Making sense of confusion: Relating performance, confidence, and self-efficacy to expressions of confusion in an introductory physics class

Jason E. Dowd,[1,*] Ives Araujo,[2,3] and Eric Mazur[1,2]

[1]*Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA*
[2]*School of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138, USA*
[3]*Physics Institute, Federal University of Rio Grande do Sul, 9500 Bento Gonçalves Avenue, Porto Alegre–RS, 91501-970, Brazil*

(Received 15 May 2014; published 3 March 2015)

Although confusion is generally perceived to be negative, educators dating as far back as Socrates, who asked students to question assumptions and wrestle with ideas, have challenged this notion. Can confusion be productive? How should instructors interpret student expressions of confusion? During two semesters of introductory physics that involved Just-in-Time Teaching (JiTT) and research-based reading materials, we evaluated performance on reading assignments while simultaneously measuring students' self-assessment of their confusion over the preclass reading material ($N = 137$; $N_{\mathrm{fall}} = 106$, $N_{\mathrm{spring}} = 88$). We examined the relationship between confusion and correctness, confidence in reasoning, and (in the spring) precourse self-efficacy. We find that student expressions of confusion before coming to class are negatively related to correctness on preclass content-related questions, confidence in reasoning on those questions, and self-efficacy, but weakly positively related to final grade when controlling for these factors ($\beta = 0.23$, $p = 0.03$).

DOI: 10.1103/PhysRevSTPER.11.010107 PACS numbers: 01.40.Fk, 01.40.Ha

I. INTRODUCTION

When instructors pose the question, "Is anyone confused?" they are asking students to reflect on a topic, consider with which parts they feel comfortable and with which parts they struggle, and then report their thoughts. A breakdown anywhere along this sequence may mislead the instructor, and, even if students accurately report their ideas, instructors may not interpret such feedback correctly [1]. Students' expressions of doubt may indicate discomfort with the material, or they may indicate that students are actually engaged and growing familiar enough with the material that it is conflicting with their prior knowledge and expectations. Conversely, the absence of confusion may indicate comprehension of the material, but it may also indicate that the student is not even aware of conflicts between new ideas and prior knowledge. Student recognition of this conflict may assist, rather than inhibit, the learning process [2–6]. Moreover, positive aspects of confusion, which indicate critical thinking and self-assessment, and negative aspects of confusion, which indicate a lack of knowledge or confidence in one's knowledge, are not mutually exclusive.

In this study, we examine how students' expressions of confusion after reading new material and before coming to class relate to other measures of learning and engagement so that we might better interpret such confusion during instruction. We use the term "expression" to emphasize that the confusion reported by students is not necessarily identical to their thought processes. Perhaps students are concerned about displaying too much or too little confusion about a subject, or motivated to reply efficiently at the expense of accuracy. We limit our analysis to confusion that students express, and we make no assumptions about how students' submitted responses may differ from their unexpressed thoughts.

From a practical point of view, the question is straightforward. However, the notion of confusion actually transects several different theoretical domains. Three related concepts—metacognition (thinking about one's own knowledge and understanding), confidence (belief in one's ability to act in a proper way), and self-efficacy (belief in one's ability to execute required actions)—each relate to different facets of confusion and therefore must combine to form a theoretical framework for the present study. Here we briefly introduce these concepts and suggest additional references for more detailed discussion.

One cannot express confusion without engaging in metacognition, which involves knowledge and cognition about cognitive phenomena [7]. Metacognition is important in learning because students who can accurately distinguish between what they have already learned and what they have

*jedowd@post.harvard.edu

yet to learn will be more strategic and effective in the educational setting. Students' evaluation of their own performance on a task is often used to assess metacognition, despite concerns about the validity of such self-reported measures [8]. Treating the similarity between students' estimates of their knowledge and their performance as a proxy for metacognition, researchers have found that this ability is generally positively related to GPA [9]. In the science classroom, numerous teaching strategies are designed to promote metacognition, as more metacognitive students are more likely to recognize inconsistencies and refine naive ideas [10]. Specifically within physics, researchers observe that adding metacognitive tasks to reading-comprehension exercises results in higher post-test scores when compared to a group of subjects who do not complete the metacognitive tasks [11]. More generally, research involving strong and weak readers, writers, and problem solvers shows skilled, successful individuals also tend to be more successful in metacognitive activities [12]. In some tasks, individuals who perform poorly do not know how poorly they are performing, which researchers attribute to links between content knowledge and the metacognitive skills related to that knowledge [13]. These studies represent only a small sample of a broad and robust field of research in which evidence consistently suggests that enhanced metacognition is positively related to learning outcomes.

The psychologist Albert Bandura introduced the term *self-efficacy* in 1977 to provide a means of explaining the effect of performance-based procedures on psychological change [14]. Simply put, self-efficacy refers to "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" [15]. A review of numerous studies measuring sources of self-efficacy in academic contexts suggests that mastery experiences are most strongly related to students' self-efficacy [16]. In the sciences, researchers have linked self-efficacy to persistence in the discipline and academic success [17–21]. Research efforts in physics education indicate that physics self-efficacy plays an important role in both learning and interest in the discipline [22–24], and that classroom experiences can change students' self-efficacies [24,25]. Researchers find that self-efficacy compares favorably to other predictive attributes for academic success and career decision making [26]. Additionally, differences in self-efficacy are associated with differences in gender performance and representation [23,27,28], and have implications for identifying academically at-risk students [29].

Confidence and self-efficacy, although distinct constructs, are very closely related and often highly correlated. Confidence refers to one's belief in one's own ability, which can be equivalent to having high self-efficacy. Confident students have been found to perform better on examination, though correct students were not always confident about their correct answers [12,30,31]. In one of these studies, the relationship between confidence and correctness is much stronger among better-performing students, whereas poorly performing students' correctness does not vary much with confidence [12]. In another study, there is a larger degree of overconfidence among lower-performing subjects [30]. Both of these findings, based upon students' reports of confidence in their own performances, capture the same link between metacognitive skills and content knowledge discussed in Ref. [13].

Some argue that self-efficacy is very specific to a task in a given situation and self-concept, a different construct, applies to more general beliefs about competence [32]. Others refer to self-efficacy as a more general sense of one's ability (e.g., in a science discipline), and that performance accomplishments in multiple contexts can effectively improve self-efficacy across contexts [14]. Thus, the lines between confidence in one's ability and confidence in one's actions in a specific circumstance are somewhat vague. In this work, we consider self-efficacy to refer to beliefs about one's ability in the discipline (i.e., physics) and confidence to refer to specific beliefs about responses to particular content-related questions.

Researchers have connected self-efficacy, confidence, and metacognition to student learning outcomes and used these connections to motivate practices for instructors [33]. In the science classroom, some argue that metacognition plays a more important role when problem solving involves grappling with unfamiliar tasks and methods than when it involves executing known procedures [10,34]. Research into productive failure and enhanced learning from unsuccessful retrieval attempts suggests that challenging students' confidence and abilities may ultimately benefit learning and performance outcomes [35–39]. However, these studies do not shed light on how instructors should interpret students' expressions of confusion. In fact, we know of only a few studies to directly address student confusion in the sciences, and these studies emphasize its complex role in student learning [1,4]. To what extent are metacognition, confidence, and self-efficacy related to confusion? Is confusion a good sign because it indicates that students engage in metacognition or a bad sign because it indicates a lack of confidence? And how do students' expressions of confusion relate to other aspects of learning and engagement? In this study, we address these questions in one context.

II. METHODS

We analyzed data collected from 137 students across a two-semester sequence of introductory physics courses at Harvard University during the fall of 2010 and the spring of 2011. Of the 137 students, 57 completed both courses; 106 students participated in the fall and 88 students participated in the spring [40]. As described below, we recorded student expressions of confusion before class, performance on reading exercises, confidence in their performance on reading exercises, pre- and post-course self-efficacy (in the spring), and performance on graded course activities.

The two-semester sequence of courses was designed for nonphysics concentrators. A different instructor taught each semester; one of us (E. M.) was the instructor for the second-semester course. Both instructors implemented the preclass components of Just-in-Time Teaching (JiTT) and emphasized the importance of completing the reading assignments to the students. JiTT is a teaching and learning strategy in which students engage with the material and submit a completed "WarmUp" activity the evening before class, so the instructor has the opportunity to incorporate this feedback and focus on student difficulties in class [41]. Specifically, as implemented here, this activity consists of two content-related questions and one "confusion" question, which requires students to reflect on their understanding of the material. In addition to these preclass activities, the instructor of the first-semester course occasionally posed in-class conceptual questions to students for discussion, but more often presented material through lectures and demonstrations. The instructor of the second-semester course employed Peer Instruction (PI), using demonstrations in conjunction with in-class conceptual questions and discussion; the use of lecture was limited [42]. In both semesters, students received credit for participation by responding to both preclass exercises and in-class questions.

In this study, we consider students' preclass responses to the confusion question as the sole metric of confusion. Students were assigned 22 reading exercises during the fall semester and 21 reading exercises during the spring semester. Students submitted responses online.

In the following sections, we detail each of the variables used to quantify students' expression of confusion, reading assignment correctness, confidence in reasoning, course performance, and physics self-efficacy. We then describe our approach to analysis.

A. Confusion index

Although the free-response format is highly conducive to rich student responses, barring automated text analysis, we could not analyze large amounts of such data. Therefore, we formatted questions so that student responses could be analyzed efficiently without sacrificing the detailed information afforded to instructors by the free-response format.

The confusion question consisted of three parts, as shown in Fig. 1. Depending on how students responded to the first part ("In the material that you just completed reading, did you find anything difficult or confusing?"), the second part appeared in one of two ways. If students responded "yes" in part one, part two asked what topics students found confusing; if students responded "no" in part one, part two asked what topics students found interesting. Part three ("Please elaborate. Do you have any questions?") allowed students to elaborate in free-response format.

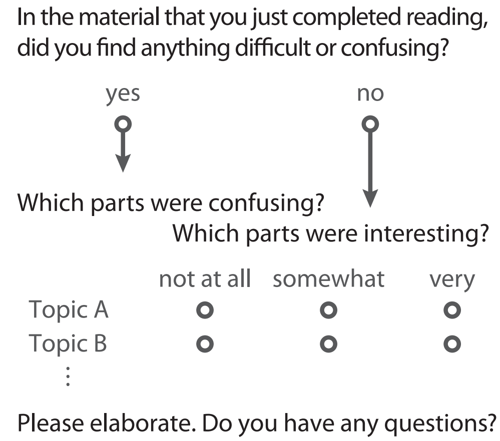

FIG. 1. Before class, students are asked to respond to this three-part confusion question regarding the reading material that they just completed as part of a reading assignment. The second part differs depending on whether students respond yes or no in the first part, as shown. The third part is free response.

Posing the question this way serves multiple purposes. The question is balanced, so students cannot answer more quickly by expressing no confusion. By asking students if they are confused and gauging their confusion on the topics within the reading assignment, the question is more thorough than simply asking which points students found confusing or *most* confusing. The free-response information is most valuable to the instructor, so the third part of the question ensures that students can be specific and thorough without too much effort. Although the inclusion of the term "difficult" in addition to "confusing" makes students' responses to the first part of the question more ambiguous than if only the term confusing were included, the second part of the question unambiguously asks students to reflect on their confusion.

We established a quantitative value for confusion based on students' responses to the second part of the question, in which they state their degree of confusion about topics within a reading assignment. Because students tend to express at least some confusion relatively frequently, students' yes or no responses are much more skewed toward yes than their degree of confusion, which is more normally distributed. We associate "not at all," "somewhat," and "very" with the numerical values 0, 1, and 2, respectively, and average these values across all topics listed within a single reading assignment. We call this average value the *confusion index*. If a student expresses no confusion in the first part of the question, the confusion index is zero. In this way, we can quantify the degree of confusion of different students. Any subsequent references to confusion should be interpreted as "average degree of confusion after reading, as calculated using the confusion index."

We emphasize that quantifying students' expressions of confusion this way is intended only to capture the magnitude of their responses in this context. We do not presume to generalize students' responses to this question to confusions in other contexts (such as in-class discussions or examinations), and we make no assumptions about what students are thinking when they submit responses. The context is authentic, in that instructors use responses to this question to assess students' ideas and prepare for class. This authenticity motivates our research questions and imparts meaning to our findings. We do not claim that this confusion question is the optimal or most valid means of assessing confusion, as we do not have data to address this.

In each assignment, the confusion question was posed *before* the two content-related questions, followed by a final opportunity to revise the response to the initial confusion question. We posed the confusion question first because we did not want challenging content-related questions to influence students' responses about their confusion. The opportunity to revise confusion was included as a separate, free-response question primarily to dissuade students from changing their responses to the initial question. Only the initial responses were analyzed. It is possible that asking students to express their confusion first may have influenced performance on the content-related questions.

B. Reading assignment correctness

The two content-related questions in each reading assignment were intended to be thought provoking and counterintuitive, so students would have to critically engage with the material before class [43]. The students were required to explain their reasoning for these two questions. At least two—and sometimes as many as three or four—researchers and instructors reviewed and discussed each content-related question before dissemination to students to ensure that they depended upon the content in the reading but could not be answered by simply locating a key passage in the text. Whenever possible, questions were designed so that response choices spanned the entire space of possible answers (e.g., *A greater than B*, *B greater than A*, *A and B are equal*). This way, each question could be posed as a multiple-choice question with a free-response field for students to explain their reasoning, limiting students' ability to use the response choices to guide their reasoning.

Students' responses were associated with numerical values of 0, 1, or 2, depending on whether they correctly answered neither, one, or both of the content-related questions, respectively.

C. Confidence in reasoning

In posing the questions this way, we also probed one more facet of student understanding: confidence in reasoning on each content-related question. After explaining their reasoning, students were asked to rate their confidence in the explanation that they provided on a scale of low, medium, and high. These levels were associated with respective sequential numerical values for quantitative analysis.

We quantify students' confidence in this way because we are interested only in the magnitude of their confidence in their reasoning on these reading assignment questions. We do not claim that students' confidence here relates to students' more general confidence in physics, or even to their confidence on similar questions in an exam setting. On a particular reading assignment, if a student expressed confusion about the material but expressed confidence in how he or she had answered the content-related questions, we are able to separate those factors by assessing confidence this way.

D. Course performance

In addition to analyzing student performance and confidence on each of the content-related questions and student expressions of confusion, we analyzed performance on course-related activities (problem sets, laboratory activities, exams, and cumulative grades). We use exploratory factor analysis to determine if the correlations among different variables describe one or more uncorrelated aspects of performance [44]. For both the fall and spring semesters, we find that the grades for reading exercises (evaluated for participation only), laboratory activities, midterm exams, final exam, and the final grade (which, scored out of 100%, is a continuous variable) all describe one factor. Therefore, we use the final grade, which is sufficiently normally distributed and is among the most heavily weighted variables in the factor, as the primary summative measure of students' performance in the course.

E. Physics self-efficacy

Additionally, in the spring semester, we analyzed pre- and postcourse performance on a survey of self-efficacy in physics, which is available in Ref. [29]. This survey consists of 25 items to which students responded on a 5-point Likert scale from "strongly disagree" to "strongly agree." Items on this survey were based on the validated Sources of Self-Efficacy in Science Courses–Physics (SOSESCP) instrument [22,24]. The survey was administered as a pretest during the first week of classes on paper and as an online post-test between the last class and the final examination.

Some items on the survey are designed to assess physics self-efficacy, and others are designed to assess Peer Instruction self-efficacy [29]; only the portion pertaining to physics self-efficacy is considered here. Of the 25 items, a subset of seven items (also identified in Ref. [29]) relates to general physics self-efficacy. This subset includes such items as "I am confident I can do the work required for this course" and "When I come across a tough physics problem, I work at it until I solve it." For these items, Cronbach's coefficient of reliability (α) was 0.85 on the pretest and 0.83 on the post-test, as reported in Ref. [29]. Students' average responses on these seven items were used to quantify their physics self-efficacy as a continuous variable. Both pre- and postcourse responses are sufficiently normally distributed for linear regression analysis.

As this survey was only administered in the spring semester, we cannot consider self-efficacy of students during the fall semester.

F. Approach to analysis

We conduct factor analysis on students' confusion indices, correctness, and confidence ratings across all of the individual assignments each semester, and we find that each of these variables is strongly described by just one factor. This does not indicate that students' expressions of confusion, for example, do not change over time. Indeed, we find that some topics, such as the propagation of electromagnetic waves, coincide with much higher average confusion indices than other topics, such as electric circuits. Instead, the single factor indicates that individual students' responses are strongly correlated; someone expressing more confusion on one topic tends to also express more confusion on other topics. Therefore, we average students' confusion indices, correctness, and confidence ratings from each reading exercise across the semester and consider these values as representative of the underlying factors.

We average the 43 independent measurements (22 from the fall semester, 21 from the spring semester) of confusion index, correctness, and confidence ratings in order to build regression models in which these average values may be included alongside course-wide measurements of self-efficacy and overall course performance. Although a student's incoming physics self-efficacy may be only slightly related to expression of confusion or confidence on any specific topic, self-efficacy may relate more strongly to these average values.

Our analysis depends on the students' ability to rate their confidence, confusion, and physics self-efficacy on scales that are not calibrated to any external standards. We did not provide descriptive rubrics or sample responses to help standardize responses out of concern that such materials would exhaust students' patience or influence their responses.

We consider only the presemester survey of physics self-efficacy in our analysis because we find that the pre- and postsemester surveys are highly correlated. More students participated in the presemester survey. We do not mean to imply that students' self-efficacies are static in general; indeed, other researchers report changes during instruction [24,25]. However, in this particular course, we do not observe statistically significant changes in students' self-efficacies, as measured by pre- and postsurveys.

In the cases of such variables as confusion index, confidence, and self-efficacy, the native scales are not readily interpretable, so we can more easily describe observations using standard scores, or z scores. These scores allow us to compare the relative strength of relationships among variables with different units, though they can also mask whether or not a relationship is meaningful on an absolute scale. Therefore, we also use the native scale for confusion index and final grade in our analysis.

III. RESULTS

A. Correlations

We first investigate the correlations among variables discussed above for both the fall and spring semesters. As shown in Table I, during the fall semester, the confusion index is negatively related to correctness on the content-related questions and confidence in reasoning on those questions, but not significantly related to final grade. In other words, students who express more confusion about reading material on average tend to express less confidence in their reasoning, and also tend to perform more poorly on content-related questions. The strengths of these relationships vary; confusion index is strongly negatively related to confidence and more moderately negatively related to correctness. However, we also see that many of the non-confusion variables are strongly related to one another, suggesting that multiple regression analysis of confusion and these other variables may reveal different relationships. The relationships during the spring semester, shown in Table II, are very similar to those observed in the fall semester. Self-efficacy is strongly negatively related to expressions of confusion, and it is also strongly positively related to all of the other characteristics under consideration here.

TABLE I. Correlations of characteristics of student learning and engagement, fall 2010 ($N = 106$).[a]

	Confusion index	Correctness	Confidence	Final grade
Confusion index	1.00			
Correctness	−0.24*	1.00		
Confidence	−0.43***	0.34***	1.00	
Final grade	−0.04	0.45***	0.24*	1.00

[a]($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

TABLE II. Correlations of characteristics of student learning and engagement, spring 2011 ($N = 88$).[a]

	Confusion index	Correctness	Confidence	Final grade	Self-efficacy (pre)
Confusion index	1.00				
Correctness	−0.37***	1.00			
Confidence	−0.52***	0.39***	1.00		
Final grade	−0.12	0.57***	0.26*	1.00	
Self-efficacy (pre)	−0.50***	0.34**	0.50***	0.37***	1.00

[a](*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

B. Regression analysis

To explore the interactions among these characteristics, we build a series of multiple regression models. Because the primary goal of this analysis is to better understand how multiple variables are related to students' expressions of confusion, we are most interested in linear regression models in which expression of confusion is the outcome variable. Specifically, we first investigate the relationship between students' expressions of confusion about reading assignments and correctness on questions related to reading assignments. Even without controlling for other variables, this relationship reveals whether students' initial confusion, upon coming to class after completing a reading assignment, is positively or negatively related to initial performance. Then, we investigate the relationship among students' expressions of confusion and the multiple other variables discussed here. We are particularly interested in the relationship between students' expressions of confusion and final grade, as final grade is our proxy for ultimate performance. The lack of significant correlations between final grades and expressions of confusion shown in Tables I and II suggest that the two variables are not related. However, it is possible that this apparent lack of relationship results from competing positive and negative relationships that combine to show no net relationship. Controlling for other variables in our regression models allows us to separate competing relationships and reveal a potential hidden relationship between our two variables of interest, expression of confusion and final grade.

Tables III and IV summarize regression models for the fall and spring semesters, respectively. Standard coefficients are displayed. These models were chosen for specific reasons, as described above. Models $1f$ and $1s$ (where the letters "f" and "s" represent fall and spring semesters, respectively) address our most immediate research question. Model $2f$ is the most comprehensive model from the fall semester, and model $2s$ is included for direct comparison. Model $4s$ is the most comprehensive model from the spring semester. Model $3s$ is included because the omission of final grade as a variable allows us to better visualize the relationship captured in model $4s$, as discussed in more detail below.

1. Simple regression models

When only the confusion index and reading exercise correctness are included in the model (models $1f$ and $1s$), there is a statistically significant negative relationship between the two variables; in other words, when no other variables are included in the model, an increase in correctness of 1 standard deviation is associated with a decrease in confusion index of approximately 0.24 standard deviations during the fall semester and 0.37 standard deviations in the spring semester. In social sciences, these might be considered roughly medium effect sizes [45]. Only 5.8% and 13.3% of the variation in confusion is explained by correctness in the fall and spring semesters, respectively.

2. Multiple regression models

When final grade, confidence in reasoning, and reading exercise correctness are included in the model (models $2f$ and $2s$), final grade is not statistically significantly related to the confusion index. When presemester self-efficacy is

TABLE III. Fitted linear regression models explaining variation in confusion index by selected variables of student learning and engagement, fall 2010 ($N = 106$).[a]

Variable	Model $1f$[b]	Model $2f$[c]
Correctness	−0.24*	−0.16
Confidence		−0.41***
Final grade		0.13

[a](*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).
[b]$R^2 = 0.058$, $RMSE = 0.98$.
[c]$R^2 = 0.208$, $RMSE = 0.90$.

TABLE IV. Fitted linear regression models explaining variation in confusion index by selected variables of student learning and engagement, spring 2011 ($N = 88$).[a]

Variable	Model $1s$[b]	Model$2s$[c]	Model $3s$[d]	Model $4s$[e]
Correctness	−0.37***	−0.27*	−0.14	−0.26*
Confidence		−0.46***	−0.32**	−0.31**
Final grade		0.15		0.23*
Self-efficacy (pre)			−0.29**	−0.34**

[a](*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).
[b]$R^2 = 0.133$, $RMSE = 0.94$.
[c]$R^2 = 0.321$, $RMSE = 0.84$.
[d]$R^2 = 0.366$, $RMSE = 0.81$.
[e]$R^2 = 0.401$, $RMSE = 0.79$.

considered alongside final grade, confidence in reasoning and reading exercise correctness (model 4*s*), final grade is positively related to the confusion index; an increase in score on final grade of 1 standard deviation is associated with an increase of 0.23 standard deviations in confusion index. This might be considered a small-to-medium effect size [45]. We see that 40.1% of the variation in confusion is explained. All four of the variables are statistically significantly related to confusion index.

To better illustrate these relationships, in Fig. 2 we highlight models 2*f*, 2*s*, and 4*s*. Because each variable is represented as a standardized *z* score, all of the variables may be represented on the same scale. We see that the relationship between confusion and final grade is positive in all of these models, albeit not necessarily statistically significant, while the relationships between confusion and each of the other variables are negative. If there were no relationship between expression of confusion and final grade, one might expect less consistency in coefficient values across different models and semesters. We do not observe any interaction effects among the variables.

To highlight the relationship between confusion and final grade, in Fig. 3 we display the difference between each student's average confusion index and the confusion index "predicted" by model 3*s*, which includes reading assignment correctness, confidence in reasoning, and self-efficacy during the spring semester. In this case, we are no longer using standard scores; the vertical axis displays the difference in confusion index on an absolute scale, and the horizontal axis displays final grades binned roughly according to letter grade. As shown, students who earn high final grades in the course express more confusion than explained by model 3*s*, and students who earn low final grades express less confusion than explained by model 3*s*.

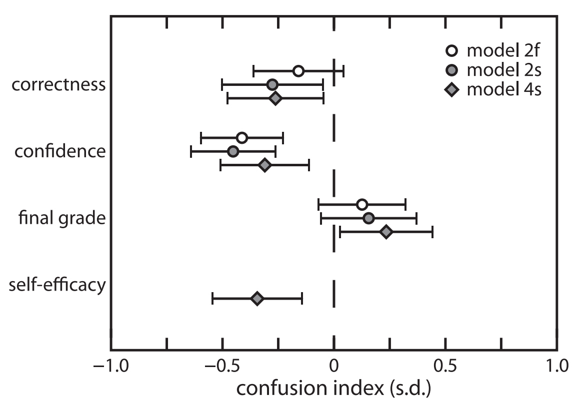

FIG. 2. Fitted linear regression models explaining variation in confusion index by selected variables of student learning and engagement. Models 2*f*, 2*s*, and 4*s* are displayed here, where the letters *f* and *s* represent fall and spring semesters, respectively. The bars represent the 95% confidence interval around each of the mean values; if the confidence interval does not overlap the "zero line," then the difference is statistically significant at the $p = 0.05$ level.

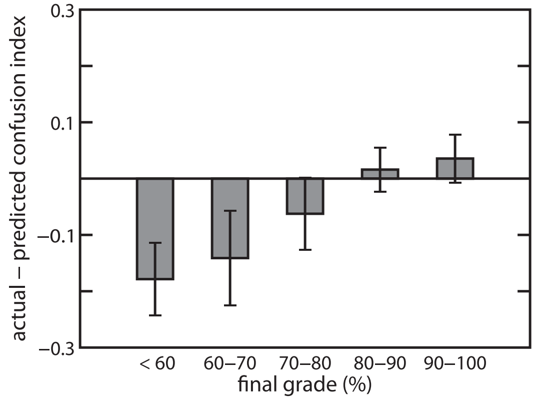

FIG. 3. Difference between each student's average confusion index and the confusion index predicted by regression model 3*s* (which includes the reading assignment correctness, confidence in reasoning, and self-efficacy during the spring semester), binned according to final grade. The vertical axis displays the *difference* in confusion index; although the absolute scale of confusion index ranges from 0 (not at all confused) to 2 (very confused), the difference between actual and predicted values can be positive or negative. The horizontal axis displays final grades binned roughly according to letter grade. The error bars represent the standard error of the mean of binned values. The numbers *N* of students in each bin are as follows: 60 or lower ($N = 2$), 60–70 ($N = 3$), 70–80 ($N = 14$), 80–90 ($N = 41$), and 90–100 ($N = 28$).

Although the difference between actual and "predicted" confusion index within each of the final grade ranges is not statistically significantly different from zero, the apparent pattern is in keeping with regression model 4*s*.

IV. DISCUSSION

The negative relationship between confusion index and correctness on reading exercises (models 1*f* and 1*s*), in conjunction with the positive relationship between confusion index and final grade (model 4*s*), suggests that the confusion index captures multiple aspects of learning and engagement, both positive and negative. When final grade, confidence in reasoning and self-efficacy are measured alongside reading assignment correctness, the positive aspects of expressing confusion about the reading emerge as being positively associated with final grade, while the negative aspects of confusion are captured by measures of self-efficacy and confidence in reasoning. Thus, we make two claims about students' expressions of confusion after reading in this introductory physics course:

(1) When relevant factors are *not* controlled for, one cannot separate positive and negative aspects of students' reported confusion, rendering such confusion negative or uninformative, depending on the learning outcome of interest.

(2) When relevant factors *are* controlled for, one may be able to separate and identify positive aspects of confusion.

In other words, confusion reflects a mixture of qualities that instructors may not be able to isolate without collecting additional information from students.

The multiple regression analyses presented above suggest that the three probes of self-regulated learning employed here—confusion index, ratings of confidence in reasoning, and evaluation of physics self-efficacy—relate to distinct aspects of students' engagement. By assessing confidence in reasoning independent of confusion, we seem to be able to separate students' doubts about the content-related questions in each reading assignment from their more general confusion about the material. In the spring semester, by assessing self-efficacy in physics independently, we also seem to be able to separate students' confusion stemming from their perception of their own abilities in physics. Thus, the positive aspects of confusion may relate to a more general, metacognitive engagement.

JiTT and PI are designed to make physics instruction more metacognitive. Frequent opportunities for reflection about sources of confusion, coupled with collaborative interactions in the classroom and a strong focus on the value of self-directed learning, are built on the notion that such activities benefit learning outcomes. It is encouraging that the relationships observed across both semesters are quite similar, despite the much more prominent role of lecture during the fall semester, but the present analysis is not intended to be immediately generalized across contexts. The specific context of this study—this student body, how instruction is carried out, how confusion and other factors are assessed—may crucially affect the observed relationships.

Additionally, the fact that students are encountering material through reading, as opposed to lecture or peer discussion, may also influence relationships. Poor performance or low confidence in reading assignments may stem from challenging content, weak reading comprehension, or a combination of these factors. However, challenges to students' encoding of new material are not unique to reading activities. Science reading, like science learning in general, involves the interpretation of presented material and integration with prior knowledge and concurrent experiences [46]. During lecture or peer discussion, classroom distractions and nuances of social engagement can also compound the inherent challenges of learning new material. The relationships among confusion after reading and the other variables highlighted here may stem in part from challenges in reading comprehension, but that does not diminish the value of recognizing that these observed relationships exist in an authentic learning environment.

Researchers have focused on improving students' reading comprehension skills through engagement in metacognition across general and even physics contexts [11,47]. Similarly, the metacognitive activities performed by students that provide data for this analysis could influence their behavior. The very act of asking students to self-assess confidence in their reasoning on each reading assignment may have strengthened the positive relationship between confidence and performance that was observed. However, because the context in which these data are collected is already quite specific, such incidental metacognitive activities do not negatively impact our claims. Our goal is not to simply assume that these findings apply broadly, but rather to provide insight about the value of confusion as a means of assessing student learning and engagement, raise caution about assuming that confusion is always negative, and raise questions about what positive aspects of confusion might actually tell instructors about student learning.

A. Implications for instruction

In the courses described here, students express their confusion after first encountering the material through reading assignments and before discussing the material in class. Asking students to describe their specific confusions before class allows instructors to tailor activities to students' prior knowledge and more effectively attend to sources of confusion. One might expect that the critical thinking and introspection required for confusion to be positive would only occur after repeated exposure to the material and some discussion. Nonetheless, we find evidence that these positive processes may take place even at this earliest point of assessment. Moreover, the absence of a negative relationship between expression of confusion and final grade invalidates the notion that confusion is simply negatively related to all learning outcomes. Although the means of assessing confusion described here taps into too many distinct aspects of engagement (initial knowledge, confidence, and self-efficacy) to provide a direct means of assessing the role of students' metacognition in learning, the small-but-positive relationship between expression of confusion and final grade suggests the potential for such a purpose.

With evidence here of potentially positive aspects of confusion in instruction, we may now ask whether these aspects change when confusion is assessed at different stages of instruction or alongside different activities in class. Perhaps students' expressions of confusion after repeated exposure to the material, rather than initial exposure, relate to critical thinking even more strongly. Or perhaps confusion expressed in different classes and different populations may convey entirely different information about learning and engagement. Confusion expressed in class may differ from confusion expressed before class. We have only begun to explore these avenues, and therefore propose that instructors assess students' confusion early, often, and without assumptions of what confusion (or the absence thereof) may convey.

V. CONCLUSION

Students who express more confusion before class tend to also display lower confidence in reasoning, lower self-efficacy, and, to a slight degree, weaker performance on reading exercises. However, when we control for all of these factors, students who express more confusion tend to also perform better overall, as measured by final grade. In other words, we are able to identify positive aspects of confusion, creating the possibility for genuine assessment of the positive impact of students' metacognition in achieving learning outcomes.

ACKNOWLEDGMENTS

Several people contributed to the work described in this paper. J. D., I. A., and E. M. conceived of the basic idea for this work. J. D. designed and implemented the study. J. D. and I. A. analyzed the results. E. M. supervised the research and the development of the manuscript. J. D. wrote the first draft of the manuscript; all authors subsequently took part in the revision process and approved the final copy of the manuscript. Brian Lukoff provided feedback on the manuscript.

APPENDIX: FALL SEMESTER, EXERCISE 20

1. Confusion topics

Forces in a fluid, buoyancy, fluid flow, surface effects, pressure and gravity, working with pressure, Bernoulli's law, viscosity, and surface tension.

2. Content-related questions

(Correct answers are indicated in bold font.)

Q1. You have two different balloons. You inflate both of them as much as you possibly can (until your lungs can't push any more air inside the balloons), and you find that one balloon is clearly much bigger than the other. You then connect the balloons to each end of a straw that you've clamped in the middle. Which way does the air flow once you remove the clamp?

(1) Air flows from the large balloon to the small one.

(2) Air flows from the small balloon to the large one.

(3) No air flows between the balloons.

(4) More information is needed to determine how the air flows.

Q2. There are two beakers, one filled with water and the other filled with vegetable oil. When you put a particular ball in the water, it just barely floats (as in, only a small fraction of the ball is above the surface of the water). When you put the same ball in the oil, it sinks to the bottom. What happens when you put the ball in the water, and then you pour the vegetable oil on top?

(1) The ball remains floating at the same exact height.

(2) The ball floats lower, just barely touching the oil-water interface.

(3) The ball floats higher, still partially within the water.

(4) The ball sinks to the bottom of the water.

(5) This cannot be determined from the given information.

[1] A. Lipson, The confused student in introductory science, Coll. Teach. **40**, 91(1992).

[2] J. Piaget, *The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development* (University of Chicago Press, Chicago, 1985).

[3] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).

[4] A. Lipson, Learning: A momentary stay against confusion, *Teach. Learn.* **4**, 2 (1990).

[5] L. C. McDermott, Guest comment: How we teach and how students learn—a mismatch?, Am. J. Phys. **61**, 295 (1993).

[6] E. F. Redish, *Teaching Physics with the Physics Suite CD* (Wiley, Hoboken, NJ, 2003).

[7] J. H. Flavell, Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry, Am. Psychol. **34**, 906 (1979).

[8] M. Ward, L. Gruppen, and G. Regehr, Measuring self-assessment: Current state of the art, Adv. Health Sci. Educ. **7**, 63 (2002).

[9] H. T. Everson and S. Tobias, The ability to estimate knowledge and performance in college: A metacognitive analysis, Instr. Sci. **26**, 65 (1998).

[10] D. Rickey and A. M. Stacy, The role of metacognition in learning chemistry, J. Chem. Educ. **77**, 915 (2000).

[11] A. Koch, Training in metacognition and comprehension of physics texts, Sci. Educ. **85**, 758 (2001).

[12] F. J. Sinkavich, Performance and metamemory: Do students know what they don't know?, *J. Instruct. Psych.* **22**, 77 (1995).

[13] J. Kruger and D. Dunning, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments, J. Personality Social Psychol. **77**, 1121 (1999).

[14] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).

[15] A. Bandura, The explanatory and predictive scope of self-efficacy theory, J. Social Clinical Psychol. **4**, 359 (1986).

[16] E. L. Usher and F. Pajares, Sources of self-efficacy in school: Critical review of the literature and future directions, Rev. Educ. Res. **78**, 751 (2008).

[17] R. W. Lent, S. D. Brown, and K. C. Larkin, Relation of self-efficacy expectations to academic achievement and persistence, J. Counsel. Psychol. **31**, 356 (1984).

[18] D. A. Luzzo, P. Hasper, K. A. Albert, M. A. Bibby, and A. Martinelli, Effects of self-efficacy-enhancing interventions

on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, J. Counsel. Psychol. **46**, 233 (1999).
[19] K. D. Multon, S. D. Brown, and R. W. Lent, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, J. Counsel. Psychol. **38**, 30 (1991).
[20] S. Andrew, Self-efficacy as a predictor of academic performance in science, J. Adv. Nurs. **27**, 596 (1998).
[21] J. Pietsch, R. Walker, and E. Chapman, The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school, J. Educ. Psychol. **95**, 589 (2003).
[22] H. S. Fencl and K. R. Scheel, Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses, AIP Conf. Proc. **720**, 173 (2004).
[23] V. Sawtelle, Ph.D. thesis, Florida International University, 2011.
[24] H. Fencl and K. Scheel, Engaging students: An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course, *J. Coll. Sci. Teach.* ***35****, 20 (2005)*.
[25] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, Phys. Rev. ST Phys. Educ. Res. **8**, 020111 (2012).
[26] R. W. Lent, S. D. Brown, and K. C. Larkin, Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking, J. Counsel. Psychol. **34**, 293 (1987).
[27] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Unpacking gender differences in students' perceived experiences in introductory physics, AIP Conf. Proc. **1179**, 177 (2009).
[28] A. L. Zeldin and F. Pajares, Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers, Am. Educ. Res. J. **37**, 215 (2000).
[29] J. Schell, B. Lukoff, and C. Alvarado, Using early warning signs to predict academic risk in interactive, blended teaching environments, *J. Internet Learning* (to be published).
[30] J. J. Shaughnessy, Confidence-judgment accuracy as a predictor of test performance, J. Res. Person. **13**, 505 (1979).
[31] J. Engelbrecht, A. Harding, and M. Potgieter, Undergraduate students' performance and confidence in procedural and conceptual mathematics, Int. J. Math. Educ. Sci. Technol. **36**, 701 (2005).
[32] E. A. Linnenbrink and P. R. Pintrich, The role of self-efficacy beliefs in student engagement and learning in the classroom, Read. Writ. Q. **19**, 119 (2003).
[33] G. Schraw, K. Crippen, and K. Hartley, Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning, Res. Sci. Educ. **36**, 111 (2006).
[34] A. H. Schoenfeld, Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics, in *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (Macmillan Publishing Co, Inc, New York, NY, 1992), p. 334–370.
[35] H. L. Roediger and J. D. Karpicke, Test-enhanced learning taking memory tests improves long-term retention, Psychol. Sci. **17**, 249 (2006).
[36] N. Kornell, M. J. Hays, and R. A. Bjork, Unsuccessful retrieval attempts enhance subsequent learning, J. Exp. Psychol. Learn. Mem. Cogn. **35**, 989 (2009).
[37] L. E. Richland, N. Kornell, and L. S. Kao, The pretesting effect: Do unsuccessful retrieval attempts enhance learning?, J. Exp. Psychol. Appl. **15**, 243 (2009).
[38] M. Kapur, Productive failure, Cognit. Instr. **26**, 379 (2008).
[39] M. Kapur, Productive failure in mathematical problem solving, Instr. Sci. **38**, 523 (2010).
[40] One student in the spring semester did not complete the presemester self-efficacy survey. Excluding this student does not change the observed relationships, so we exclude this student in order to maintain a consistent data set of 88 students across analyses.
[41] G. M. Novak, E. T. Patterson, A. Gavrin, and W. Christian, *Just-in-Time Teaching: Blending Active Learning with Web Technology*, 1st ed. (Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999).
[42] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
[43] The content-related questions from one of the reading assignments are included in the appendix.
[44] A. Costello and J. Osborne, Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, *Pract. Assess. Res. Eval.* **10** (2005).
[45] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (Rev. ed.)* (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, 1977), Vol. XV.
[46] L. D. Yore and J. A. Shymansky, Reading in science: Developing an operational conception to guide instruction, J. Sci. Teach. Educ. **2**, 29 (1991).
[47] A. S. Palinscar and A. L. Brown, Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, Cognit. Instr. **1**, 117 (1984).