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

Louis Deslauriers[a,1], Logan S. McCarty[a,b], Kelly Miller[c], Kristina Callaghan[a], and Greg Kestin[a]

[a]Department of Physics, Harvard University, Cambridge, MA 02138; [b]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; and [c]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved August 13, 2019 (received for review December 24, 2018)

We compared students' self-reported perception of learning with their actual learning under controlled conditions in large-enrollment introductory college physics courses taught using 1) active instruction (following best practices in the discipline) and 2) passive instruction (lectures by experienced and highly rated instructors). Both groups received identical class content and handouts, students were randomly assigned, and the instructor made no effort to persuade students of the benefit of either method. Students in active classrooms learned more (as would be expected based on prior research), but their perception of learning, while positive, was lower than that of their peers in passive environments. This suggests that attempts to evaluate instruction based on students' perceptions of learning could inadvertently promote inferior (passive) pedagogical methods. For instance, a superstar lecturer could create such a positive feeling of learning that students would choose those lectures over active learning. Most importantly, these results suggest that when students experience the increased cognitive effort associated with active learning, they initially take that effort to signify poorer learning. That disconnect may have a detrimental effect on students' motivation, engagement, and ability to self-regulate their own learning. Although students can, on their own, discover the increased value of 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.

scientific teaching | undergraduate education | evidence-based teaching | constructivism

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 forced to interact with one another (15, 17, 24), they resent the increase in responsibility for their own learning (21, 22), and they complain that "the blind can't lead the blind" (19). More recent literature shows that if instructors explain and facilitate active learning, student attitudes toward it can improve over the course of a semester (25, 26). However, these studies do not measure students' inherent, unbiased response to being actively engaged with the material. 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.

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

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.

Author contributions: L.D., L.S.M., and K.C. designed research; L.D., L.S.M., K.M., K.C., and G.K. performed research; L.D., L.S.M., and K.M. analyzed data; and L.D., L.S.M., K.M., and G.K. wrote the paper.

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).

[1]To 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.

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

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 mode, 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 students, leaving 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 about 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 specific background 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 static equilibrium, instructor A used active learning, while instructor B taught the same topic using a passive lecture. For the second class meeting, on the topic of fluids, instructor A used a passive lecture while instructor B used active learning. At the end of each class period, students completed a brief survey about their perceptions of the class and their FOL, followed by a multiple-choice test of learning (TOL). Table 2 summarizes the experimental design. As this study involved classroom-based research using normal educational practices, it was exempt from Institutional Review Board oversight. We informed students that they would be learning the same material in both groups with different instructional methods, that they would all experience both instructional approaches, and that their scores on the TOL would not have any impact on their course grades. Nearly all students consented to participate, so attrition was negligible: only 8 out of 157 opted out or failed to complete the study.

The study design featured a number of controls to ensure consistency and avoid bias: 1) Both instructors had extensive, identical training in active learning, using best practices as detailed in prior research (3, 6, 36). 2) Both instructors also had comparable experience in delivering fluent, traditional lectures. 3) The lecture slides, handouts, and written feedback provided during each class were identical for active instruction and for passive lecture. 4) Students were individually randomly assigned to 2 groups, and these groups were indistinguishable on several measures of physics background and proficiency (Table 1). 5) Each student experienced both types of instruction in a crossover study design that controls for other possible variation between students. 6) Students had no exposure to either of the instructors before the experimental intervention. 7) The entire protocol was repeated in 2 different courses with the same results; a total of 149 students participated. 8) The instructors did not see the TOLs, which were prepared independently by another author. 9) The author of the TOLs did not have access to the course materials or lecture slides and wrote the tests based only on a list of detailed learning objectives for each topic.

Students in both groups received identical paper handouts with key concepts and equations along with example problems targeting specific learning objectives. The handouts had blank space for students to take notes and fill in answers to these sample problems. (All materials are provided in *SI Appendix*.) In the control group, the instructor presented slides based on the handouts, gave explanations and demonstrations, and solved the example problems while students listened and filled in the answers along with the instructor. Emphasis was placed on maximizing the fluency with which the information was delivered. The use of handouts and focus on problem-solving was different from the usual lectures in these courses. Using the taxonomy of Stains (12), these classes in the control group were strictly didactic in approach, with none of the supplemental group activities found in the usual class meetings. In the experimental group, the instructor actively engaged the students using the principles of deliberate practice (3, 36, 37): students were instructed to solve the sample problems by working together in small groups while the instructor roamed the room asking questions and offering assistance. After the students had attempted each problem, the instructor provided a full solution that was identical to the solution given to the control group. Students were actively engaged throughout the class period, making the experimental group fully student-centered (12). The crucial difference between the 2 groups was whether students were told directly how to solve each problem or were asked to try to solve the problems themselves in small groups before being given the solution. In other words, students in both groups received the exact same information from 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 1. Descriptive statistics for the randomized groups used in the study

	Group A		Group B				
Measure of background and proficiency	Mean	SD	Mean	SD	*t*	*df**	*p*
Spring semester (enrollment: 65 students)							
FCI pretest score (0–30)	23.3	6.57	24.8	4.35	−0.96	49	0.34
CLASS pretest score (%)	75.7	10.04	73.4	11.93	0.82	58	0.41
Average of first 2 midterms (%)	77.9	13.4	78.1	14.4	−0.07	63	0.94
Fall semester (enrollment: 92 students)							
FCI pretest score (0–30)	23.6	4.19	24.6	3.40	−1.05	70	0.30
CLASS pretest score (%)	77.7	9.84	75.5	14.04	0.75	78	0.46
Average of first 2 midterms (%)	70.5	13.0	72.02	13.9	−0.54	84	0.59

*Some study participants did not have pretest data; all 149 participants had midterm scores.

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

Table 2. Randomized experimental design for the study

Class topic	Group A: instructor A in classroom A	Group B: instructor B in classroom B
Static equilibrium	Active (treatment)	Passive (control)
Fluids	Passive (control)	Active (treatment)

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 FOL and TOL 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

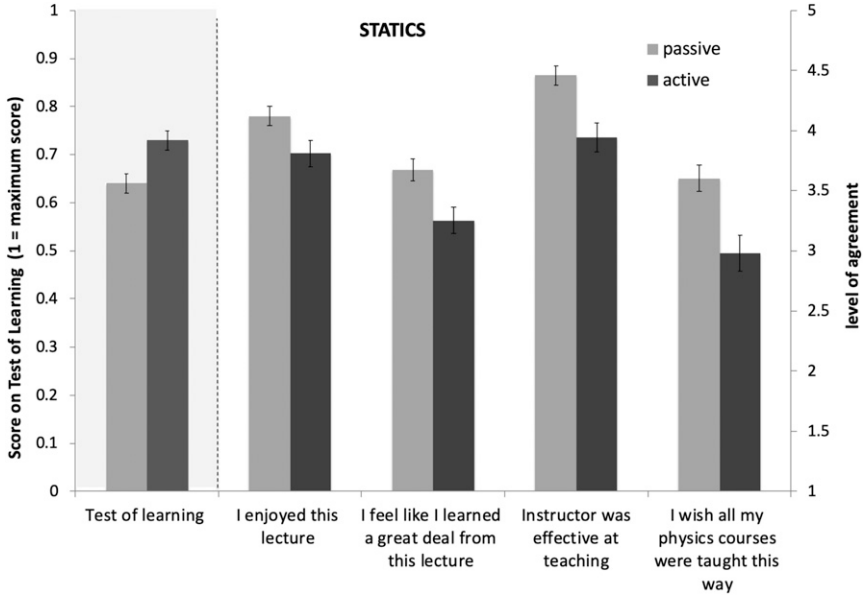

Fig. 1. A comparison of performance on the TOL and FOL responses between students taught with a traditional lecture (passive) and students taught actively for the statics class. Error bars show 1 SE.

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

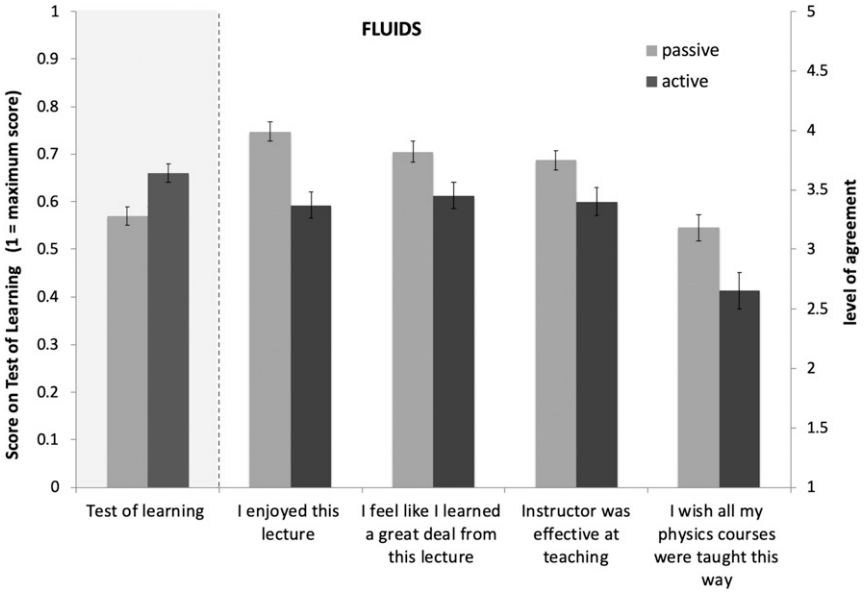

Fig. 2. A comparison of performance on the TOL and FOL responses between students taught with a traditional lecture (passive) and students taught actively for the fluids class. Error bars show 1 SE.

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.

Table 3. Standardized coefficients for linear regression models predicting students' overall FOL (model 1) and performance on the TOL (model 2)

Regression parameter	Model 1: FOL (standardized *z* score)	Model 2: TOL (standardized *z* score)
Constant	0.34	−0.46
Passive (0) versus active (1)	−0.56***	0.46***
Topic (fluids = 0; statics = 1)	0.44**	0.44***
Semester (spring = 0; fall = 1)	−0.37*	−0.29*
Instructor (A = 0; B = 1)	−0.03	−0.12
CLASS pretest (*z* score)	0.01	0.00
FCI pretest (*z* score)	−0.07	0.25***
Average of first 2 midterms (*z* score)	−0.20*	0.26***
Gender (female = 0; male = 1)	−0.05	0.33*
R^2	0.17	0.39
RMSE	0.97	0.77

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).

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

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 disjointed 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, we investigated the connection between 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 intense style of active learning 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

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

successfully. In addition, during group work, poor attitudes or low engagement of a few students can have negative effects on other students in their groups. Thus, although students may eventually, on their own, discover the value of active learning during a semester-long course, their learning will be impaired during the first part of the course while they still feel the inherent disfluency associated with in-class activities.

We recommend that instructors intervene early on by explicitly presenting the value of increased cognitive efforts associated with active learning. Instructors should also give an examination (or other assessment) as early as possible so students can gauge their actual learning. These strategies can help students get on board with active learning as quickly as possible. Then, throughout the semester, instructors should adopt research-based explanation and facilitation strategies (26), should encourage students to work hard during activities, and should remind them of the value of increased cognitive effort. Instructors should also solicit frequent feedback such as "one-minute papers" throughout the course (43) and respond to students' concerns. The success of active learning will be greatly enhanced if students accept that it leads to deeper learning—and acknowledge that it may sometimes feel like exactly the opposite is true.

These recommendations should apply to other student populations and to other disciplines as the cognitive principles underlying these effects are not specific to physics or to the well-prepared students in this course. To illustrate this point, imagine a course with a different group of students, or in a different subject, that uses a highly effective interactive pedagogy with course materials tailored to its own student audience. Now bring in a fluent and charismatic lecturer with special knowledge of student thinking who uses the same materials but eliminates all interactive engagement from the course, consistent with the design of this study in which active learning alone is toggled on and off. As a specific example, consider Peer Instruction (2) with well-honed clicker questions that target common student difficulties and misconceptions. Instead of allowing students to answer and discuss these questions, the lecturer would describe and explain each of the answers. From the research reviewed in ref. 4, it is clear that students would learn less in the passive lecture environment. For instance, students deprived of active engagement with clicker questions could not discover their own misconceptions or construct their own correct explanations. Yet based on the cognitive principles discussed above, the fluent lecturer could address student difficulties and misconceptions in such a way as to make students feel like they learned a lot from the lecture. Indeed, given our observation that highly proficient students are better able to judge their own learning, it is reasonable to expect that students who are less well prepared than those in our study would show even larger discrepancies between actual learning and FOL.

In conclusion, we find that students' perception of their own learning can be anticorrelated 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 benefitting 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.

1. R. R. Hake, Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* **66**, 64–74 (1998).
2. C. H. Crouch, E. Mazur, Peer instruction: Ten years of experience and results. *Am. J. Phys.* **69**, 970–977 (2001).
3. L. Deslauriers, E. Schelew, C. Wieman, Improved learning in a large-enrollment physics class. *Science* **332**, 862–864 (2011).
4. S. Freeman *et al.*, Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410–8415 (2014).
5. J. M. Fraser *et al.*, Teaching and physics education research: Bridging the gap. *Rep. Prog. Phys.* **77**, 032401 (2014).
6. L. Deslauriers, C. Wieman, Learning and retention of quantum concepts with different teaching methods. *Phys. Rev. ST Phys. Educ.* **7**, 010101 (2011).
7. W. K. Adams *et al.*, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Phys. Rev. ST Phys. Educ.* **2**, 010101 (2006).
8. E. Brewe, L. H. Kramer, G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS. *Phys. Rev. ST Phys. Educ.* **5**, 013102 (2009).
9. J. Watkins, E. Mazur, Retaining students in science, technology, engineering, and mathematics (STEM) majors. *J. Coll. Sci. Teach.* **42**, 36–41 (2013).
10. C. Henderson, M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Phys. Rev. ST Phys. Educ.* **3**, 020102 (2007).
11. J. Handelsman *et al.*, Education. Scientific teaching. *Science* **304**, 521–522 (2004).
12. M. Stains *et al.*, Anatomy of STEM teaching in North American universities. *Science* **359**, 1468–1470 (2018).
13. C. Henderson, T. Stelzer, L. Hsu, D. Meredith, Maximizing the benefits of physics education research: Building productive relationships and promoting institutional change. *American Physical Society Forum on Education Newsletter*, Fall 2005, pp. 11–14. https://www.aps.org/units/fed/newsletters/fall2005/maximize.html. Accessed 20 June 2019.
14. M. Dancy, C. Henderson, Framework for articulating instructional practices and conceptions. *Phys. Rev. ST Phys. Educ.* **3**, 010103 (2007).
15. R. M. Felder, R. Brent, Navigating the bumpy road to student-centered instruction. *Coll. Teach.* **44**, 43–47 (1996).
16. D. U. Silverthorn, P. M. Thorn, M. D. Svinicki, It's difficult to change the way we teach: Lessons from the Integrative Themes in Physiology curriculum module project. *Adv. Physiol. Educ.* **30**, 204–214 (2006).
17. A. P. Fagen, C. H. Crouch, E. Mazur, Peer instruction: Results from a range of classrooms. *Phys. Teach.* **40**, 206–209 (2002).
18. C. Turpen, M. Dancy, C. Henderson, Faculty perspectives on using peer instruction: A national study. *AIP Conf. Proc.* **1289**, 325–328 (2010).
19. J. W. Belcher, Improving Student Understanding with TEAL [TEAL = Technology Enhanced Active Learning], *The MIT Faculty Newsletter*, vol. XVI, no. 2, 2003. http://web.mit.edu/fnl/vol/162/belcher.htm. Accessed 20 June 2019.
20. M. H. Dancy, C. Henderson, Beyond the individual instructor: Systemic constraints in the implementation of research-informed practices. *AIP Conf. Proc.* **790**, 113–116 (2005).
21. R. M. Felder, Random thoughts: Sermons for grumpy campers. *Chem. Eng. Educ.* **41**, 183–184 (2007).
22. R. M. Felder, Random thoughts: The link between teaching and research. 2. How to strengthen each without weakening the other. *Chem. Eng. Educ.* **44**, 213–214 (2010).
23. C. Henderson, M. Dancy, M. Niewiadomska-Bugaj, Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Phys. Rev. ST Phys. Educ.* **8**, 020104 (2012).
24. M. Vuorela, L. Nummenmaa, How undergraduate students meet a new learning environment? *Comput. Human Behav.* **20**, 763–777 (2004).
25. K. Nguyen *et al.*, Students' expectations, types of instruction, and instructor strategies predicting student response to active learning. *Int. J. Eng. Educ.* **33**, 2–18 (2017).
26. S. Tharayil *et al.*, Strategies to mitigate student resistance to active learning. *Int. J. STEM Educ.* **5**, 7 (2018).
27. J. Kruger, D. Dunning, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**, 1121–1134 (1999).
28. J. D. Bransford, A. L. Brown, R. R. Cocking, Eds., *How People Learn: Brain, Mind, Experience, and School* (National Academy Press, 1999).
29. S. R. Porter, Self-reported learning gains: A theory and test of college student survey response. *Res. High. Educ.* **54**, 201–226 (2013).
30. S. K. Carpenter, M. M. Wilford, N. Kornell, K. M. Mullaney, Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychon. Bull. Rev.* **20**, 1350–1356 (2013).

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.

31. D. M. Oppenheimer, The secret life of fluency. *Trends Cogn. Sci.* **12**, 237–241 (2008).
32. D. Hestenes, M. Wells, G. Swackhamer, Force concept inventory. *Phys. Teach.* **30**, 141–158 (1992).
33. K. K. Perkins, M. Gratny, Who becomes a physics major? A long-term longitudinal study examining the roles of pre-college beliefs about physics and learning physics, interest, and academic achievement. *AIP Conf. Proc.* **1289**, 253–256 (2010).
34. E. Gire, B. Jones, E. Price, Characterizing the epistemological development of physics majors. *Phys. Rev. ST Phys. Educ* **5**, 010103 (2009).
35. S. P. Bates, R. K. Galloway, C. Loptson, K. A. Slaughter, How attitudes and beliefs about physics change from high school to faculty. *Phys. Rev. ST Phys. Educ.* **7**, 020114 (2011).
36. D. J. Jones, K. W. Madison, C. E. Wieman, Transforming a fourth-year modern optics course using a deliberate practice framework. *Phys. Rev. ST Phys. Educ.* **11**, 020108 (2015).
37. K. A. Ericsson, R. Th. Krampe, C. Tesch-Römer, The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* **100**, 363–406 (1993).
38. H. White, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838 (1980).
39. A. Abadie, S. Athey, G. W. Imbens, J. Wooldridge, When should you adjust standard errors for clustering? (NBER Working Paper 24003, National Bureau of Economic Research, Cambridge, MA) https://dx.doi.org/10.3386/w24003 (November 2017).
40. C. Diemand-Yauman, D. M. Oppenheimer, B. E. Vaughan, Fortune favors the **bold** (*and the Italicized*): Effects of disfluency on educational outcomes. *Cognition* **118**, 111–115 (2011).
41. M. R. Lepper, M. Woolverton, "The wisdom of practice: lessons learned from the study of highly effective tutors" in *Improving Academic Achievement*, J. Aronson, Ed. (Academic Press, 2002), pp. 135–158.
42. W. B. Wood, K. D. Tanner, The role of the lecturer as tutor: Doing what effective tutors do in a large lecture class. *CBE Life Sci. Educ.* **11**, 3–9 (2012).
43. D. R. Stead, A review of the one-minute paper. *Active Learn. High. Educ.* **6**, 118–131 (2005).
44. B. Uttl, C. A. White, D. W. Gonzalez, Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* **57**, 22–42 (2017).
45. N. Kornell, H. Hausman, Do the best teachers get the best ratings? *Front. Psychol.* **7**, 570 (2016).

Downloaded from https://www.pnas.org by "WEBER STATE UNIVERSITY, LIBRARY SERIALS" on September 17, 2024 from IP address 137.190.85.112.