FCI (34), the CLASS (7), and the average scores on 2 midterm examinations that took place prior to the study. The descriptive statistics summarized in Table 1 confirm successful randomization at the student level for these measures.

Table 3 summarizes these statistical models. Model 1 predicts students’ overall FOL, which is a composite of the FOL survey responses weighted according to a principal components analysis. (The entire analysis is virtually identical if the primary FOL question 2 is used alone in place of this composite variable.) The students in active classrooms reported more than half an SD (0.56) lower FOL compared with those in passive lectures. Model 2 predicts students’ performance on the TOL. In this case, students in active classrooms scored almost half an SD (0.46) higher on the examination. These results are highly significant (\( P < 0.001 \)). In addition, the crossover study design allows us to control for any additional person-level variation by adding a categorical variable for each individual student (treating each student as his or her own control); we find no meaningful change using these additional covariates. Conversely, as expected for a randomized experiment, if we remove from the statistical model all student-level covariates (CLASS score, FCI score, midterm average, and gender) the point estimates of the effects of active learning also show no meaningful change (less than half the SE).

In educational research, a question often arises whether to analyze the data at the individual student level or at the group level (typically by classroom or by school). The convention in recent research on higher education, e.g., ref. 4, is that if preexisting

### Table 2. Randomized experimental design for the study

<table>
<thead>
<tr>
<th>Class topic</th>
<th>Group A: instructor A in classroom A</th>
<th>Group B: instructor B in classroom B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static equilibrium</td>
<td>Passive (control)</td>
<td>Active (treatment)</td>
</tr>
<tr>
<td>Fluids</td>
<td>Passive (control)</td>
<td>Active (treatment)</td>
</tr>
</tbody>
</table>

Fig. 1. A comparison of performance on the TOL and FOL responses between students taught with a traditional lecture (passive) and students taught actively for the statics class. Error bars show 1 SE.
groups are exposed to treatment versus control conditions, the statistical analysis should account for these clusters, since both randomization and treatment are applied at the group level. Many studies of college science courses do not correctly account for clustering, and indeed Freeman et al. (4) had to correct for this oversight in their metaanalysis. On the other hand, if students are individually randomized, or the experiment is a crossover study in which each student receives both conditions, then an individual-level analysis is appropriate, even if the treatment is (inevitably) delivered at the class level. This convention is rigorously justified (39) as long as peer effects are negligible. In our study, the crossover design controls for peer effects at the linear level since students have the same peer group under both active and passive conditions. A remaining concern could be a nonlinear interaction between peer effects and the 2 styles of teaching—for instance, if students openly expressed disdain for the pedagogy only in the active classroom. The physics courses used in this study are routinely video-recorded, and videos of the experiment show no overt peer interactions that could affect the outcomes in active versus passive classrooms. Students took the FOL and TOL surveys immediately at the end of each class period, so there could be no peer effects outside the classroom. Moreover, as shown in SI Appendix, even if we postulate an extremely large unobserved peer effect on active versus passive learning, our results would still remain highly significant ($P < 0.001$).

Having observed this negative correlation between students’ FOL and their actual learning, we sought to understand the causal factors behind this observation. A survey of the existing literature suggests 2 likely factors: 1) the cognitive fluency of lectures can mislead students into thinking that they are learning more than they actually are (30, 31) and 2) novices in a subject have poor metacognition and thus are ill-equipped to judge how much they have learned (27–29). We also propose a third factor: 3) students who are unfamiliar with intense active learning in the college classroom may not appreciate that the increased cognitive struggle accompanying active learning is actually a sign that the learning is effective. We describe below some evidence suggesting that all 3 factors are involved and propose some specific strategies to improve students’ engagement with active learning.

![Figure 2](image.png) 

**Fig. 2.** A comparison of performance on the TOL and FOL responses between students taught with a traditional lecture (passive) and students taught actively for the fluids class. Error bars show 1 SE.

<table>
<thead>
<tr>
<th>Table 3. Standardized coefficients for linear regression models predicting students’ overall FOL (model 1) and performance on the TOL (model 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression parameter</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Passive (0) versus active (1)</td>
</tr>
<tr>
<td>Topic (fluids = 0; statics = 1)</td>
</tr>
<tr>
<td>Semester (spring = 0; fall = 1)</td>
</tr>
<tr>
<td>Instructor (A = 0; B = 1)</td>
</tr>
<tr>
<td>CLASS pretest ($z$ score)</td>
</tr>
<tr>
<td>FCI pretest ($z$ score)</td>
</tr>
<tr>
<td>Average of first 2 midterms ($z$ score)</td>
</tr>
<tr>
<td>Gender (female = 0; male = 1)</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

Both models control for class content (fluids versus statics), semester, instructor, and student data (CLASS score, FCI score, midterm performance, and gender). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$. Results are unaffected by the choice of ordinary least-squares or robust SEs (38). The raw FOL and TOL scores were pooled before standardization, which accounts for the effect of the “topic” covariate (fluids vs. statics).
One of the most important metacognitive cues is the apparent fluency of cognitive tasks. Perceived fluency has broad impacts on judgment and perception (31). In the laboratory context, previous research has compared students’ perceived ability to recall facts from a 5-min video from a fluent versus a disfluent lecturer (30). The disfluent lecturer—who avoided eye contact, did not speak clearly, and lacked flow—led to lower perceived retention even though students’ actual recall was the same as it was with the fluent lecturer. Research has also shown that when students are forced to struggle through something that is difficult, the consequent disfluency leads to deeper cognitive processing (31, 40). In our study, students in the actively taught groups had to struggle with their peers through difficult physics problems that they initially did not know how to solve. The cognitive effort involved in this type of instruction may make students frustrated and painfully aware of their lack of understanding, in contrast with fluent lectures that may serve to confirm students’ inaccurately inflated perceptions of their own abilities.

To learn more about our students’ perceptions, we conducted follow-up one-on-one, structured interviews with a subset of students from the study (17 students total). The students were drawn from both semesters and provided a representative sample of the entire population as measured by their CLASS scores, FCI scores, and final course grades. Consistent with the literature, most students (15 of 17) found the instruction in the active classrooms disjoined and lacking in flow when compared with the more fluent passive lecture. Students also cited the frequent interruptions that accompanied each transition from group activities to instructor feedback (14 responses), a concern that their errors made during class would not be corrected (10 responses), and a general feeling of frustration and confusion (14 responses) when discussing their concerns about the actively taught classes. In addition, although conventional wisdom suggests that students do not always enjoy working in groups, none of the students raised group work as an issue during interviews. In contrast, all but 1 of the students found the passive lecture more enjoyable and easier to follow. At the end of each interview, students were shown the results of the study. After commenting on the results, each student was asked “If seeing these results will impact the way you study,” and 14 out of 17 students said that it would.

In addition to suggesting that active learning produces better FOL and perceived fluency with a linear regression model predicting students’ FOL, given by FOL question 2: “I feel like I learned a great deal from this lecture.” Students who perceived the instructor to be highly fluent, as measured by agreement with the statement “The instructor was effective at teaching,” reported more than half an SD (0.51) higher FOL compared with those who perceived the instructor as disfluent (P < 0.001). Notably, the type of instruction (active vs. passive) was not significant in predicting FOL; only the perceived fluency of the instructor was relevant. We conducted additional one-on-one, structured interviews to validate that students interpret the question about teaching effectiveness as a measure for fluency of instruction. These interviews revealed that students interpret this question primarily as 1) clarity of explanations, 2) organization of presentation, and 3) smooth flow of instruction. In addition, students presented several scenarios in which they could imagine reporting that a teacher was highly effective even if they personally did not feel they learned very much—for instance, if they were not sufficiently prepared for a class or too tired to pay close attention. The strong correlation between students’ FOL and the effectiveness/fluency of instruction suggests that greater perceived fluency is related to higher perceived FOL.

A second factor that could account for our observed results is that novices (such as the students in our study) generally have poor metacognition and are not good at judging their own learning. “The same knowledge that underlies the ability to produce correct judgment, is also the knowledge that underlies the ability to recognize correct judgment. To lack the former is to be deficient in the latter.” (27) Although this well-known effect predicts that students’ FOL may be unreliable, it does not predict whether these feelings should be biased in favor of active versus passive styles of teaching. We investigated this hypothesis by adding a nonlinear interaction term to model 2, described above, that predicts students’ performance on the TOL. We found a moderately significant (P < 0.05) interaction between students’ background physics knowledge as measured by the FCI and their FOL as measured by question 2: “I feel like I learned a great deal from this lecture.” The sign of this interaction was positive, which means that students with more prior expertise had a stronger (more positive) correlation between FOL and actual performance on the test. Combining this observation with that in the previous paragraph, we propose that novice students are poor at judging their actual learning and thus rely on inaccurate metacognitive cues such as fluency of instruction when they attempt to assess their own learning. These 2 factors together could explain the strong, overall negative correlation we observed in this study.

A final factor could be that the students in this study had little prior experience with fully student-centered classrooms in a college environment (12). As suggested by the interviews described above, when students experienced confusion and increased cognitive effort associated with active learning, they perceived this disfluency as a signal of poor learning, while in fact the opposite is true. It is unlikely that the sheer novelty of student-centered active learning alone can account for students’ negative response to this mode of instruction. First, as mentioned above, both the experimental (active) and control (passive) groups experienced a change from the usual instructional approach in these courses: in the passive group, students experienced none of the small-group activities that were interspersed in the usual course lectures. Second, one can imagine a thought experiment in which students are given one-on-one tutoring with an expert tutor for 1 wk of a course. This would constitute a dramatic change from their usual classroom experience, but nearly all students would likely prefer this style of instruction—which is demonstrably superior (41, 42)—to their familiar lectures.

Based on the 3 factors discussed above, it is likely that a significant part of students’ comparably negative response to this type of active style of instruction is a result of the disfluency they experience in this cognitively demanding environment. We carried out a semester-long intervention to see if these attitudes could be changed. Near the beginning of a physics course that used the same active learning strategy described here, the instructor gave a 20-min presentation that started with a brief description of active learning and evidence for its effectiveness. He then presented additional detail about the connections between perceived fluency, FOL, and actual learning, including a discussion of the negative correlations we observed in this study. (The transcript for this presentation can be found in SI Appendix.) Students’ questions and discussion following the presentation indicated that they were most interested in the idea that fluency and FOL can often be misleading. Students indicated that this knowledge would be useful for understanding how to approach active learning. At the end of the semester, over 65% of students reported on a survey that their feelings about the effectiveness of active learning significantly improved over the course of the semester. A similar proportion (75%) of students reported that the intervention at the beginning of the semester helped them feel more favorably toward active learning during lectures.

As the success of active learning crucially depends on student motivation and engagement, it is of paramount importance that students appreciate, early in the semester, the benefits of struggling with the material during active learning. If students are misled by their inherent response into thinking that they are not learning, they will not be able to self-regulate, and they will not learn as
In conclusion, we find that students’ perception of their own learning can be antecedent with their actual learning under well-controlled implementations of active learning versus passive lectures. These results point to the importance of preparing and coaching students early in the semester for active instruction and suggest that instructors should persuade students that they are benefiting from active instruction. Without this preparation, students can be misled by the inherent disfluency associated with the sustained cognitive effort required for active learning, which in turn can have a negative impact on their actual learning. This is especially important for students who are new to fully student-centered active learning (12), as were the students in this study. These results also suggest that student evaluations of teaching should be used with caution as they rely on students’ perceptions of learning and could inadvertently favor inferior passive teaching methods over research-based active pedagogical approaches (44, 45)—a superstar lecturer could create such a positive FOL that students would choose those lectures over active learning. In addition, given the powerful general influence of fluency on metacognitive judgments (31), we expect that these results are likely to generalize to a variety of college-level subjects.

ACKNOWLEDGMENTS. We acknowledge significant contributions from Eric Mazur and David J. Morin; along with valuable discussions with Gary King, Erin Driver-Linn, Andrew Ho, Edward J. Kim, Jon R. Star, Federico Capasso, Dustin Tingley, Philip M. Sadler, and Melissa Franklin.