Understanding the relationship between student attitudes and student learning

Michael J. Cahill,[1,*] Mark A. McDaniel,[1,2] Regina F. Frey,[1,3,4] K. Mairin Hynes,[5]
Michelle Repice,[3] Jiuqing Zhao,[1] and Rebecca Trousil[6]

[1] *Center for Integrative Research on Cognition, Learning and Education, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, USA*
[2] *Department of Psychological and Brain Sciences, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, USA*
[3] *The Teaching Center, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, USA*
[4] *Department of Chemistry, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, USA*
[5] *Department of Physics, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, USA*
[6] *Department of Mathematics and Computer Science, Skidmore College, 815 N. Broadway, Saratoga Springs, New York 12866, USA*

(Received 6 July 2017; published 6 February 2018)

Student attitudes, defined as the extent to which one holds expertlike beliefs about and approaches to physics, are a major research topic in physics education research. An implicit but rarely tested assumption underlying much of this research is that student attitudes play a significant part in student learning and performance. The current study directly tested this attitude-learning link by measuring the association between incoming attitudes (Colorado Learning Attitudes about Science Survey) and student learning during the semester after statistically controlling for the effects of prior knowledge [early-semester Force Concept Inventory (FCI) or Brief Electricity and Magnetism Assessment (BEMA)]. This study spanned four different courses and included two complementary measures of student knowledge: late-semester concept inventory scores (FCI or BEMA) and exam averages. In three of the four courses, after controlling for prior knowledge, attitudes significantly predicted both late-semester concept inventory scores and exam averages, but in all cases these attitudes explained only a small amount of variance in concept-inventory and exam scores. Results indicate that after accounting for students' incoming knowledge, attitudes may uniquely but modestly relate to how much students learn and how well they perform in the course.

DOI: 10.1103/PhysRevPhysEducRes.14.010107

I. INTRODUCTION

Recent years have seen a growing interest in attitudes toward physics as an important component of student experiences and outcomes in introductory physics classes [1–7]. These attitude measures, such as the Colorado Learning Attitudes about Science Survey (CLASS) [2,8] and the Maryland Physics Expectations Survey (MPEX) [1], measure not necessarily the valence of one's feelings toward physics, but rather the extent to which one holds expertlike beliefs about and approaches toward physics. These attitude measures have garnered a great deal of research attention both because they are viewed as important outcomes in their own right [1,7] and because of the implicit assumption that these attitudes can influence student learning [1,4–9]. The attitude-learning link is important theoretically because confirming that attitudes predict learning (independent of potential confounds such as prior knowledge) would reinforce the importance of developing and implementing curricula that enhance student attitudes (i.e., lead to more expertlike perceptions of physics). Accordingly, the current study is designed to answer the following primary research question: Do incoming attitudes predict learning, after controlling for the potentially confounding influence of prior knowledge?

A. Associations between attitudes and learning

A general trend in the literature is that the types of curricula that produce higher learning gains on concept inventories also positively impact student attitudes.

*Corresponding author.
cahillmj@wustl.edu

Specifically, interactive-engagement curricula, relative to traditional-lecture curricula, produce both higher learning gains [10–11] and positive attitudinal effects [2–6], and these effects have been demonstrated within the same study [7]. These trends may contribute to the implicit assumption that student attitudes contribute to learning, but surprisingly little empirical evidence supports this assumption. A recent meta-analysis [12] of the CLASS and MPEX examined, among other issues, the relationship between these attitude measures and conceptual gains. This meta-analysis included studies examining correlations among various combinations of prescores, postscores, and gains on attitude measures and concept inventories, but only one study [2] directly analyzed the relationship between presemester attitudes and learning gains in the course. In this study [2], presemester (and postsemester) attitudes correlated significantly with learning gains on a mechanics concept inventory, the Force and Motion Conceptual Evaluation (FMCE) [13], suggesting that students' incoming attitudes may contribute to their learning [2].

The just-mentioned result provides preliminary support for the idea that student attitudes relate to learning, and the current study seeks to further solidify this finding in several key ways. First, replication of any finding is necessary, and in classroom research it is particularly important to understand how a given result may change across different learning contexts. In fact, the aforementioned meta-analysis [12] concludes that more work is necessary to understand this relationship between student beliefs and learning of physics. The previously reported correlations between incoming attitudes and learning gains [2] came from a single course, which was a large interactive-engagement course with a heavy emphasis on conceptual understanding. The current study seeks to conceptually replicate this finding with a different student population and across different learning contexts that vary in terms of content (classical mechanics versus electromagnetism, thermodynamics, and quantum mechanics) and instructional style (traditional lecture versus active physics). This allows us to examine the possibility that the attitude-learning link might vary across different types of courses. For example, one possibility is that attitudes may be more influential in interactive engagement courses, which place more emphasis on certain expertlike aspects of learning physics such as developing coherent conceptual structures. Additionally, in order to increase the reliability and generalizability of our findings, each analysis includes three semesters of data, with each semester including multiple sections and instructors.

Second, although the use of normalized gains as the measure of learning [2] indirectly accounts for the influence of prior knowledge, it is still possible that the rate of learning (as measured by normalized gain) is influenced by prior knowledge. In this study, we directly account for the influence of prior knowledge by including early-semester concept-learning scores as covariates and using measures of conceptual knowledge (as opposed to gains) as our dependent measure. Finally, we further tested the robustness of the attitude-learning link by using knowledge measures that differ from those used in previous research. Instead of the FMCE as a concept inventory, the current study used the Force Concept Inventory (FCI) [14] (for mechanics courses) and Brief Electricity and Magnetism Assessment (BEMA) [15] (for electricity and magnetism courses). The FCI was used instead of the FMCE both because it allowed us to extend previous research [2] with a new instrument, and because the FCI covers a wider range of concepts than the FMCE, making it more representative of the material covered in the mechanics courses that were part of this study. Exam averages also were included as an additional measure of physics-knowledge outcomes.

B. Concept inventories versus exam averages as measures of conceptual knowledge

In the current study, we used two types of physics knowledge measures—concept inventories and exam averages, which have complementary sets of advantages and disadvantages. Concept inventories are validated, general-use instruments, which means that the same instrument can be used across different courses and contexts, and that normative data exist for comparison. However, because they are general-use instruments, the items may not align exactly with what is taught or the way things are taught by a particular instructor in a particular course. Additionally, these are research instruments with no practical impact on grades and so results based only on concept inventory scores may not hold high value for students and instructors. In contrast, exams are course and instructor specific, so the items should align well with the material as taught in the course. Further, results based on exam averages should have a high degree of practical meaning for instructors and students. The weakness of exams as a measure of knowledge is that they are written by individual instructors and are not validated. Accordingly, direct comparisons across courses are not meaningful and results can be influenced by the way an instructor writes exams. However, if concept inventories and exam averages yield similar results, then both the validity and practical meaning of our findings will be supported by these complementary measures.

In addition, concept inventories and exams may measure different types of knowledge, both of which are important for success in physics. The FCI and BEMA consist of multiple-choice questions designed to assess the correctness and coherence of students' conceptual understanding. Distractor items relate to common conceptual misunderstandings, so correctly responding requires sound conceptual understanding, and no calculations are required. In contrast, exams in these courses heavily emphasize quantitative problem solving. All problems involve computation, although the exams are designed such that conceptual understanding also is necessary to set up appropriate problem-solving strategies and correctly apply equations. Thus, including both exam averages and concept inventories allows examination of how attitudes relate to both conceptual understanding and quantitative problem-solving skill. Capturing both of these learning outcomes is important because previous theoretical work (modeling) and empirical findings suggest that for introductory physics students, acquisition of conceptual understanding and of quantitative problem-solving skill is somewhat independent [16,17].

C. Study Overview

In the present study, we investigate whether students' incoming attitudes predict students' learning and course performance across four courses—active physics of the fall semester (AP I), traditional-lecture version of the fall semester (TP I), active physics of the spring semester (AP II), and traditional-lecture version of the spring semester (TP II). We first examine the attitude-learning association using the overall CLASS index. Subsequently, we attempt to pinpoint more specific aspects of attitudes that are tied to learning, first using the original 8 subscale structure and then using an alternative factor structure that we developed, which has been reported previously [7] and is very similar to another recently reported CLASS factor structure [18].

II. RESEARCH METHODS

Data were collected during six consecutive semesters (Fall 2009, Spring 2010, Fall 2010, Spring 2011, Fall 2011, and Spring 2012) of the introductory physics courses at Washington University in Saint Louis. Each fall course (AP I and TP I) primarily covered mechanics, and each spring course (AP II and TP II) covered electricity and magnetism, thermodynamics, and quantum mechanics.

A. Study Instruments

1. Knowledge assessments

Force Concept Inventory (FCI).—The FCI [14] is a measure assessing conceptual knowledge for topics covered in the fall semester (i.e., mechanics). The FCI consists of 30 multiple-choice items, and students completed it near the beginning and end of the fall semester. The FCI score is the proportion of items a student answered correctly.

Brief Electricity and Magnetism Assessment (BEMA).—The BEMA [15] assesses conceptual knowledge for some of the topics covered in the spring semester (i.e., electricity and magnetism). The BEMA includes 30 multiple-choice items, and students completed it at the beginning and end of the spring semester. The BEMA score is the proportion of items a student answered correctly.

Exams.—Each course included three noncumulative exams. The exam score is the proportion correct, averaged across these three exams. For a given course within a given semester, all sections had common exams, but exam problems changed every semester. Exams differed across active and traditional-lecture courses. All of the exam problems for all of the courses involved quantitative problem solving. The problems in traditional-lecture physics tended to be relatively context-free, "classic" physics problems, whereas the problems in the Active courses tended to be framed in terms of real-world situations or applications (see Figs. 1–4 for a representative sample problem from each course). No formal validity or reliability assessment was conducted on these exams.

2. Colorado Learning Attitudes about Science Survey

Of the 42 items, 36 are scored in terms of whether the participant responds in line with the expert consensus. Of the remaining 6 items, 5 lack expert consensus and one item (item 31) is used to discard students who are not reading the questions ("We use this statement to discard the survey of people who are not reading the questions. Please select agree option 4 for this question to preserve your answers."). The 36 items with expert consensus were used in our analysis. Students responded on a 5-point scale, but in line with previous research [2,7,8] we coded responses on these items dichotomously as either in agreement with expert consensus ("positive") or in disagreement ("not positive"). If experts agreed with an item, student responses of 4 or 5 were coded as positive and all other responses as not positive; if experts disagreed, student responses of 1 or 2 were coded as positive and all other responses as not positive. The 36 items were used to compute an overall CLASS score, and scores for various subscales, described below, were also computed. For the overall score and each subscale score, the score represents the *percent positive*—the percentage of items within the scale that were coded as positive.

Traditional subscales.—In line with previous literature, CLASS items were assigned to 8 overlapping subscales—Real-World Connection, Personal Interest, Sense-Making Effort, Conceptual Understanding, Applied Conceptual Understanding, Problem-Solving General, Problem-Solving Confidence, and Problem-Solving Sophistication. The score for each subscale represents the percent positive for items included in the scale.

Newly developed factors.—Through a process that was both theoretically driven and statistically driven (described in a previous paper [7]), we extracted two factors from the CLASS: Learning Approach and Solving Approach. Appendix A includes all of the details of the factor development process and Appendix B shows the items associated with each factor. One item ("I do not expect physics equations to help my understanding of the ideas;

1) (20 points) On December 24, 1968, after a 69 hour voyage, an engine burn inserted the Apollo 8 spacecraft into lunar orbit. After two (elliptical) orbits, another engine burn placed Apollo 8 into a very nearly circular ($\varepsilon = 0.001$) orbit at a distance $h = 111.4$ km above the Moon's surface ($R_{Moon} =$ 1738.1 km). For 16 hours and 10 minutes Apollo 8 stayed in this circular orbit, and executed 8 total orbits of the Moon, after which a final engine burn put the spacecraft on a trajectory that brought it back to Earth. The Apollo 8 craft was the first of the Apollo series to successfully orbit the Moon. Its pilot James A. Lovell, Frank Borman and William Anders were the first three humans ever to see in person the far side of the Moon. The Apollo 8 command module is now on display at the Chicago Museum of Science and Industry.

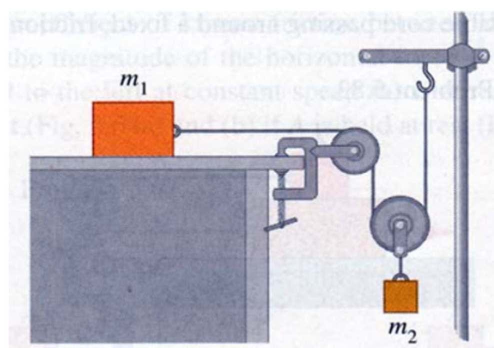
a) (10 points) From the data given above, determine the mass M_{Moon} (in kg) [6 points] and its surface gravity g_{Moon} (in m/s^2) [4 points].

b) (10 points) The Apollo 8 spacecraft was not equipped with a real lunar module (LM) capable of landing on the Moon. However, it did carry a Lunar Module Test Article (LMTA) of mass $M_{LMTA} = 9027$ kg to simulate the ballast of a real LM that would be used to put men on the Moon's surface in subsequent missions. How much *total energy* E_{tot} (in units of J) would it take to put this LM into the lunar orbit given in part (a), starting from rest on the surface of the Moon? [Ignore the KE of the LM caused by the rotation of the Moon itself. This calculation gives an estimate of the energy (thus fuel) requirements to get a LM off the Moon's surface and into an orbit where it could dock with the waiting command module. The command module remains in orbit around the Moon while two of the three astronauts are on the surface. For the present problem, ignore the mass of the astronauts and other complicating factors. To get an idea of whether your answer is plausible, it may be helpful to note that the energy in 1 gallon of gasoline (~3.8 kg) is 1.25×10^8 J. But note that you do <u>not</u> have to evaluate plausibility here.].

FIG. 1. A sample exam problem from AP I.

6. [17 Points] In the drawing to the right two masses are connected by massless and frictionless pulleys and string. The coefficients of friction are $\mu_s = 0.35$, $\mu_k = 0.25$ between the mass m_1 and the table, and the masses are $m_1 = 8.0$ kg and $m_2 =$ 8.0 kg.

a) [3 Points] Carefully define the 2 "systems" you need to solve this problem and make a clear <u>Free-Body diagram</u> for each "system" clearly showing all of the forces.



b) [14 Points] Find the acceleration of each mass and the tension in the string. [Note: Because of the pulley, if m_1 moves a distance L, m_2 moves half that amount, a distance L/2]

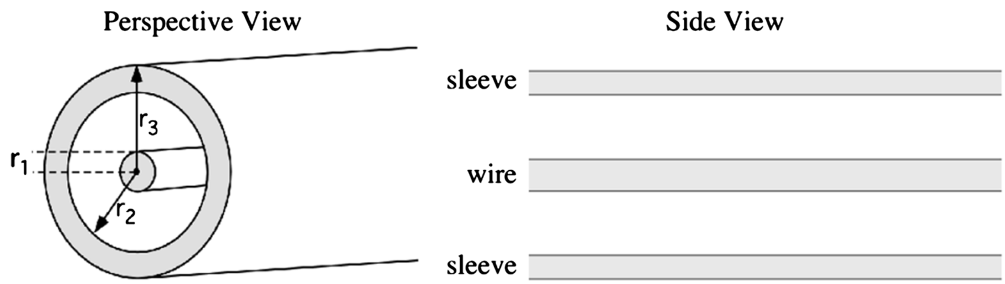
FIG. 2. A sample exam problem from TP I.

they are just for doing calculations") was included in both factors in our previous manuscript [7] and has now been removed from both factors. This item loaded equally on both factors; therefore, it was removed to enhance the clarity of our factor structure. It is worth noting that although we performed our factor analysis independently, our factor structure is very similar to a structure recently reported by another group of researchers [18]. Their previously reported Problem Solving or Learning factor is a subset of our Learning Approach factor, and the Effort and Sense Making factor is a subset of our Solving Approach factor [18].

B. Courses

The active physics courses (AP I and AP II) differed from the traditional-lecture courses in many ways and were designed to encourage deeper conceptual-thinking skills. Each AP class typically included 2 to 3 interactive engagement (IE) activities, which were completed in groups of 2 to 3 students. A number of IE activities centered around demonstrations, in which students were given several minutes to predict the results of the demonstration and discuss with their partners. After having the students discuss among themselves, the instructor would perform

1) (15 points) The figure below shows a portion of a long copper *wire* of length L enclosed by a cylindrical copper *sleeve*. [This is the basic configuration of the beloved coaxial cable that brings cable TV, internet and other delights into millions of homes each day]. The radius of the wire is r_1, and the radii of the inner and outer surfaces of the sleeve are r_2 and r_3, respectively, where $r_3 << L$. The space between r_1 and r_2 is filled with a plastic dielectric (electrical insulator). Now suppose that, initially, the wire and copper sleeve are electrically neutral, but then a charge Q is put on the *wire* (only), giving the wire a charge per unit length of $\lambda = Q/L$.



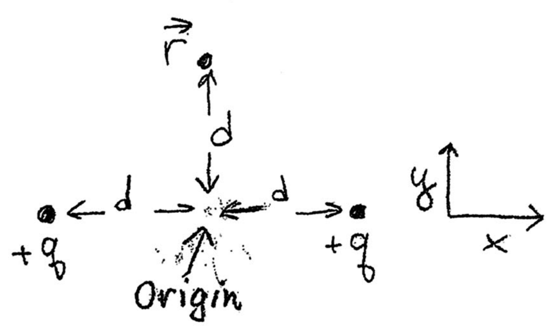
a) [5 points] Using Gauss's law, find the *magnitude* of the electrical field $\vec{E}$ at a distance $r_1 < r < r_2$ in terms of λ and r. [3 points] Assume that the field is measured very far from either end of the wire/sleeve, and ignore any influence of the dielectric on the electric field in the vicinity of the wire. In the "Perspective View" figure above, sketch with dashed lines the (imaginary) closed Gaussian surface used in your calculation [1 point], and in the "Side View" draw some vectors indicating the directions of the electrical field $\vec{E}$ at the surface of the wire. [1 point]

b) [6 points] If the total charge on the surface of the *wire* is $+Q_1$, find the charge Q_2 on the inner surface of the copper sleeve at radius r_2 [2 points]. Then, find the charge Q_3 on the outer surface of the copper sleeve at radius r_3 [2 points]. For both calculations, give your reasoning. [2 points]

c) [4 points] Suppose that we run a conventional current I in the wire and simultaneously run the same current I, but in the *opposite* direction, in the sleeve. Use Ampere's law to find the magnitude of the resulting magnetic field strength B (in terms of μ_o, I and r) at a distance $r_1 < r < r_2$ [2 points], and at a distance $r > r_3$. [2 points]

FIG. 3. A sample exam problem from AP II.

1. **(14 points) Consider a molecule consisting of two equal positive charges q at the positions -di and di, along the x-axis, indicated below in the figure below.**



a) **[4 points] Calculate the interaction force between the charges, in terms of q, d, and fundamental constants.**

b) **[5 points] Calculate the electric field vector at the point r=dj in terms of q, d, and fundamental constants.**

c) **[5 points] A negative charge of magnitude -q and mass m is placed at the point r with zero initial velocity. How fast is it going when it reaches the origin (halfway between the two positive charges? Give your answer in terms of q, d, m, and fundamental constants.**

FIG. 4. A sample exam problem from TP II.

the demonstration and then facilitate discussion with the entire class about why the given result occurred. Other IE activities were two-minute problems, which are multiple-choice or true or false questions that target difficult concepts from the chapter. Students would discuss their answer and reasoning before sharing with the class as a whole by voting. The instructor would then present the correct answer and facilitate further discussion. Within a class period, activities were supplemented with one or more mini lectures, which were limited to 10–15 min each. In contrast, the majority of time in traditional-lecture physics (TP I and TP II) involved students listening to lecture or following along as the instructor solved problems on the board.

Another major difference between the curricula was that active physics had daily homework covering the required reading but traditional lecture did not. Finally, as described above, the exams differed across curricula, with active physics exam problems typically framed in the context of a real-world situation or application and traditional-lecture physics exams containing more "classic" context-free problems (see Figs. 1–4). More details about these curricula can be found in previously published papers [7,17].

C. Procedure

At the beginning of each semester, introductory physics students were asked to provide informed consent to participate in the study procedures for that semester. Those who provided informed consent and then completed all study instruments during the semester were allowed to replace one nonzero lab grade with a perfect score. All instruments were completed online. During the fall semester, the FCI and CLASS were both completed during weeks 2 and 3 of the semester, and both were completed again 2 to 3 weeks before the end of the fall semester. This CLASS score was used as a predictor for the spring semester because the CLASS was not administered at the beginning of the Spring 2010 semester. The use of late-fall versus early-spring CLASS scores will be addressed in the results section. Students completed the BEMA during the first 2 to 3 weeks of the spring semester and again 2 to 3 weeks before the end of the spring semester. In addition, students in each course took three noncumulative exams throughout the semester. See Fig. 5 for a summary timeline of study procedures.

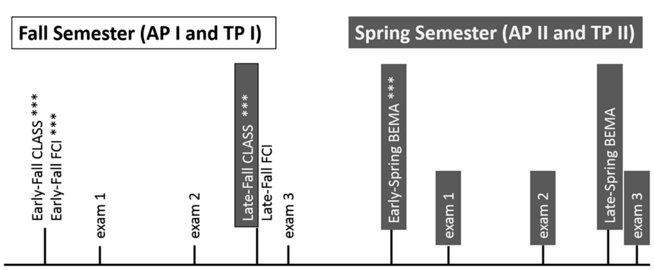


FIG. 5. Summary timeline of study procedures. Measures with dark font and no background are part of the analysis of fall semester courses (AP I and TP I). Measures with white font and dark background are part of analysis of spring semester courses (AP II an TP II). Measures followed by *** are predictor variables and other measures are dependent variables.

D. Samples

The criteria for inclusion in the sample were a completed CLASS (early fall for the fall samples; late fall for the spring samples), a completed early-semester knowledge measure (FCI for the fall samples; BEMA for the spring samples), and valid scores on a knowledge measure used as the dependent measure in the particular analysis (late fall FCI, late-spring BEMA, or all three exams from the given semester). The relationship between CLASS and student performance was analyzed separately for four different (but overlapping) samples: fall active physics [$N = 725$ (59% of course) for both FCI and exam analyses], fall traditional lecture [$N = 341$ (50%) for FCI analyses, 339 (50%) for exam analyses], spring active physics [$N = 526$ (47%) for BEMA analyses, 525 (47%) for exam analyses], and spring traditional lecture [$N = 195$ (34%) for BEMA analyses, 194 (34%) for exam analyses].

E. Analytic strategy

1. Overview

First, to confirm that attitudes and prior knowledge were, in fact, associated, we computed zero-order correlations between early-semester concept inventory scores and CLASS scores (overall CLASS and all categories) for all four of our courses. Next, we conducted regression analyses to address our central interest—the degree to which students' attitudes and beliefs toward physics predicted course performance and late-semester knowledge, particularly after controlling for early-semester knowledge. For these analyses, we used stepwise multiple regression with a measure of conceptual knowledge (either late-semester concept inventory or exam performance) as the dependent measure. Exam performance was operationalized as the students' proportion correct on exams. Each course had three noncumulative exams, and proportion correct was averaged across these three exams. The primary predictors were attitudes (overall CLASS score) and early-semester concept-inventory scores, also treated as proportions. For fall-semester analyses, attitudes were assessed with early-fall CLASS scores and prior knowledge was assessed with the early-fall FCI; for spring semester analyses, attitudes were assessed with late-fall CLASS scores and prior knowledge was assessed with the early-spring BEMA. For our primary analysis, separate regression models (a total of eight in all) were implemented within each of the four courses and for each measure of conceptual knowledge: late-semester concept inventory scores and exam performance.

Additionally, to try to identify CLASS categories that were associated most highly with learning, we conducted the same set of 8 regression analyses two more times, with CLASS categories entered as predictors instead of the overall CLASS score. In one set, the 8 traditional CLASS categories were entered as predictors, and in another set, the two factors we extracted, Learning Approach and Solving Approach, were entered as predictors.

2. *Stepwise regression procedure*

For each of our regression analyses, we utilized a two-step stepwise regression procedure in which our prior knowledge measure (early-fall FCI or early-spring BEMA) was entered in step 1, and the CLASS (the overall score for our primary analyses or all categories in our secondary analyses) was entered in step 2. Step 1 provides the model R^2 when only prior knowledge is included in the model, and step 2 provides the model R^2 when prior knowledge and attitude measures were included. Crucially, step 2 also provides the change in R^2 resulting from the addition of the attitude variable(s) and the test of whether that change is statistically significant. This test of change in R^2 directly addresses our primary research question—whether attitudes predict learning after factoring out the influence of prior knowledge. A significant change in R^2 at step 2 indicates that attitudes predict course performance above and beyond the influence of prior knowledge. The regression coefficients for CLASS variables at step 2 represent the strength of the association between each predictor and the dependent measure, after factoring out the influence of all other predictors. Specifically, the coefficient values represent the expected change in the dependent measure for every 1-unit difference in the predictor. For our supplementary analyses including multiple CLASS categories as predictors, these coefficients (and associated significance values) indicate which aspects of the CLASS most strongly relate to learning in a particular course.

III. RESULTS

A. Correlating CLASS and prior knowledge

Table I shows, for each course, the Pearson correlation (r) between prior knowledge (early-fall FCI or early-spring BEMA) and CLASS (overall scale and all subscales). Pearson's correlation r is a measure of association between two continuous variables, and convention in the social sciences [19] is that a value of 0.1 indicates a small association, 0.3 indicates a medium association, and 0.5 indicates a large association. For interpretation purposes, note also that the square of r is equal to the variation in one variable accounted for by the variation in the other variable. The correlations with overall CLASS range from 0.23 to 0.39, and the correlations between prior knowledge and all of the subscales range from 0.08 to 0.40, with 17 considered medium associations ($0.3 \leq r < 0.5$) and 26 considered small associations ($0.1 \leq r < 0.3$). All correlations reached significance ($p < 0.05$), except for the correlation between BEMA and Conceptual Understanding in TP II, which was below 0.1 and did not reach statistical significance. Although the associations are not large, these results show that attitudes are associated with prior knowledge and confirm the importance of accounting for prior knowledge when analyzing the relationship between attitudes and learning.

B. Primary analyses using overall CLASS as predictor

Parallel analyses were conducted for all four courses in the study– Active Physics I (AP I), Active Physics II (AP II), Traditional Physics I (TP I), and Traditional Physics II (TP II). For each course, we conducted stepwise

TABLE I. Correlations between prior knowledge and CLASS measures. Superscripts indicate p values. The notations on the table are as follows: Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); Learning Approach (LA); Solving Approach (SA); personal interest (PI); real-world connection (RC); problem-solving general (PSG); problem-solving confidence (PSC); problem-solving sophistication (PSS); sense-making effort (SME); conceptual understanding (CU); applied conceptual understanding (ACU). For AP I and TP I, the measure of prior knowledge is the FCI. For AP II and TP II, the measure of prior knowledge is the BEMA.

	AP I ($N = 725$)	AP II ($N = 525$)	TP I ($N = 339$)	TP II ($N = 194$)
CLASS	0.39^{c}	0.24^{c}	0.39^{c}	0.23^{b}
LA	0.37^{c}	0.28^{c}	0.39^{c}	0.16^{a}
SA	0.28^{c}	0.19^{c}	0.30^{c}	0.23^{b}
PI	0.34^{c}	0.18^{c}	0.35^{c}	0.21^{b}
RC	0.23^{c}	0.14^{b}	0.25^{c}	0.18^{a}
PSG	0.38^{c}	0.25^{c}	0.40^{c}	0.19^{b}
PSC	0.33^{c}	0.16^{c}	0.32^{c}	0.20^{b}
PSS	0.39^{c}	0.24^{c}	0.40^{c}	0.20^{b}
SME	0.24^{c}	0.14^{b}	0.27^{c}	0.21^{b}
CU	0.31^{c}	0.24^{c}	0.32^{c}	0.08
ACU	0.33^{c}	0.25^{c}	0.34^{c}	0.16^{a}

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

TABLE II. Regression-model summaries predicting late-semester concept inventory (FCI and BEMA) scores from prior knowledge measures and overall CLASS scores. The notations on the table are as follows: Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observations (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II). Prior knowledge alone was entered in step 1 of each model, and then overall CLASS was added in step 2.

	AP I ($N = 725$)		TP I ($N = 341$)		AP II ($N = 526$)		TP II ($N = 195$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.607[c]	0.612[c]	0.566[c]	0.570[c]	0.478[c]	0.506[c]	0.316[c]	0.349[c]
ΔR^2		0.005[b]		0.004		0.028[c]		0.033[b]
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.73[c]	0.70[c]	0.73[c]	0.72[c]	0.88[c]	0.79[c]	0.78[c]	0.72[c]
CLASS		0.09[b]		0.08		0.20[c]		0.19[b]

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

regression analyses in which prior knowledge (early-semester FCI for AP I and TP I; early-semester BEMA for AP II and TP II) was entered as a predictor in step 1 and overall CLASS score was added as an additional predictor in step 2.

1. Predicting late-semester concept inventory scores

Table II shows the results for each course when concept inventories (late-semester FCI for AP I and TP I; late-semester BEMA for AP II and TP II) were used as the dependent measure. For AP I, AP II, and TP II, the addition of the CLASS score significantly improved the model, as indicated by a significant ΔR^2. These results show that attitudes predict late-semester conceptual knowledge above and beyond what is explained by early-semester conceptual knowledge. An examination of the regression coefficients shows that prior knowledge is a much stronger predictor than the CLASS, but that the CLASS has a meaningful association with late-semester knowledge, particularly in AP II and TP II. For example, the 0.20 coefficient in AP II means that a 1-unit difference in CLASS (from 0 [complete disagreement with experts] to 1 [complete agreement with experts]) increases the predicted BEMA score by 0.2. In other words a student in complete agreement with experts would be expected, on average, to have a BEMA score that is 0.20 higher than a student in compete disagreement with experts.

2. Predicting exam performance

Table III shows the results for each course when exam performance (proportion correct) was used as the dependent measure. Again, for AP I, AP II, and TP II, the addition of

TABLE III. Regression-model summaries predicting exam scores from prior knowledge measures and overall CLASS scores. The notations on the table are as follows: Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observations (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II). Prior knowledge alone was entered in step 1 of each model, and then overall CLASS was added in step 2.

	AP I ($N = 725$)		TP I ($N = 339$)		AP II ($N = 525$)		TP II ($N = 194$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.191[c]	0.197[c]	0.235[c]	0.238[c]	0.133[c]	0.145[c]	0.077[c]	0.099[c]
ΔR^2		0.006[a]		0.003		0.012[b]		0.022[a]
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.17[c]	0.16[c]	0.33[c]	0.32[c]	0.19[c]	0.16[c]	0.29[c]	0.25[b]
CLASS		0.04[a]		0.05		0.05[b]		0.11[a]

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

the CLASS score significantly improved the model, as indicated by a significant ΔR^2. These results show that attitudes predict exam performance above and beyond what is explained by early-semester conceptual knowledge. Examination of coefficients shows that both prior knowledge and CLASS scores were less predictive of exam performance than they were of late-semester concept inventory scores. However, the overall pattern of results was the same for both measures of learning (concept inventories and exam averages)—prior knowledge significantly predicts learning in all courses, with the CLASS explaining additional unique variance for AP I, AP II, and TP II.

3. A note about the use of late-fall CLASS scores

The CLASS predictor for AP I and TP I is the early-semester score from the beginning of the semester, whereas the CLASS predictor for AP II and TP II is the late-semester score from the previous semester. It is worth discussing whether the use of late-fall scores could have influenced the pattern of results for AP II and TP II. Whether late-fall and early-spring attitudes differ (and there are reasons such as stress, fatigue, etc. to think they might) is a separate question from whether late-fall attitudes and early-spring attitudes differentially predict spring conceptual learning. The most likely effect would be that, relative to early-spring scores, late-fall scores would be weaker predictors of spring conceptual learning because of the larger time gap between when the attitude is measured and when the conceptual learning is measured. In this case, the strength of the CLASS-conceptual learning association would be artificially reduced in AP II and TP II because of the use of late-fall CLASS scores. However, this does not appear to be the case, as the CLASS-conceptual learning associations were higher for AP II and TP II than for AP I and TP I, despite this larger time gap. Although it is still possible that the use of late-fall scores produced an underestimate of the attitude-conceptual learning associations in the spring, it is unlikely that the use of late-fall CLASS scores substantively changed the pattern of results.

C. Analyses including CLASS categories as predictors

We next attempted to identify particular aspects of the CLASS that might best predict learning. To that end, we conducted additional regression analyses in which CLASS categories, rather than the overall CLASS score, were entered as predictors. First, we used the original 8 categories reported by the authors of the CLASS [8]. Then, in a separate set of analyses we used the two factors that we had developed as predictors [7].

1. Analyses with 8 CLASS categories

In these analyses, prior knowledge was entered as the predictor in step 1 and then all 8 categories were added as additional predictors in step 2. In Appendix C, Table IV shows the results for each course when concept inventories were used as the dependent measure. Adding these predictors significantly improved the model for AP II and TP II but not AP I or TP I. None of the individual predictor coefficients was significant in AP II or TP II. In Appendix C, Table V shows the same analyses with exam performance as the DV. In these analyses, adding the CLASS categories in step 2 significantly improved the model for AP I and AP II. Inspection of coefficients shows that Applied Conceptual Understanding was the only significant predictor for AP I and none of the individual CLASS categories were significant predictors in AP II. Although the CLASS categories did not significantly improve the model in TP II, this was likely due to the relatively small sample size, as the actual change in R^2 was higher in this course than in any of the other courses. And within TP II, Problem-Solving Sophistication is the only CLASS category that is a significant predictor, and its association with exam performance is relatively large ($B = 0.17$).

2. Analyses with 2 newly developed factors

In these analyses, prior knowledge was entered as the predictor in step 1 and then both factors (Learning Approach and Solving Approach) were entered as additional predictors in step 2. In Appendix C, Table VI shows the results for each course with concept inventories as the DV. Similar to the analyses including overall CLASS, the addition of the CLASS variables significantly improved the model for AP I, AP II, and TP II. Out of these models, the only significant coefficient was the coefficient for Learning Approach within AP II (0.16). In Appendix C, Table VII shows the same analyses with exam performance as the DV. Again, the addition of these two CLASS factors significantly improved the model in AP I, AP II, and TP II. Learning Approach was a significant predictor of exam performance in AP I and AP II.

3. Summary

Overall, the results show that, after accounting for prior knowledge, CLASS scores predict late-semester concept inventories and exam performance in three of the four courses we examined. Entering subscales of the CLASS (either 8 categories or 2 categories) suggested that, generally, particular subfactors within the CLASS do not alone reliably associate with learning. With the two-factor structure, Solving Approach was not a significant predictor in any analyses, and Learning Approach was a significant predictor in only 2 of the 4 courses when exam performance was used as the dependent measure and in only 1 of 4 courses when FCI or BEMA was used as the dependent measure. With the eight-factor structure, even fewer significant predictors emerged and no predictors were consistent across courses. Specially, regardless of whether exam performance or concept inventories were used as the DV, none of the eight categories was a significant predictor in more than one course. Even Applied Conceptual Understanding, which was a relatively strong predictor of exam performance in TP II ($B = 0.17$), was not a significant predictor of exam performance in any of the other courses and was not a significant predictor of late-semester BEMA in TP II.

IV. DICUSSION

The notion that student attitudes influence learning is an assumption in the physics education literature that has been rarely tested [12]. The current study directly examined this attitude-learning association after controlling for prior knowledge. Our results suggest that, independent of incoming physics knowledge, students who come into physics with more expertlike attitudes toward physics may learn more and perform better than students with more novicelike attitudes, at least in certain courses. However, despite the CLASS being a significant predictor in three of the four courses we analyzed, it is noteworthy that the size and significance of CLASS's influence varied considerably across course and dependent measure (concept inventory versus exam average). In the electricity and magnetism courses (AP II and TP II), with late-semester BEMA as the dependent measure, the contribution of CLASS to performance is relatively large ($B = 0.20$ and $B = 0.19$ for AP II and TP II, respectively), but in all other analyses the relationship is smaller or even nonsignificant.

The overall pattern of results suggests that while attitudes (as measured by the CLASS) may relate to learning, the relationship is relatively modest and inconsistent across courses and measures. These results may shed light on important variables that influence the strength of the association between attitudes and learning and suggest that, in certain situations, the relationship between attitudes and learning may be rather weak. However, another possibility is that attitudes generally have a strong impact on learning, but the use of CLASS and similar instruments to measure attitudes can produce underestimates of this relationship. The following sections elaborate on these possibilities.

A. Systematic effects of other variables

A recent meta-analysis [12] concludes that there exists "a small significant correlation between students' incoming beliefs about learning physics and their conceptual gains", but that understanding this relationship requires consideration of other variables, such as course characteristics. Potentially, our data reveal important variables that influence the strength of the attitude-learning relationship. First, the assessment measure appears to be important, as attitudes contributed more to concept inventory performance than they did to exam scores. Although exams were designed to necessitate conceptual understanding, they were centered on quantitative problem solving. Previous theoretical and empirical work suggests that for introductory physics students, conceptual physics knowledge and quantitative problem-solving knowledge (in physics) can be somewhat independent [16,17]. Perhaps expertlike attitudes (revealed by the CLASS) are mostly related to conceptual learning outcomes, with relatively more novicelike approaches (e.g., approaching quantitative problems in terms of focusing on equations and computations) proving sufficient to support the quantitative problem solving required by the exams. This interpretation bears further exploration, however, as it remains possible that the concept inventories and the exams differ in terms of psychometric properties (e.g., reliability), which could influence the observed strength of the attitude-learning relationship.

Second, course content may be an important variable, as attitudes relate more strongly to learning in the electricity and magnetism courses. Perhaps attitudes play a larger role as material becomes more challenging or less grounded in personal experiences. Researchers have theorized that students struggle to learn electricity and magnetism because, relative to mechanics, the concepts are abstract and outside of everyday experience [20]. Mechanics problems often involve familiar objects such as balls and ramps, and concepts like velocity and acceleration relate strongly to everyday experiences. In contrast, electricity and magnetism requires reasoning about microscopic entities or abstractions with which students have no tangible experiences. Developing coherent conceptual models for this material may require substantial effort, and thus attitudes may be an important factor in whether students put in the effort to learn these concepts, in the approach students take to learn the concepts (e.g., see Ref. [21]), or both.

One variable that does not appear to exert much influence on the attitude-leaning association in our data is course format (active versus traditional). At the outset, we speculated that attitudes may be more influential in interactive engagement (IE) courses because these courses are designed to place more emphasis on certain expertlike aspects of learning physics such as developing deeper understanding of core concepts and activating those concepts in the service of problem solving. Thus, the IE format might align particularly well for students with expert-like attitudes. Alternatively, in line with the current patterns, it seems that a student's initial attitude is not any more important for determining learning in IE formats relative to traditional formats. Of course, along with our other findings regarding the attitude-learning link (as discussed above), it could be the current patterns are limited to the particular features of our courses and measures. That is, a more extreme IE format (completely reversed classroom) might show a different pattern, as might different measures. At the least, however, the current patterns identify some variables that may merit further investigation in determining an attitude-learning link.

B. CLASS and other surveys may not measure the right kind of attitudes and beliefs

If attitudes influence learning and performance, the likely mechanism would be that students with more expertlike attitudes use more successful approaches to learning physics. Students with more novice attitudes perceive physics as a set of facts and equations and equate learning physics with knowing these facts and equations. Thus their attempts to learn physics might be expected to center on memorization and practicing the use of certain equations and problem-solving procedures. In contrast, students with more expert-like attitudes perceive physics as interconnected and coherent, so they might be expected to learn physics by connecting new knowledge to previous knowledge and attempting to build coherent conceptual frameworks. Connected to this idea is the recent finding from chemistry that abstraction learners, who tend to relate and find patterns among examples, outperform exemplar learners, who tend to focus on learning the features of individual examples [21].

However, recent work suggests that student attitudes, as measured by surveys such as the CLASS, may not strongly align with the way students actually approach learning science in practice. With regard to beliefs about the nature of science (NOS), Sandoval [22] distinguishes between *formal* epistemology, the beliefs students express about knowledge and knowledge production in professional science, and *practical* epistemology, the beliefs students hold about their own knowledge production in science courses. Recent research demonstrates that formal epistemology and practical epistemology do not always align. For example, previous research has demonstrated a dissociation between students' beliefs about the NOS as measured by self-report survey instruments (i.e., declarative NOS understanding or formal epistemology) and how students actually approach knowledge construction during collaborative work (i.e., procedural NOS understanding or practical epistemology) [23]. Specifically, while procedural NOS understanding became more expertlike across the semester of an inquiry-based course, declarative NOS understanding changed very little.

With regard to the CLASS, Rowe and Phillips [24] coded homework assignments for various aspects of expertlike thinking and sense making and demonstrated that these codes did not significantly correlate with expertlike beliefs as measured with the CLASS. The small sample size ($N = 26$) may have accounted for the lack of significance, especially in the correlation between these codes and percent of unfavorable (novicelike) responses on the CLASS ($r = 0.33$, $p = 0.10$), but these preliminary results nonetheless suggest that the CLASS may be a measure of formal epistemology with little direct bearing on how students approach their own physics learning.

Potentially, then, attitudes do strongly predict learning, but especially if those attitudes relate to practical epistemology and directly influence how students approach their own physics learning. In other words, our results do not necessarily show that attitudes weakly relate to learning, but rather they may reflect that the CLASS is not a good measure of the type of attitudes that more directly influence learning. Future research should work to develop and use assessments that relate more closely to students' practical epistemology, such as the homework coding developed by Rowe and Phillps [24] to (a) continue to investigate the relationship between formal and practical epistemologies and (b) determine whether measures of practical epistemology are stronger predictors of learning than typically-used attitude measures such as the CLASS.

C. Study limitations

Although this study included a variety of courses, and data were collected across multiple years and sections taught by multiple instructors, the generalizability of the results may be limited by several factors. First, all data were collected from a single, highly selective university, so the patterns of data may be driven, in part, by the way these courses are taught or by characteristics of the student body. Future research should investigate whether the patterns in our data replicate at other universities with different student characteristics. Further, because the study instruments were not required parts of the course, the response rate for the study was relatively low, ranging from 34% to 59% across the four courses. In addition, responding was likely not random, with certain types of students (e.g., highly conscientious students) likely overrepresented in the sample. Because our samples may not have been completely representative of the populations from which they were drawn, we acknowledge the possibility that our results may not generalize to all of the students in these courses. However, we have no *a priori* reason to suspect that the variables that influence one's likelihood of participating in the study would influence the associations between attitudes and learning.

D. Conclusion

The CLASS and similar measures have received a great deal of attention in physics education research, often acting as primary outcome measures in studies investigating the impact of new curricula. Although attitudes are worthwhile outcomes to study in their own right, their importance would undoubtedly be enhanced by confirmation that they directly and strongly relate to student learning. Our data contain some evidence of a unique relationship between student attitudes and learning, but do not demonstrate that this is a strong, consistent relationship. Much research is still needed to better understand the link between student attitudes and learning and how this relationship plays out across various learning contexts.

ACKNOWLEDGMENTS

The data for this project was collected under HHMI Grants No. 52005911 and No. 52006961.

APPENDIX A: FACTOR DEVELOPMENT PROCESS

Our factor development (described in a previous paper [7]) was both theoretically driven and statistically driven. First, we identified two dimensions of students' approaches toward physics that we believed would be particularly influential on performance: Learning Approach and Solving Approach. We conceptualized learning approach as capturing whether students focused more on acquiring rote knowledge or building deeper conceptual understanding [21] and we viewed the Solving Approach dimension as capturing whether students tended to solve problems by focusing on formulas or step-by-step algorithms or by formulating strategies from a conceptual basis. Independently, all members of the research team coded each of the 36 questions from the overall CLASS for whether it potentially belonged to the Learning Approach dimension and/or the Solving Approach dimension (items could be coded as being part of both dimensions). In this initial pass, we included a question in the dimension if any members of the research team coded it this way, leaving 24 questions in the Learning Approach scale and 15 in the Solving Approach scale. Ten questions overlapped across the scales, so the two scales included a combined 29 items.

We next used factor analysis to statistically derive our final factor structure. We performed exploratory factor analyses with varimax rotation. For each analysis, to determine the number of factors, we first eliminated all factors with eigenvalues below 1.0 (known as the Kaiser stopping rule), and then we used the scree test to determine which of the remaining factors should be kept. In this test, the eigenvalue of each factor is plotted with eigenvalues on the y axis and factor number on the x axis. Factor 1 always has the highest eigenvalue with the value decreasing for each subsequent factor. In the scree test, the researchers determine, visually, the point at which the eigenvalues "level off." The factor at which the leveling off begins and all subsequent factors are removed. For each factor analysis, we retained the factors with eigenvalues above 1.0 that passed the scree test.

We initially conducted two separate factor analyses—one including the items from the Learning Approach scale and another including items from the Solving Approach scale in order to see whether each group of items loaded into a single coherent factor. Overlapping items were included in both analyses. The analyses revealed that both groups of items cohered into a single factor, but several of the items had loadings < 0.2 (1 item from the rote or conceptual scale, 2 from the active or passive scale, and 1 common item that loaded < 0.2 on both scales). Removing these four items results in 25 total remaining items—22 rote or conceptual and 12 active or passive, with 9 questions overlapping between the scales. We next conducted a factor analysis with all of 25 remaining items, and as expected, it produced a two-factor solution. We placed each item into the factor on which it loaded the most highly, and one ambiguous item, which loaded equally on both factors, was removed. This resulted in two 12-item factors that are both internally consistent (Cronbach's α was 0.81 for factor 1 and 0.76 for factor 2). Inspection of the questions in these two factors suggested that the two factors corresponded to our two proposed dimensions, and we labeled factor 1 Learning Approach and factor 2 Solving Approach. It is worth noting that although we performed our factor analysis independently, our factor structure is very similar to a structure recently reported by another groups of researchers [18]. The previously reported Problem Solving or Learning factor is a subset of our Learning Approach factor, and the Effort and Sense Making factor is a subset of our Solving Approach factor [18].

APPENDIX B: TWO-FACTOR STRUCTURE (LISTED BY CLASS ITEM NUMBER)

Learning Approach (Rote versus Conceptual)

1. A significant problem in learning physics is being able to memorize all the information I need to know.
5. After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.
6. Knowledge in physics consists of many disconnected topics.
8. When I solve a physics problem, I locate an equation that uses the variables given in the problem and plug in the values.
10. There is usually only one correct approach to solving a physics problem.
12. I cannot learn physics if the teacher does not explain things well in class.
17. Understanding physics basically means being able to recall something you've read or been shown.
20. I do not spend more than five minutes stuck on a physics problem before giving up or seeking help from someone else.
21. If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.
22. If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations.
34. I can usually figure out a way to solve physics problems.

40. If I get stuck on a physics problem, there is no chance I'll figure it out on my own.

Solving Approach (algorithmic versus concept-based)

2. When I am solving a physics problem, I try to decide what would be a reasonable value for the answer.
11. I am not satisfied until I understand why something works the way it does.
15. If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.
23. In doing a physics problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem.
24. In physics, it is important for me to make sense out of formulas before I can use them correctly.
26. In physics, mathematical formulas express meaningful relationships among measurable quantities.
29. To learn physics, I only need to memorize solutions to sample problems.
32. Spending a lot of time understanding where formulas come from is a waste of time.
36. There are times I solve a physics problem more than one way to help my understanding.
37. To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.
39. When I solve a physics problem, I explicitly think about which physics ideas apply to the problem.
42. When studying physics, I relate the important information to what I already know rather than just memorizing it the way it is presented.

APPENDIX C: ANALYSES USING CATEGORY STRUCTURES OF THE CLASS

Tables IV and V present regression analyses with 8 class categories, rather than an overall CLASS score, used as predictors. Analyses were conducted both with late-semester concept inventories as the dependent measure (Table IV) and with exam performance as the dependent measure (Table V). Tables VI and VII present parallel regression analyses with 2 CLASS factors (Learning Approach and Solving Approach) entered as predictors. Table VI shows analyses with

TABLE IV. Regression-model summaries predicting late-semester concept inventory (FCI and BEMA) scores from prior knowledge measures and eight CLASS categories. Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observations (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II); personal interest (PI); real-world connection (RC); problem-solving general (PSG); problem-solving confidence (PSC); problem-solving sophistication (PSS); sense-making effort (SME); conceptual understanding (CU); applied conceptual understanding (ACU). Prior knowledge alone was entered in step 1 of each model, and the eight CLASS variables were added in step 2.

	AP I ($N = 725$)		TP I ($N = 341$)		AP II ($N = 526$)		TP II ($N = 195$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.607[c]	0.614[c]	0.566[c]	0.584[c]	0.478[c]	0.516[c]	0.316[c]	0.385[c]
ΔR^2		0.007		0.018		0.038[c]		0.069[b]
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.73[c]	0.70[c]	0.73[c]	0.72[c]	0.88[c]	0.78[c]	0.78[c]	0.69[c]
PI		−0.01		−0.06		−0.01		−0.03
RC		0.02		0.09[b]		−0.01		0.01
PSG		0.05		0.02		0.12		0.07
PSC		0.03		−0.04		−0.06		−0.01
PSS		0.00		−0.02		−0.01		0.12
SME		−0.01		0.05		0.03		0.04
CU		0.00		−0.09		0.04		−0.01
ACU		0.00		0.13		0.10		0.07

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

TABLE V. Regression-Model Summaries predicting exam scores from prior knowledge measures and eight CLASS categories Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observations (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II); personal interest (PI); real-world connection (RC); problem-solving general (PSG); problem-solving confidence (PSC); problem-solving sophistication (PSS); sense-making effort (SME); conceptual understanding (CU); applied conceptual understanding (ACU). Prior knowledge alone was entered in step 1 of each model, and the eight CLASS variables were added in step 2.

	AP I ($N = 725$)		TP I ($N = 339$)		AP II ($N = 525$)		TP II ($N = 194$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.191[c]	0.220[c]	0.235[c]	0.246[c]	0.133[c]	0.183[c]	0.077[c]	0.144[c]
ΔR^2		0.029[b]		0.011		0.050[c]		0.066
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.17[c]	0.16[c]	0.33[c]	0.32[c]	0.19[c]	0.15[c]	0.29[c]	0.25[b]
PI		−0.02		0.04		<0.01		−0.08
RC		−0.01		−0.01		−0.02		0.02
PSG		−0.01		−0.09		0.04		0.13
PSC		<0.01		0.04		−0.03		−0.11
PSS		−0.01		<0.01		0.04		0.17[a]
SME		0.01		0.04		−0.02		0.05
CU		<0.01		0.07		−0.02		−0.06
ACU		0.07[a]		−0.03		0.06		−0.05

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

TABLE VI. Regression-model summaries predicting late-semester concept inventory (FCI and BEMA) scores from prior knowledge measures and CLASS factors from our two-factor solution. Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observation (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II); Learning Approach (LA); Solving Approach (SA). Prior knowledge alone was entered in step 1 of each model, and the two CLASS factors were added in step 2.

	AP I ($N = 725$)		TP I ($N = 341$)		AP II ($N = 526$)		TP II ($N = 195$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.607[c]	0.612[c]	0.566[c]	0.569[c]	0.478[c]	0.518[c]	0.316[c]	0.352[c]
ΔR^2		0.005[a]		0.003		0.040[c]		0.036[b]
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.73[c]	0.71[c]	0.73[c]	0.72[c]	0.88[c]	0.77[c]	0.78[c]	0.73[c]
LA		0.04		0.04		0.16[c]		0.10
SA		0.05		0.02		0.06		0.07

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

TABLE VII. Regression-model summaries predicting exam scores from prior knowledge measures and CLASS factors from our two-factor solution Fall Active Physics (AP I); Fall Traditional Physics (TP I); Spring Active Physics (AP II); Spring Traditional Physics (TP II); number of observation (N); change in R^2 between step 1 and step 2 (ΔR^2); prior knowledge (PK) (FCI in AP I and TP I, BEMA in AP II and TP II); Learning Approach (LA); Solving Approach (SA). Prior knowledge alone was entered in step 1 of each model, and the two CLASS factors were added in step 2.

	AP I ($N = 725$)		TP I ($N = 339$)		AP II ($N = 525$)		TP II ($N = 194$)	
	Model-level statistics							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
R^2	0.191[c]	0.214[c]	0.235[c]	0.239[c]	0.133[c]	0.170[c]	0.077[c]	0.106[c]
ΔR^2		0.022[c]		0.004		0.036[c]		0.030
	Predictor coefficients (unstandardized)							
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step1	Step 2
PK	0.17[c]	0.15[c]	0.33[c]	0.32[c]	0.19[c]	0.15[c]	0.29[c]	0.26[c]
LA		0.07[c]		0.04		0.09[c]		0.10
SA		−0.01		<0.01		−0.03		0.01

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

late-semester concept inventories as the dependent measure, and Table VII shows analyses with exam performance as the dependent measure.

[1] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).

[2] K. K. Perkins, W. K. Adams, S. J. Pollock, N. D. Finkelstein, and C. E. Wieman, Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey, AIP Conf. Proc. **790**, 61 (2005).

[3] V. K. Otero and K. E. Gray, Attitudinal gains across multiples universities using the Physics and Everyday Thinking curriculum, Phys. Rev. ST Phys. Educ. Res. **4**, 020104 (2008).

[4] E. Brewe, L. Kramer, and G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS, Phys. Rev. ST Phys. Educ. Res. **5**, 013102 (2009).

[5] E. Brewe, A. Traxler, J. de la Garza, and L. Kramer, Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction, Phys. Rev. ST Phys. Educ. Res. **9**, 020116 (2013).

[6] B. A. Lindsey, L. Hsu, H. Sadaghiani, J. W. Taylor, and K. Cummings, Positive attitudinal shifts with the Physics by Inquiry curriculum across multiple implementations, Phys. Rev. ST Phys. Educ. Res. **8**, 010102 (2012).

[7] M. J. Cahill, K. M. Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey, Multiyear, multi-instructor evaluation of a large-scale interactive-engagement curriculum, Phys. Rev. ST Phys. Educ. Res. **10**, 020101 (2014).

[8] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006).

[9] D. Hammer, Epistemological beliefs in introductory physics, Cognit. Instr. **12**, 151 (1994).

[10] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[11] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bohham, M. H. Dancy, and J. S. Risley, The student-centered activities for large enrollement undergraduate programs (SCALE-UP) project, in *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007).

[12] A. Madsen, S. B. McKagan, and E. C. Sayre, How physics instructions impacts students' beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).

[13] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion

Conceptual Evalutation and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66**, 338 (1998).

[14] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992).

[15] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating and electricity and magnetism tool: Brief electricity and magnetism assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).

[16] R. Ploetzner and K. Van Lehn, The acquisition of qualitative physics knowledge during textbook-based physics training, Cognit. Instr. **15**, 169 (1997).

[17] M. A. McDaniel, S. M. Stoen, R. F. Frey, Z. E. Markow, K. M. Hynes, J. Zhao, and M. J. Cahill, Dissociative conceptual and quantitative problem solving outcomes across interactive engagement and traditional format introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 020141 (2016).

[18] K. A. Douglas, M. S. Yale, D. E. Bennett, M. P. Haugan, and L. A. Bryan, Evaluation of Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **10**, 020128 (2014).

[19] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences,* 2nd ed. (Erlbaum, Hillsdale, NJ, 1988).

[20] R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, Am. J. Phys. **74**, 329 (2006).

[21] R. F. Frey, M. J. Cahill, and M. A. McDaniel, Students' concept-building approaches: A novel predictor of success in chemistry courses, J. Chem. Educ. **94**, 1185 (2017).

[22] W. A. Sandoval, Understanding students' practical epistemologies and their influence on learning through inquiry, Sci. Educ. **89**, 634 (2005).

[23] I. Y. Salter and L. J. Atkins, What students say versus what they do regarding scientific inquiry, Sci. Educ. **98**, 1 (2014).

[24] G. Rowe and A. M. Phillips, A comparison of students' written explanations and CLASS responses, in *Proceedings of the Physics Education Research Conference 2016, Sacramento, CA* (AIP, New York, 2016), pp. 280–283.