Student satisfaction in interactive engagement-based physics classes

Jon D. H. Gaffney*

Eastern Kentucky University, Department of Physics and Astronomy, Richmond, Kentucky 40475, USA

Amy L. Housley Gaffney†

University of Kentucky, School of Information Science, Lexington, Kentucky 40506, USA

(Received 16 March 2016; published 23 August 2016)

Interactive engagement-based (IE) physics classes have the potential to invigorate and motivate students, but students may resist or oppose the pedagogy. Understanding the major influences on student satisfaction is a key to successful implementation of such courses. In this study, we note that one of the major differences between IE and traditional physics classes lies in the interpersonal relationships between the instructor and students. Therefore, we introduce the interpersonal communication constructs of instructor credibility and facework as possible frameworks for understanding how instructors and students navigate the new space of interactions. By interpreting survey data ($N = 161$ respondents in eight sections of an IE introductory algebra-based physics course), we found both frameworks to be useful in explaining variance in student ratings of their satisfaction in the course, although we are unable to distinguish at this point whether instructor credibility acts as a mediating variable between facework and course satisfaction.

DOI: [10.1103/PhysRevPhysEducRes.12.020125](https://doi.org/10.1103/PhysRevPhysEducRes.12.020125)

I. INTRODUCTION

Physics has seen numerous pedagogical reforms that move instruction from a teacher-centered to a learner-centered model; this shift in focus has led to numerous research-based instructional strategies (RBIS) and to interactive engagement-based (IE) learning environments for physics, such as Student-Centered Active Learning Environment for Undergraduate Programs (SCALE-UP) [1]. These IE learning environments are known to yield learning gains better than those of traditional lecture or lab classes, and RBIS exist in the literature as resources for instructors who want to improve their students' understanding of physics. However, implementation of RBIS within the context of an IE learning environment is a highly complicated process, involving many factors [2]. One key factor that decision makers need to understand is that *the adoption of an IE learning environment changes the fundamental nature of the interpersonal relationships between the instructor and the students.*

Traditional teacher-based courses adopt a mode of communication that is largely one directional; while instructors may receive feedback from students, that feedback is typically limited to short in-class answers and out-of-class interactions.

However, in learner-centered models, the communication becomes more conversational during class time itself. Instructors, more than in traditional classes, must listen to students and respond to their concerns immediately. Such interactions are inherently more risky, as they involve more vulnerability from both the instructor and the student [3]. Therefore, while student learning of physics content may be the ultimate goal of IE courses, instructors cannot ignore affective outcomes. Indeed, instructors may give up RBIS implementation attempts due to pushback from their students [4]. Consequently, understanding the cause of student resistance is vital for instructors who wish to mitigate such pushback.

Investigating student resistance necessarily involves both understanding the phenomenon of resistance and creating a theoretical framework with which to approach it. One study addresses the former by developing an observation protocol to classify and articulate instances of student resistance in active-learning engineering classrooms [5]. The approach of Shekhar *et al.* includes observation that focuses on both student pushback and instructor approaches to introducing activities. We take a complementary approach by postulating that students may experience discomfort due to the new forms of interaction between instructor and student. Moreover, if we can detail the nature of that discomfort by understanding student-instructor interactions from a student perspective, we will have a theoretical framework on which to cast the results of empirical observations.

We begin by noting that instructors and students both bring their own experiences and expectations into the classroom. For instance, instructors vary in terms of their prior experiences with respect to both physics and any RBIS they plan to implement. Consequently, there are

*jon.gaffney@eku.edu

†Present address: Hamilton College, Oral Communication Center, Clinton, NY 13323, USA.
algaffne@hamilton.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

substantial differences in how classrooms look and feel, even among classes at the same institution; this difference is emphasized, rather than muted, by the IE environment [6]. While implementing at least one RBIS generally improves students' learning, the amount of improvement has considerable variation [7]. One factor is that instructors establish different norms in the class and thus student experiences differ [8]. Furthermore, the alignment of an instructor's individual epistemological beliefs about student-centered learning with the classroom in which the course is taught may substantially impact the effectiveness of the course, especially for sociotechnical (that is, student-centered and technology-enhanced) classrooms [9].

Additionally, student backgrounds need to be considered when implementing IE classes, as the expectations of students at an implementation site may vary greatly from those where the IE environment was developed. For example, most of the published research regarding IE environments was conducted in the United States. Hitt *et al.* found that an implementation of an IE course in the United Arab Emirates was largely successful, but that some minor adaptations were necessary to account for students' different expectations [10]. It is worth pointing out that the opposing expectations were suspected to largely come from prior classroom experiences rather than directly from cultural factors. Thus, it is reasonable to expect that different experiences between student populations may account for some of the variation of success in implementations of IE courses; for example, we might anticipate that the population of students in an IE physics class at different institutions may have substantially different distributions of learning profiles [11].

However, we should not only consider the students and instructors in isolation; the impact of new interactions between students and instructors is also a major factor in understanding IE physics classrooms. Here, we set out to explore interpersonal interactions in several sections of active-learning introductory physics classes from the student perspective. In Sec. II, we introduce theoretical constructs that will help us explore these interactions: credibility and facework. Instructor credibility refers to students' perceptions of instructors' competence, trustworthiness, and goodwill. Facework refers to actions taken to protect one's own, or another's, desired self-image. Facework is examined here in terms of the behaviors instructors use to help protect students. We intersperse research questions and hypotheses to explore how those constructs relate to one another.

In Sec. III we describe the particular IE classes that we studied, and we explain the measures used to probe student expectations and experiences. In Sec. IV, we present our data and results, which we discuss at length in Sec. V, where we also discuss limitations of the study and implications for instructors who currently teach IE physics courses or who hope to implement RBIS.

II. THEORETICAL FRAMEWORK AND RESEARCH QUESTIONS

This project uses a theoretical framework we introduced to physics education research several years ago, known as expectancy violations theory. We relate that framework to other instructional constructs (i.e., credibility and facework) in order to propose research questions and hypotheses aimed at understanding student affect in IE classes.

A. Expectancy Violation Theory

In building a line of research on reformed classrooms, we have drawn on Expectancy Violations Theory (EVT) as a way of understanding students' experiences. In our everyday interactions, violations of our expectancies—that is, someone interacting in a way that we do not expect—may be viewed positively or negatively [12]. We assess violations based on our perceptions, not on an objective reality.

When we assess violations, we draw on both *who* committed the violation and *what* the violation was. Our affect toward the situation is also impacted by the extent of the violation. For example, imagine a classroom where the instructor has not sought feedback or questions for several weeks. Suddenly, that instructor calls on a student, unsolicited, to provide insight into a particular situation. The student is very likely to see that interaction as a violation. However, whether that violation is seen as positive or negative by the student depends on many things: how does the student feel about the instructor (does the student admire or disdain the instructor)? Is the instructor's request for a simple agreement, or is it more extensive, requiring a long, logically connected exposition? Is the student prepared to respond? Does the student's response receive praise or criticism from the instructor? Those variables impact how the student will leave the encounter, and they may impact future expectations as well. In cases where we have limited information about the other person or our current situation, we operate based on our previous related experiences. For example, at the start of a course, students tend to rely on their general schema of how previous courses proceeded.

Although much of the literature on EVT is in interpersonal communication contexts (e.g., nonverbal behaviors such as physical closeness [12] or response time to emails [13]), the theory has been useful in examining instructional settings as well. For example, when instructors presented information with less clarity than students expected, students reported less cognitive learning and motivation [14]. A more recent study focused on sharing personal information in class. The extent to which classmates expected personal information to be shared was a significant predictor of how much classmates liked the person who disclosed information [15].

Students' expectations of physics instructors are significantly different compared to instructors in other disciplines

such as tax accounting and business. For example, students reported that instructor difficulty in explaining the scientific method was more problematic for a physics instructor compared to others, while “inability to control class discussion” was only minimally problematic for a physics instructor [16]. Notably, that study did not distinguish based on instructional setting or approach, but instead focused on general impressions of instructors in various disciplines.

Our previous work focused on EVT in IE physics classes (e.g., SCALE-UP) has indicated that experiences were a driving force in students’ ultimate satisfaction with a course. For example, we found that the students’ reported experiences and attitudes toward those experiences played a significant role in student satisfaction. No measures of student expectations of the activities significantly contributed to satisfaction, but the expectations generally shifted after they had been oriented to the course. However, we also found that students’ attitudes did shift on some aspects during the semester, suggesting that students’ expectations and affect toward experiences are somewhat malleable [17].

In order to set the foundation for our further modeling of students’ expectations and experiences, we begin with a research question intended to describe the overall data:

RQ1: What are students’ (a) expectations, (b) experiences, and (c) attitudes toward different activities in an IE physics course?

Next, we pose a question that explores the impact of these expectations, experiences, and attitudes on an outcome:

RQ2: To what extent do (a) students’ expectations about how often they will do certain activities, (b) their reported frequency of experiences with those activities, and (c) attitudes toward those activities predict satisfaction with a course?

Previously, we noted that instructor behaviors are a major predictor of satisfaction. To help us understand how satisfaction develops, we turned to two concepts drawn from research on the role of communication in teaching and learning. The first, credibility, deals with students’ perceptions of an instructor, while the second, facework, deals with specific behaviors that an instructor may utilize to enhance students’ learning experience.

B. Credibility

Instructor credibility refers to students’ perceptions of the extent of an instructor’s competence, trustworthiness, and goodwill [18]. The first component, competence, refers to perceived expertise and intelligence. Trustworthiness refers to honesty and overall character of a person. Finally, goodwill refers to perceived caring toward students. A person who exhibits understanding, empathy, and

responsiveness will be seen as more caring compared to others who do not exhibit such characteristics. This conceptualization of credibility has been applied to a variety of contexts, including college teaching in diverse disciplines.

Instructor credibility is significantly related to a variety of student outcomes. Instructors who are seen as credible are more likely to have students who are more motivated to learn [19], have increased cognitive learning [20], and are more respectful of the instructor [21]. Most relevant here, students who perceive their instructor as more credible also report greater affective learning; that is, they are more satisfied with their learning experience and report a more positive attitude toward the content and course [20]. In one meta-analytic approach [22], credibility accounted for approximately 20% of the variance in outcomes such as affect. Typically, all three facets are included in correlational studies of instructor credibility.

Owing to over three decades of research on instructor credibility, we expect to see a similar pattern and thus propose the following hypothesis about instructor credibility and course satisfaction for our IE physics courses:

H3: Each of the three facets of instructor credibility is significantly correlated with course satisfaction.

C. Facework

Although instructor credibility is in the eye of the beholder, there may be particular instructor behaviors that fuel such perceptions. A more recent contribution to this literature is the concept of facework. Deriving from research on politeness and intercultural communication, face refers to the desired self-image that one seeks to portray in interactions [23]. For example, a new teaching assistant in her first recitation section may want to be construed as competent and in control while in the classroom. Facework, then, refers to specific behaviors taken to protect one’s own or others’ face. In the teaching assistant example, a supervising faculty member may preface criticism in such a way to recognize the assistant’s expertise and abilities. Likewise, instructors may frame comments to students to help them feel competent and successful. Typically, facework centers on one or more foci: solidarity, approbation, and tact. Solidarity refers to strategies that reinforce commonalities, in-group identity, and otherwise highlight the need to be included. Approbation refers to comments and actions that are intended to avoid potential disagreement or doubt. Tact strategies refer to those behaviors that are indirect and tentative as a way to show respect for the other’s autonomy.

Classroom communication that engages in effective facework includes interactions that encourage students’ investment in the class, maintain a safe climate for risk taking, respect students’ contributions, encourage higher order thinking, encourage autonomous thought,

build students' abilities to apply course content outside of class, and focus on improvement [24].

Instructor use of facework is positively correlated to positive student outcomes such as increased motivation and preference for productive learning approaches [25]. Likewise, instructors who engaged in more facework were seen as more fair compared to instructors who engaged in fewer behaviors [26]. Instructors who engage in behaviors to mitigate potential threats to face are perceived as significantly more credible than their counterparts [27]. Though not framed through the lens of facework, a recent study found that when physics students felt that their instructors were more supportive of their autonomy, they generally had an overall more positive experience in their physics classes. This result was used to argue that course design should include opportunities for autonomy-supportive interactions [28].

We propose that facework is a correlate of satisfaction:

H4: Instructor facework is significantly related to course satisfaction.

Presuming instructor credibility and facework do have the anticipated relationships with course satisfaction, we propose creating a new model for student satisfaction. Current theorizing in communication research positions credibility in two ways: as an outcome of teacher actions or attitudes, and as a predictor of student outcomes [22]. Thus, we propose that credibility acts as an interim step between activities and facework and satisfaction:

H5: Activities and facework together predict credibility, which in turn predicts satisfaction.

Finally, we propose modeling our variables in a way that predicts satisfaction, beginning with our previously used control for anticipated final grades [17]

RQ6: Controlling for anticipated final grades, how do expectations, experiences, attitude, facework, and credibility predict satisfaction?

This final hypothesis provides the opportunity to better understand the underlying structure of variables as they relate to predicting satisfaction.

III. METHODS

A. Environment

This study focused on introductory algebra-based physics courses at a regional Appalachian University. The classes were IE and similar to SCALE-UP [1]. As in SCALE-UP, the classroom had round tables with nine chairs each, and the daily routine of the classroom emphasized collaborative active learning rather than lecture. However, there were substantial differences between

SCALE-UP and the IE classes studied here. Notably missing were key features like name tags, individual white boards, and computers (except for data collection during laboratory activities) [29]. Either individually or in groups, students often solved problems during class on paper work sheets that were developed in-house. When groups were used, each instructor did so differently: some randomly assigned tables and allowed *ad hoc* groups to form, while others specifically assigned and reassigned groups throughout the semester. Class size was limited to 40 students, and each lead instructor had an assistant (often a part-time faculty member, but sometimes a tenured colleague). Thus, the student-teacher ratio in the classroom was more favorable than typical of SCALE-UP, and many students received one-on-one instruction during class time.

A total of eight sections of the introductory algebra-based physics course were studied over two academic semesters. Each section was led by one of five tenured faculty members (two of whom are female), and no instructor taught more than two sections. We had no instructional responsibility in any of the sections.

During their second week and fourteenth week of instruction, the lead instructors invited their students to respond to online surveys for extra credit. Course instructors did not see the data. Additionally, one of us (J.G.) visited each classroom to ask for access to course artifacts and, in the Fall 2015 semester, to recruit interview volunteers. Data from the different sections were then pooled; it was not our intention to use collected data to compare instructors' approaches. Instead, we were looking for patterns within the students' experience that could explain their variance in satisfaction in the course.

B. Instruments

Students responded to online surveys early (week 2) and late (week 14) in the semester. Those surveys contained multiple instruments, which are further explained below. Additionally, basic demographic information including students' gender and expected final grade in the course were collected at week 14.

Response rates varied by section but ranged from approximately 50% to 80%. Out of a population of about 300 students, we had 259 unique respondents; 173 completed the satisfaction instrument, and 147 completed the PEVA at both week 2 and week 14. Likewise, 140 completed the credibility and facework instruments at both times. Contributing to the low response rate were the online nature of the survey, the variability in instructor support in announcing and reminding students about the surveys, and the timing: during week 14, the students tended to be preparing for final projects and exams.

1. Course satisfaction

We provided a set of eight questions, two of which were reverse coded (see Table I), to measure each student's

TABLE I. Items in the course satisfaction measure, on the week 14 survey. Starred items are reverse coded.

Satisfaction items
1. The way the course was structured supported my learning
2. I looked forward to coming to class each day
3. I was very disappointed with the instruction in this course*
4. I was pleased by the variety of activities we did in class
5. I would recommend this class to a friend
6. I did not enjoy how we spent our time in class*
7. I feel like I learned very much in this class
8. Overall, I am very satisfied with this course

satisfaction in the course. We envisioned “satisfaction” as encompassing both strong instruction and a pleasurable classroom experience; the questions were designed with those goals in mind. We began with affect-related questions in the original PEVA [30] and incorporated questions of the type found on end-of-semester student evaluations. We fine-tuned the scale in subsequent work [31], removing two questions that compared their experience to lecture and rephrasing others for clarity [17]. Responses to the measure were reliable ($N = 173$; $\alpha = 0.91$), and they were summed to make a “satisfaction score,” which we used as the target construct for this study, for each participant. The satisfaction score ranged from 0 (very negative) to 48 (very positive), with a mean score of 26. We treated the distribution as continuous, but we did not assume normality because it failed the Shapiro-Wilk test ($p = 0.022$).

2. PEVA

During week 2, students were asked how often they expected certain activities to occur in class on a scale from 1 (“almost never”) to 7 (“almost constantly”). During the week 14 survey, they were asked to reflect on how often those activities occurred in class, responding with the same scale. On both surveys, they were also asked how they felt about those activities, choosing one of three options from a drop-down menu: “it is enjoyable”; “I don’t enjoy it but I don’t mind it”; and “I do not enjoy it and I don’t like when we do it.” In follow-up interviews, students reported that they interpreted those responses as approximately positive, neutral, and negative, respectively. We subsequently refer to these responses as “affect.” Additionally, students were asked to select all activities that they believed were “essential for you in this course”; they were permitted to select any or all of the activities.

The list of 14 activities included in the survey were based on those in the original PEVA [30]. However, instead of treating the activities as separate from one another as in previous works, we explored the idea that those activities were manifestations of underlying components or general behaviors that were typical of an IE physics class. After modifying the measure somewhat (see Refs. [32,33], and [17] for details on revisions to the PEVA), we used a factor

analysis to determine that the frequency with which students reported experiencing the activities clustered into four factors. By looking at the activities that principally loaded onto each factor, we identified these factors as related to instructor support, instructor-led class time, working with classmates, and working individually. Table II shows the activities and loadings onto each factor for our collected data, using a Varimax rotation. Only loadings of 0.4 and higher are reported.

We postulate that for any IE physics class, those components should emerge, although the loadings may be considerably different, depending on how the class time is spent. The two components related to the instructor represent the two roles that the instructor plays in an active-learning classroom. On the one hand, the instructor typically plans class time and chooses what material to cover; on the other hand, he encourages students, helps them solve problems individually and in groups, and provides feedback in real time as the students struggle with the content. Theoretically, students may perceive that their instructor interacts with the class in those two ways independently; that is, some instructors may spend a considerable amount of class time explaining difficult concepts in lecture but less time directly correcting students’ efforts.

In terms of communication, the four components may be interpreted as “instructor to students as a class,” “instructor with students individually,” “students with each other,” and “each student with self.” It is possible that some active learning classes could also have a fifth component, “students to class as a whole,” represented by item No. 14, but because students here reported that presenting their work to the class was fairly infrequent, we did not observe evidence of such a component.

Most items cleanly load onto a single component, but some items represent more than one component to some substantial degree. For example, “discussing physics concepts directly with the instructor” seems to depend on both how supportive the instructor is and, to a somewhat lesser degree, how she chooses to lead class time. Similarly, “asking questions that I have” depends on both how supportive the instructor is and how often students work with classmates, which follows since the statement is open ended regarding of whom the questions are being asked. Since we cannot conclude that the items on the survey necessarily map onto a single component of class time, we combined the items using a component score coefficient matrix resulting from a least-squares regression analysis, a process that ensured that the factor scores were maximally correlated to the estimated factor [34]. In much of the subsequent analysis, we collapsed the PEVA Experiences measure into four variables corresponding to those factors.

At this point we should emphasize that all measures in this study are explicitly about student *perceptions*, not about external observations. For example, we did not ask

TABLE II. Items in the PEVA and loadings onto each of four theoretical components.

PEVA items	Instructor support	Instructor class time	Work with classmates	Individual effort
1. Solving problems individually				0.83
2. Solving problems in groups			0.83	
3. The instructor introducing new material to the whole class in lecture		0.75		
4. Discussing physics concepts directly with the instructor	0.67	0.43		
5. Discussing physics concepts with classmates			0.74	
6. Doing structured laboratory activities			0.63	
7. Asking questions that I have	0.64		0.41	
8. The instructor demonstrating how to solve problems to the whole class		0.76		
9. Getting individualized feedback from the instructor	0.82			
10. Getting encouragement and support from the instructor	0.65	0.48		
11. The instructor leading whole-class discussion about conceptual questions		0.77		
12. Working to understand physical concepts by myself				0.82
13. The instructor providing guidance for how to do activities	0.47	0.61		
14. Presenting my problem solutions to the class	0.71			

students to quantify the time spent in class doing particular activities; rather, we allowed them to report a qualitative score that represented how often they *felt* like they were doing that activity. Asking questions in this way introduces an additional layer between instructor actions (e.g., spending more time lecturing) and student perceptions (e.g., *feeling* like the instructor spent more time lecturing), a point that will be explored more in Sec. V. However, by keeping all questions about student perceptions, we were able to make consistent claims: after all, student satisfaction is a *perceptive* construct, not an empirical one. Therefore, claims within the framework of student perception have more validity, at the expense of their implications being potentially more unclear.

3. Credibility

To measure students' perception of instructor credibility, we used an established measure from the communication literature [18], changing some words to focus the respondent's attention on the expertise of the instructor as a *teacher* rather than as a *physicist*. We chose to include this emphasis to highlight that we were most interested in students' perceptions of their instructors as physics *teachers*, not as physicists. In fact, this concern bore out, as explored in Sec. VA 3.

The items included in the measure are listed in Table III. Students selected a number on a sliding scale between the two possible descriptors of the instructor. In keeping with the literature, we treated instructor credibility as consisting of three facets (competence, goodwill, and trustworthiness) that represent unique but highly related aspects of instructor credibility. In Table III, we identify which facet each item is measuring. Before computing scores, we reversed the

coding of some items so that a higher number (right item) always indicated higher credibility. We computed a score for each facet by averaging each participant's responses to items within that facet. We emphasize that it is not appropriate to combine these facets as if they were a basis

TABLE III. Items in the credibility instrument, on both the week 2 and week 14 surveys. Students selected a number ranging from 1 (the left item) to 7 (the right item) identifying where on the spectrum they would place their instructor. Each item is identified as belonging to the competence, goodwill, or trustworthiness facet. Item ordering was scrambled for presentation. Starred items are reverse coded.

Credibility items	Facet
Intelligent: Unintelligent*	Competence
Untrained (teacher): Trained (teacher)	Competence
Inexpert (teacher): Expert (teacher)	Competence
Informed: Uninformed*	Competence
Incompetent (teacher): Competent (teacher)	Competence
Bright: Stupid*	Competence
Cares about me: Doesn't care about me*	Goodwill
Has my interests at heart: Doesn't have my interests at heart*	Goodwill
Self-centered: Not self-centered	Goodwill
Concerned with me: Not concerned with me*	Goodwill
Insensitive: Sensitive	Goodwill
Not understanding: Understanding	Goodwill
Honest: Dishonest*	Trustworthiness
Untrustworthy: Trustworthy	Trustworthiness
Honorable: Dishonorable*	Trustworthiness
Moral: Immoral*	Trustworthiness
Unethical: Ethical	Trustworthiness
Phony: Genuine	Trustworthiness

set of variables; rather, each facet was highly correlated with the others at the $p < 0.001$ level: competence was correlated with goodwill ($r = 0.74$); competence was correlated with trustworthiness ($r = 0.84$); and goodwill was correlated with trustworthiness ($r = 0.80$). Each facet was reliable: alpha values, whether measured at week 2 or week 14, were all greater than 0.87.

Responses to each of the facets were considered to be continuous in the analysis; however, the distributions of responses were not nearly normal. For all three facets, distributions were skewed toward positive perceptions of credibility; in the aggregate, students felt that their instructors were generally credible.

4. Facework

On the week 14 survey, students were asked to reflect upon their direct interactions with the lead instructor in their physics class by responding to an existing facework scale [25]. Based on theoretical considerations in the literature, we expected three components to emerge (see Sec. II C). However, a factor analysis quickly revealed that the items loaded onto two factors distinct from those suggested by the literature (see Table IV).

Rather than items clustering based on categories of solidarity, approbation, and tact, the items clustered more generally based on whether the statement highlighted instances where instructors acted to protect students' face or instances where instructors acted to harm students' face, whether intentional or not. Unlike the PEVA instrument, there were no instances where items loaded onto both factors. Thus, we used a less complicated procedure, averaging the scores of the items that loaded onto each factor with at least a score of 0.6. One item, "Leave you free to choose how to proceed with an activity," did not load onto either factor; upon reflection, the structure of the course was such that the activities in class were quite constrained, with detailed instructions. Therefore, we have no reason to believe that students would ever feel free to choose how to proceed in an activity regardless of interactions with their instructor. This item was removed from the analysis.

It is worth noting that the factors were precisely determined by whether the statement was written in a way that sounded positive or negative. Students in interviews noted the difference in tone between the items. One student (Deborah) claimed that while she understood that her instructor would never see the responses, she worried about her responses coming off as too harsh for an instructor that she perceived to be essentially friendly and helpful. Indeed, from that perspective, it is easy to see how students may have cushioned their responses to protect their instructors [35]. Consequently, aggregate responses to the "harming face" component are skewed positive (that is, toward instructors not harming students'

TABLE IV. Items in the facework instrument, on the week 14 survey. Students responded to the prompt, "Think about the direct interactions you had with the lead instructor in your physics class (e.g., during individual or whole-class discussions). With those in mind, please respond to all of the questions below (even if some seem redundant) indicating the degree to which each describes your interactions." Responses ranged from 0 (not at all) to 6 (very much). Each item is identified as relating to either "protecting" or "harming" face.

Facework items	Variable
1. Work to avoid making you look bad	Protecting
2. Leave you free to choose how to proceed with an activity	(Removed)
3. Make sure that s/he doesn't cast you in a negative light	Protecting
4. Seem unconcerned about your feelings	Harming
5. Let you know that he/she thinks highly of you	Protecting
6. Leave you without a choice of how to respond to questions	Harming
7. Make an effort to understand your needs	Protecting
8. Sound like s/he disapproves of you	Harming
9. Make you feel pushed into agreeing with his/her suggestions	Harming
10. Show that he/she cares about your learning experience	Protecting
11. Express blunt criticism of your ideas or abilities in class	Harming
12. Make you feel like you can choose how to respond to instructors' ideas	Protecting
13. Make you feel like an important member of the class	Protecting
14. Make it hard for you to propose your own ideas	Harming
15. Seems attentive to you as an individual	Protecting

face), while responses to the "protecting face" component display more variation and are more nearly normal.

C. Validation interviews

Late in the semester, one of us (J.G.) visited each of the four sections to recruit students for voluntary interviews. Initially, more than 12 students expressed interest in participating, but due to timing challenges at the semester's end, interviews were scheduled with five students who represented three of the sections. One of those students also brought written comments from a classmate, which were used as part of the interview conversation.

The interviews lasted approximately 20 min to 1 h. We asked questions to ensure that the students understood the survey and that they could provide justification from instances in class for their claims. For example, students were asked to what degree they felt their instructor worked to help protect students' self-image, and if they could provide examples (or counterexamples). They were also shown the facework questions from the survey and asked if

they could understand them and if they believed that those questions adequately assessed how they felt about their instructor's interactions with respect to their face.

Because of the voluntary nature of these interviews and the small number that were completed, we used the interview data only as evidential support for what we saw in the quantitative data, not as examples of typical experiences. It is important to note that the comments made by students tended to focus on frustrations and disappointments; they did not generally expound on comments related to satisfactory parts of the course. These interviews should not be seen as representative of all students; indeed, the average satisfaction score of all students was 26 out of 48, indicating that students were overall modestly satisfied (or at worst, neutral) about the course. Nonetheless, the interviews provide an additional layer of insight that helps contextualize the data.

IV. RESULTS

A. RQ1: Expectations and experiences

To answer RQ1, which asked about students' expectations, experiences, and attitudes toward different activities in an IE physics course, we explored students' responses to the PEVA. Table V shows aggregate responses to the PEVA at week 2. The question of consequences resulting from differences between how each individual instructor prepares students' expectations for the semester is relevant but beyond the scope of this project [36].

Overall, students expected to encounter most of the activities fairly frequently in class, including having the instructor introduce new material in class, which is typically deemphasized in IE classrooms. Students felt that they

would present their problem solutions to the class less frequently than the other activities. Additionally, students generally reported disliking presenting their problem solutions to the class. Another item, asking questions, had a fairly neutral affect, as students apparently disliked the activity while understanding its necessity. Both of those activities represent potentially face threatening situations: in both cases, students are vulnerable to criticism. In contrast, the activities with highest affect are either those that do not create face threatening situations or those that actively mitigate potential face threat (such as getting encouragement and support from the instructor or solving problems in groups).

Two items, solving problems individually and working to understand physical concepts by oneself, demonstrated relatively low affect. Two weeks into the semester, most students did not enjoy working on physics individually, although perhaps many resigned themselves to doing it.

Table VI shows how often students reported doing the same activities in class later in the semester (week 14). Overall, students reported that most of the activities happened about as frequently as they had expected. A few activities were reported as happening somewhat less, and those activities (items No. 4, 8, 9, 11, and 13) involved the instructor and had been associated with a generally positive affect at the start of the semester.

Every activity exhibited a negative affective shift. That is, on every item, the proportion of respondents claiming that they liked the activity decreased while simultaneously the proportion claiming they did not like doing it increased. In some cases, the differences were negligible and not statistically significant, but the overall trend should be noted. Perhaps the biggest decrease in affect was on doing

TABLE V. Students ($N = 205$) reported their expectations at week 2 about how frequently certain activities will occur in the classroom (a rating of 7 means, almost constantly; a rating of 1 means, almost never), their attitude toward those activities (1 means, it is enjoyable; 2 means, I don't enjoy it but I don't mind it; 3 means, I do not enjoy it and I don't like when we do it), and whether they feel those activities are essential (reported is the percent of respondents to the attitude survey selecting the activity as essential).

PEVA items	Expectations median	Attitude			Essential?
		1	2	3	
1. Solving problems individually	5	35%	57%	8%	81%
2. Solving problems in groups	6	72%	24%	4%	81%
3. The instructor introducing new material to the whole class in lecture	6	62%	37%	1%	83%
4. Discussing physics concepts directly with the instructor	5	60%	38%	2%	61%
5. Discussing physics concepts with classmates	6	68%	29%	3%	70%
6. Doing structured laboratory activities	6	69%	25%	6%	78%
7. Asking questions that I have	5	41%	55%	4%	74%
8. The instructor demonstrating how to solve problems to the whole class	6	85%	14%	1%	89%
9. Getting individualized feedback from the instructor	5	74%	25%	<1%	64%
10. Getting encouragement and support from the instructor	5	85%	15%	<1%	55%
11. The instructor leading whole-class discussion about conceptual questions	6	68%	30%	1%	72%
12. Working to understand physical concepts by yourself	5	35%	52%	13%	68%
13. The instructor providing guidance for how to do activities	6	80%	20%	0%	84%
14. Presenting my problem solutions to the class	4	12%	49%	40%	17%

TABLE VI. Students ($N = 176$) reported their experiences at week 14 about how frequently certain activities occurred in the classroom (a rating of 7 means, almost constantly; a rating of 1 means, almost never), their attitude toward those activities (1 means, it is enjoyable; 2 means, I don't enjoy it but I don't mind it; 3 means, I do not enjoy it and I don't like when we do it), and whether they feel those activities are essential (reported is the percent of respondents to the attitude survey selecting the activity as essential).

PEVA items	Experiences median	Attitude			Essential?
		1	2	3	
1. Solving problems individually	5	27%	57%	15%	73%
2. Solving problems in groups	6	65%	28%	7%	78%
3. The instructor introducing new material to the whole class in lecture	6	55%	43%	2%	79%
4. Discussing physics concepts directly with the instructor	4	48%	47%	6%	55%
5. Discussing physics concepts with classmates	6	60%	36%	5%	59%
6. Doing structured laboratory activities	6	41%	40%	19%	49%
7. Asking questions that I have	5	39%	55%	6%	69%
8. The instructor demonstrating how to solve problems to the whole class	5	77%	20%	2%	86%
9. Getting individualized feedback from the instructor	4	69%	30%	1%	64%
10. Getting encouragement and support from the instructor	5	77%	23%	<1%	47%
11. The instructor leading whole-class discussion about conceptual questions	5	63%	34%	3%	72%
12. Working to understand physical concepts by yourself	5	26%	55%	20%	60%
13. The instructor providing guidance for how to do activities	5	70%	26%	3%	72%
14. Presenting my problem solutions to the class	4	11%	45%	43%	12%

structured laboratory activities; while students generally enjoyed them at the start of the semester, their overall opinion was more mixed toward the end of the semester. The proportion of respondents claiming that laboratory activities were an essential part of the course also decreased. These trends toward negative affect are similar in some ways to other trends noted in the literature of introductory physics classes having an overall negative impact on students' attitudes over the term of the course [37].

B. RQ2: Experiential and affective variables

The second research question attempted to predict course satisfaction from expectations, experiences, and attitudes. Because we expected that students who believed they were earning a higher grade in a course would tend to rate their satisfaction in that course higher [38], we began by controlling for that effect. Thus, our first step was to regress expected grade against course satisfaction, $F(1, 171) = 24.87$, $p < 0.001$. The multiple correlation coefficient was $R = 0.36$, adjusted $R^2 = 0.122$, meaning that approximately 12% of the variance in students' satisfaction was explained by their expected course grade alone.

To build a model from students' expectations, experiences, and affect toward the activities that occurred in class, we first determined the correlations between student satisfaction and each of those variables. Because of the large number of possible correlations, we choose a conservative level of significance, $p < 0.01$, to minimize artifacts.

The frequency with which students expected to receive encouragement and support from their instructor (item

No. 10) early in the semester (week 2) was significantly correlated with their satisfaction at week 14 ($r = 0.25$, $p < 0.005$), but there were no other significant correlations between students' expectations early in the semester and their satisfaction in the course. We also found no correlations between students' affect towards individual activities early in the semester and their satisfaction in the course.

However, we found that how often students experienced each activity by week 14 was correlated with their satisfaction in the course, with three exceptions: No. 1 (individual problem solving), No. 3 (instructor introducing new concepts in class), and No. 12 (working to understand concepts by oneself). Students' affect towards seven of the 14 activities were significantly correlated with course satisfaction (items No. 1, 2, 4, 6, 7, 9, and 12).

Because it is unwieldy and generally uninformative to carry so many variables through further analysis, we considered the underlying component factors in the classroom, as discussed in Sec. III B 2. Students' experiences on three of the four components were positively and significantly related to course satisfaction: the frequency of instructor support ($r = 0.42$, $p < 0.001$), the frequency of instructor-led class activities ($r = 0.33$, $p < 0.001$), and the frequency of interaction with their peers ($r = 0.29$, $p < 0.001$). The amount of individual effort was negatively correlated with course satisfaction ($r = -0.17$, $p < 0.05$), but this result did not meet our conservative threshold for significance.

Similarly, we wished to reduce the number of relevant variables in the affective measure. Because we could not ascribe affect toward the underlying factors directly, we averaged the scores for the items that had a principal loading onto each factor. Consequently, we created four

TABLE VII. Experiential and affective variables regressed against course satisfaction ($N = 173$). The improvement to adjusted R^2 is shown for the addition of each variable. Each variable made a significant ($p < 0.01$) contribution to the model.

Variable added	R	Adjusted R^2	Adj. R^2 change	F change	Std. error
Expected grade	0.356	0.122	0.122	24.87	10.30
Freq. of instructor support	0.553	0.298	0.176	43.85	9.21
Freq. of instructor-led class	0.667	0.435	0.137	42.18	8.26
Affect for group	0.719	0.517	0.071	25.19	7.73
Affect for indiv. effort	0.736	0.528	0.022	9.09	7.55
Freq. of group	0.750	0.547	0.019	8.09	7.39

variables that tell us how students felt about the individual activities that *principally* constituted each factor. When we did so, three of the four variables were positively correlated with student satisfaction: affect towards group activities ($r = 0.36$, $p < 0.001$), individual effort activities ($r = 0.32$, $p < 0.001$), and instructor support activities ($r = 0.28$, $p < 0.001$). The fourth, affect towards instructor-led classtime activities, was not significantly correlated ($r = 0.15$, $p = 0.052$).

Next, to answer RQ2, we selected all variables significantly correlated with satisfaction as possible predictors, using the conservative criterion of significance at $p < 0.01$. We began by entering expected grades again, then entered the other variables in descending order of the correlation with satisfaction. We only entered a variable if it significantly improved the model ($p < 0.01$) after the previous variables were entered. Table VII shows how the model improved with the addition of each variable. Neither affect towards instructor support activities nor student expectations of the frequency with which they would receive encouragement or support were added to the model because those predictors explained no significant additional variance in student satisfaction.

C. H3: Credibility and satisfaction

H3 predicted that course satisfaction would be significantly correlated with each component of instructor credibility. This hypothesis was supported for each component of credibility with significance at the $p < 0.001$ level: competence, $r = 0.53$; goodwill, $r = 0.65$; and trustworthiness, $r = 0.55$.

Additionally, we expected that when we controlled for students' expected course grades, instructor credibility would be a good predictor of course satisfaction. We began with the component most strongly correlated with satisfaction (instructor goodwill). Adding that variable to the model explained an additional 35.5% of the variance in course satisfaction over expected course grades alone [$F(2, 163) = 77.9$, $p < 0.001$], making the total adjusted $R^2 = 0.482$ for the model containing expected course grades and instructor goodwill. Because of the degree to which the facets of credibility were correlated with each

other, we found that adding instructor competence or trustworthiness to the model after adding goodwill did not significantly improve the model.

D. H4: Facework and satisfaction

Our hypothesis H4 predicted that instructor facework would be significantly related to course satisfaction. Using the division into harming and protecting facework that we outlined in Sec. III B 4, this hypothesis was supported. Harming facework was negatively correlated with satisfaction, $r = -0.42$, $p < 0.001$, while protecting facework was positively correlated, $r = 0.63$, $p < 0.001$. As expected, students who reported more encounters of face threatening situations (that is, a higher frequency of the harming experiences) reported lower satisfaction in the course; on the other hand, those who reported that their instructor did more protective facework had higher satisfaction in the course.

As expected, we found that protecting facework was a good predictor of course satisfaction when we controlled for expected course grades [$F(2, 159) = 72.071$, $p < 0.001$]. The total adjusted $R^2 = 0.469$ for the model containing expected course grades and protecting facework; adding in that facework variable explained an additional 33.8% of the variance beyond expected course grades alone. Adding harming facework only marginally improved the model and did not meet our $p < 0.01$ criterion, so that variable was excluded.

E. H5: Credibility as a mediating variable

Our hypothesis H5 addressed the possibility of situating instructor credibility as an interim step toward satisfaction. We predicted that instructor facework would predict credibility (specifically, the *goodwill* facet of credibility, which we found in Sec. IV C to be a good predictor of course satisfaction). We anticipated this relationship from Sec. II C, where we noted that instructors who actively mitigate face threats tend to be perceived as more credible. We have replicated that result; protecting facework is highly correlated with the goodwill facet of credibility ($r = 0.79$, $p < 0.001$), and harming facework is strongly

negatively correlated with the goodwill facet of credibility ($r = -0.56$, $p < 0.001$).

One interpretation of this finding is that instructors can gain credibility through the mechanism of protecting facework, and instructors lose credibility through the mechanism of harming facework. Indeed, a linear regression model of goodwill with both facework variables explains 68% of the variance of goodwill credibility [$F(2, 165) = 179.7$, $p < 0.001$].

We also predicted that student experiences and affect toward in-class activities would also predict credibility, but those variables did not significantly improve the model of goodwill predicted by facework variables alone. Specifically, while the frequency of instructor support and class-led activities were correlated with goodwill credibility ($r = 0.41$, $p < 0.001$ and $r = 0.33$, $p < 0.001$, respectively), those variables each only marginally enhanced the model and failed to meet our $p < 0.01$ criterion and were thus subsequently dropped. Because those variables were correlated with protecting facework ($r = 0.38$, $p < 0.001$ and $r = 0.28$, $p < 0.001$, respectively), it appears that the frequency of instructor support and instructor-led class time variables had mostly an *indirect* effect on credibility, perhaps mediated by facework. That is, students who experienced more instructor support and instructor-led class time likely also had (or observed) more instances where potentially face-threatening situations were mitigated or defused by their instructor. That instructor's facework then predicted credibility, rather than the in-class activities themselves.

We might further ask whether the only implication of facework is the gain or detriment to goodwill credibility; that is, does facework explain any additional variance in course satisfaction beyond credibility, or is its effect entirely mediated through the credibility construct? To test this question, we regressed both protecting and harming facework against course satisfaction while controlling for both expected course grade (as before) and goodwill credibility. We found that protecting facework did explain a significant, albeit small, portion of the variance beyond that explained by goodwill (see Table VIII). Thus, we conclude that while protecting facework actions seem to

improve an instructor's goodwill credibility, those actions must also have some other positive effect on course satisfaction. Consequently, our hypothesis H5 was only partially supported; the implications of this finding are explored in Sec. V.

F. RQ6: Building a predictive model of course satisfaction

Our final research question, RQ6, asks how all of our variables (expectations, experiences, attitude, facework, and credibility) together predict course satisfaction. To answer this research question, we added variables into the model in order of strength of correlation (after controlling for expected final grade). Thus, the first step was to include instructor goodwill and protecting facework, as at the end of Sec. IV E and shown in the first three rows of Table VIII.

Next, because we noted in Sec. IV B that the frequency with which students experienced both the instructor support and the instructor-led class time components of the course were both correlated with course satisfaction, we added those into the model. Apparently, while those variables only indirectly predicted credibility, they held a direct relationship with student satisfaction; when we controlled for course grade, instructor goodwill, and protecting facework, those variables explained an additional 6.8% of the variance of course satisfaction, a modest improvement over the model without those variables (see Table VIII).

We also found in Sec. IV B that two affective variables, related to working in groups and working individually, were significant predictors of course satisfaction. We have no theoretical reason to believe that those variables should be correlated to our other variables, so we simply postulate that they, too, should predict additional variance in student satisfaction. Indeed, we see that they both significantly (albeit modestly) improve the model, explaining an additional 5.8% of the variance (see Table VIII).

However, with the inclusion of the affective variables, instructor goodwill credibility no longer contributes significantly to the model (see Table IX). In other words, we no longer gain any advantage by controlling for students' perception of the instructor as caring; instead, the variance associated with goodwill occurs in other

TABLE VIII. Facework, experiential, and affective variables regressed against course satisfaction ($N = 161$). The improvement to adjusted R^2 is shown for the addition of each variable. Each variable made a significant ($p < 0.01$) contribution to the model.

Variable added	R	Adjusted R^2	Adj. R^2 change	F change	Std. error
Expected grade	0.370	0.131	0.131	25.33	10.39
Instructor goodwill	0.698	0.480	0.349	108.34	8.04
Protecting facework	0.722	0.512	0.032	11.41	7.79
Freq. of instructor support	0.744	0.543	0.031	11.55	7.54
Freq. of instructor-led class	0.770	0.580	0.037	15.12	7.22
Affect for group	0.798	0.624	0.044	18.92	6.84
Affect for indiv. effort	0.808	0.638	0.014	7.07	6.71

TABLE IX. Regression analysis for course satisfaction, with instructor goodwill removed ($N = 161$). The improvement to adjusted R^2 is shown for the addition of each variable.

Variable added	R	Adjusted R^2	Adj. R^2 change	F change	Std. error
Expected grade	0.370	0.131	0.131	25.33	10.39
Protecting facework	0.690	0.469	0.338	102.71	8.13
Freq. of instructor support	0.724	0.515	0.046	16.12	7.77
Freq. of instructor-led class	0.760	0.567	0.052	20.16	7.33
Affect for group	0.794	0.618	0.051	21.70	6.89
Affect for indiv. effort	0.804	0.633	0.015	7.54	6.75

places. Specifically, the affective variables are correlated with goodwill credibility (affect toward group work: $r = 0.29$, $p < 0.001$; affect toward individual work: $r = 0.20$, $p < 0.001$). Thus, while goodwill credibility itself is easily measured and a strong predictor of course satisfaction, its contribution to that target construct is largely negligible when expected grade, protecting facework, experiential factors, and affect toward experiential factors are all considered.

V. DISCUSSION

When building linear regression models, additional terms in the model both explain more of the target construct's variance and increase the model's complexity. To address these opposing pressures, we highlight two models that shine in contrasting ways. First, we describe the simple model from Sec. IV C that provides substantial explanatory power by utilizing the intuitive construct of instructor credibility. Next, we discuss a more complex model from Sec. IV F that explains more of the variance in course satisfaction by emphasizing instructor actions associated with facework. Finally, we address limitations to the study and discuss future directions for exploration.

A. Instructor credibility

In Sec. IV C, we see that, when controlling for students' expected final grade in their course, the instructor goodwill facet of credibility (which we use as a proxy for credibility itself) is a substantial predictor of course satisfaction, explaining an additional 33.5% of the variance. Consequently, an intuitive conclusion is that anything that an instructor might do to increase their credibility in the eyes of the students is likely to also increase their students' satisfaction with the course. While this theoretical approach is simplistic (perhaps overly so), it has the advantage of plentiful support in the literature. Credibility is a well-established construct, and numerous studies have demonstrated its usefulness in understanding interpersonal interactions (Sec. II B). In order to better understand students' assessment of instructor credibility in these IE classes, we sought additional support from student interviews.

First, we note that in interviews, discussions about the individual facets tended to blur together (for example,

while discussing trustworthiness, a student might talk about how a certain violation of classroom norms demonstrated a lack of caring). This blurring emphasizes how the facets are not independent component factors of credibility, but different aspects that are related in complex ways [18]. Thus, our use of a single facet as a proxy for all of credibility is a reasonable approach, but the other facets should not be entirely ignored; all three facets of credibility were highly correlated with course satisfaction, and the literature suggests that the three facets are so deeply entangled that they are unlikely to change independently of one another.

1. Goodwill

We asked students in interviews about ways in which they felt instructors demonstrated caring (or lack thereof) in order to get a sense of how students judged the goodwill facet of credibility. William [39], a highly successful student, said that he felt like his instructor (whom we call "A") cared little to none about the success of the students in the class. He further detailed that A did not know anyone's name in the class. A also introduced obstacles, such as not posting in-class slides unless at least one student emailed such a request. Additionally, A provided no clear way for students to calculate their own class grades. William also felt like the instructor delighted in asking questions that were of the form, "when is X possible?" to which the response was "X is not possible," just to see the students struggle.

Lisa, Charlie, and Elise highlighted challenges they experienced in instructor B's class. On day one, instructor B came across as caring and understanding, but as the semester went on, the students perceived an inconsistent tone. For example, Lisa reported two interactions that made her feel that B was acting condescending: in an email, B told Lisa that a particular homework problem was just like an in-class example and that she should be able to do it; during an in-class exchange, B made a comment about making good use of class time that Lisa took as an insult. Charlie stated that B did work toward redirecting incorrect answers in class so as to try to protect the student responding, and that such efforts did help him seem more caring. However, Lisa pointed out that despite those efforts,

“when I saw everybody in the classroom looking at me and I got the wrong answer, I did not like the feeling that I got, and knowing that I was struggling as well didn’t help at all.” The three students also felt that assigning students to heterogeneous groups [40] did not help learning, but just made the “low-performing” person felt like a burden on the rest of the group; this situation promoted embarrassment while making it difficult to ask for help.

Meanwhile, Deborah discussed an overall positive experience with instructor *C*. Specifically, she noted that *C* had a genuine enthusiasm for the subject, and she could tell that *C* was authentically interested in getting students to experience that enthusiasm as well. Additionally, *C* worked at creating a safe environment where students could ask questions. Deborah mentioned that she had personal issues interfering with her ability to get schoolwork done on time and approached *C* about them; *C* was receptive and willing to work with her on making suitable arrangements. However, Deborah tempered those positive aspects with one specific criticism: Deborah did not think that *C* knew the students’ names, which Deborah interpreted as *C* lacking a personal connection with the class, a detached approach to interactions with groups (for example, not telling a group whether they correctly solved a problem), and a difficulty in providing personalized feedback to students.

These examples demonstrate the strong role that goodwill plays in students’ perceptions. As our quantitative data bore out, goodwill is at the forefront for creating a satisfactory experience for students in the classroom.

2. Trustworthiness

Trustworthiness refers to how ethical, genuine, and honest the instructor appears. William stated that he did not feel his instructor was trustworthy. As an example, William cited that *A* promised a certain problem would be on the exam but then proceeded to change it so much that the students no longer considered it to be similar. Additionally, William pointed out that certain homework problems required calculus (not a prerequisite for the course) and that it felt like *A* was trying to trick the students on questions. In general, he claimed that instructor *A* was not following the “confines of the course.” Similarly, Charlie said that he felt that his instructor (*B*) reneged on a first-day statement that process and conceptual understanding were more important than mathematics: when Charlie visited *B* to discuss his second exam, *B* stated that Charlie had no conceptual errors and would have earned a higher grade had he not made certain math mistakes; if conceptual understanding was so much more important, why had math mistakes reduced his grade so much? Deborah claimed that her instructor (*C*) would sometimes promise to extend online homework assignments but then forget to do so.

Those examples all reflect ways an instructor’s trustworthiness may be compromised, and it is easy to see how they also contribute to dissatisfaction in the classroom. Trustworthiness is not an abstract quality; in the daily interactions of an IE classroom, the trustworthiness of an instructor is tested. If students do not feel that their instructor can be trusted, they will be hesitant to engage with the instructor. When students then refuse to interact with the instructor, the instructor may interpret those actions as resistance to the pedagogy.

3. Competence

Competence was perhaps least understood by the students. In his interview, William confused being an expert *physicist*, which he believed *A* to be, with being an expert *teacher*, which he did not believe *A* to be. He then provided the example of his high school teacher, whom he felt was a more competent teacher despite less expertise in the field of physics itself. This confusion may raise some questions about the validity of the competence scale, but that scale was not used in the analysis.

Although there were not many comments specifically about competence in the interviews, Lisa noted that *B* had some expertise teaching, such as *B* relating the physics in the classroom to real-world settings. Deborah pointed out that *C* explained things well and took the time to summarize in-class activities after they were completed. Those examples show how perceived competence as a teacher might positively affect student satisfaction: if the instructor is clear and well connected to the real world, the content may make more sense and have more relevance. However, the competence facet seems to be ambiguous, and it is unclear exactly what makes instructors appear competent to students.

4. Limitations of credibility

While credibility refers in a broad sense of how trustworthy, caring, and competent the instructor is perceived to be, there is no mechanism for change within the construct itself; that is, in order to understand how to be perceived as more competent, we must understand first what makes students believe instructors appear competent and then test whether affecting those variables can in turn affect credibility. In this sense, the construct of credibility is fairly limited. However, credibility is also easy for instructors to grasp intuitively, and it may be a useful construct for that purpose: although we have not secured a causal relationship, students who are more satisfied tend to find their instructors more credible. Thus, a seemingly sensible place for instructors to begin thinking about their interactions in IE physics classes is to interact with students in ways that boost their credibility, by intentionally acting trustworthy, caring, and competent.

B. Facework in the classroom

In contrast to the previous simple model, we can think of satisfaction in the classroom as being described by a more nuanced, complex model that highlights particular actions by the instructor that might directly impact student satisfaction. Table IX shows that students were more satisfied with their IE physics course when they felt that they were earning a high grade, their instructor was perceived to be doing protecting facework, they were being actively supported by their instructor, they were spending more class time on instructor-led activities, they had a positive attitude toward working with their classmates, and they had a positive attitude about putting forth individual effort. Most of those factors are directly related to the instructor's decisions and efforts in the classroom, and the effect of protecting facework is clearly the most powerful.

Because the construct of facework is made up of generalized behaviors (see Table IV), we can provide direct advice to instructors regarding how to manage their interpersonal interactions with their students. To illustrate, we can turn to student accounts of instructor behaviors to identify particular interactions that impacted course satisfaction. Notably, the same interactions that we explored in Sec. VB with respect to instructor goodwill can be reassessed with the additional layer of facework. When students cite examples of instructors not caring, they describe particular actions made by the instructor; those actions can be mapped onto the construct of facework and provide insight about a causal mechanism for satisfaction.

Recall William's observations about instructor A, who did not provide emotional support or protect students from potentially face threatening situations in class. A made no visible effort to try to understand the particular needs of the students (item No. 7; see Table IV). Instead, from William's perspective, A even seemed to *create* threatening situations, leaving the students uncomfortable and without a legitimate option to express their ideas or respond in an open way (item No. 12). Consequently, the classroom environment fostered one-sided vulnerability and face-degrading interactions.

Likewise, students' experiences with instructor B can be understood in terms of facework. Lisa, Charlie, and Elise stated that B provided appropriate framing for the start of the semester, but they felt that instructor apparently fell short of implementing those ideals. Instead, they reported interactions that made the students look bad (item No. 1). Condescending comments directly damaged Lisa's face, and being singled out as a poor performer (even unintentionally) reinforced face threat regarding the competence of the students by casting them in a negative light (item No. 3). The students repeatedly reported a level of inconsistency in B's tone, discussing that his level of caring seemed to go up and down throughout the semester; that inconsistency led to an overall distrust in whether B was actually trying to understand where the students were coming from (item No. 7).

According to Deborah's account, C demonstrated caring for students by working to prevent them from looking bad (item No. 1), making an effort to understand them (item No. 7), and demonstrating a concern for students' learning experience (item No. 10). However, the lack of a personal connection implies a certain detachment that might be interpreted as C not demonstrating appreciation towards individuals or their contributions to the class (items No. 5, 13, and 15).

We emphasize that William and Deborah noticed when their instructors did not appear to know the names of their students. The students interpreted that lack of personal connection as indicative of a lack of caring (items No. 13 and No. 15). Making an effort to learn the names of students—and demonstrating that the names are known—is a straightforward, actionable way that instructors can indicate to their students that they care about them.

It is worth noting that the instructors may not have been aware of the effect their interactions had, but that lack of awareness is essentially the point: mitigation of face threat does not necessarily result from natural behaviors taken by instructors in the setting of the IE classroom; instead, we believe that such instructors must attend to the new forms of interpersonal communication that they undertake with their students. Specifically, *instructors should make additional, directed efforts to mitigate threats to face when interacting with their students*. These actions may be learned from expert IE instructors (from physics or other fields), or they may evolve over time as the instructors respond to their student evaluation feedback. In any case, having the framework of facework should guide efforts of instructors who want to improve both their interactions with their students and those students' satisfaction in their courses.

These two models, credibility and facework, yield similar and complementary understandings. Students' assessment of instructor credibility seems to impact satisfaction with the course, suggesting that it behooves instructors to build credibility whenever possible, with particular attention toward building a sense of caring and goodwill. Facework added to our ability to predict satisfaction and also highlighted specific actions instructors could take. Together, understanding such affective components of the learning process enhances instructors' interactions in the classroom and ultimately students' learning experience.

C. Limitations and future directions

This study was constrained by a number of factors that limit the scope of the conclusions we might otherwise draw. All of the students who participated in the study were in sections of the introductory algebra-based physics course at the same institution. The population of that physics course tends to be fairly diverse in terms of gender and degree program, including representation from chemistry, biology, construction management, and preprofessional programs, but the students tended to be white non-Hispanic students

from Kentucky or neighboring states. Additionally, many of the students were from the socially depressed Appalachian region and as such may have different socio-cultural norms acting upon them than in other regions. This lack of diversity, coupled with a small number of overall respondents, may cast some doubt about the reproducibility of the influence of some variables that are more weakly predictive of student satisfaction. Additionally, the small class size and student to teacher ratio may have amplified the number of interactions that students in IE courses are likely to experience. However, we believe that the overall conclusions of the study, which align with existing literature, are not likely to be influenced by those limitations of student population. Nonetheless, in future studies, we hope to see replication among various class sizes as well as more diverse groups, including other underrepresented minorities and students in different levels of physics classes. These underrepresented groups may particularly benefit from IE because of the unique interpersonal relationships that can be built, and previous research in SCALE-UP has emphasized its success among those groups [1].

We only report on data collected at two times during the semester, and as such we have no sense of the malleability of credibility or the direct causal impact of instructor facework. We would also like to understand the ebb and flow of students' attitudes. Particularly, to what degree can instructors recover from damaged credibility, and how might they do so? Future work could explore those key features so that we can more carefully articulate causes and preventative measures for student pushback, while at the same time emphasizing the positive aspects of student affect to help instructors create powerfully rich classroom experiences. Potentially, results from ongoing observational studies [5]

could yield additional insights into credibility building and facework actions in the classroom.

VI. CONCLUSIONS

We have presented multiple constructs that can explain much of the variability in student satisfaction: activities in the classroom (broken down by component and then by frequency or affect), instructor credibility (broken down by facet), and facework (broken down by efforts to protect or threaten). Many of these constructs are correlated together, but they are conceptually distinct.

For students in our classes, the model that best explained student satisfaction combined information from both the activities in the classroom and facework to provide a nuanced look at possible causes for student pushback. Alternatively, a very simplified model using instructor credibility (specifically, goodwill credibility) may be conceptually useful for faculty looking for quick ways to reduce resistance in IE classes.

Both models were constructed based on the assumption that, unlike non-IE classes, a wide variety of interactions happen in the IE classroom. By exploring those interactions from the perspective of the students, we validate the expectations they bring into the classroom. Based on these expectations, students may perceive certain interactions in the classroom as potentially threatening to their face. How the instructor navigates around or through those situations is strongly correlated with—and may even partially determine—how satisfied students are satisfied in the course.

ACKNOWLEDGMENTS

This project was funded by Eastern Kentucky University.

-
- [1] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project, in *Research-Based Reform of University Physics*, Reviews in PER Vol. 1, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), <http://www.compadre.org/portal/items/detail.cfm?ID=585>.
 - [2] K. Foote, A. Knaub, C. Henderson, M. Dancy, and R. J. Beichner, Enabling and challenging factors in institutional reform: The case of SCALE-UP, *Phys. Rev. Phys. Educ. Res.* **12**, 010103 (2016).
 - [3] J. D. H. Gaffney and J. T. Whitaker, Making the most of your first day of class, *Phys. Teach.* **53**, 137 (2015).
 - [4] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020102 (2007).
 - [5] P. Shekhar, M. Demonbrun, M. Borrego, C. Finelli, M. Prince, C. Henderson, and C. Waters, Development of an observation protocol to study undergraduate engineering student resistance to active learning, *Int. J. Eng. Educ.* **31**, 597 (2015).
 - [6] E. A. West, C. A. Paul, D. Webb, and W. H. Potter, Variation of instructor-student interactions in an introductory interactive physics course, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010109 (2013).
 - [7] S. J. Pollock and N. D. Finkelstein, Sustaining educational reforms in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010110 (2008).
 - [8] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020101 (2009).

- [9] N. Lasry, E. Charles, and C. Whittaker, When teacher-centered instructors are assigned to student-centered classrooms, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010116 (2014).
- [10] G. W. Hitt, A. F. Isakovic, O. Fawwaz, M. S. Bawa'aneh, N. El-Kork, S. Makkiyil, and I. A. Qattan, Secondary implementation of interactive engagement teaching techniques: Choices and challenges in a Gulf Arab context, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020123 (2014).
- [11] J. Pond and J. J. Chini, Exploring student learning profiles in algebra-based studio physics: A person-centered approach, in *2015 PERC Proceedings*, edited by A. D. Churukian, D. L. Jones, and L. Ding (College Park, MD, 2015), pp. 263–266, <http://www.compadre.org/per/items/detail.cfm?ID=13887>.
- [12] J. K. Burgoon, A communication model of personal space violations: Explication and an initial test, *Hum. Commun. Res.* **4**, 129 (1978).
- [13] Y. M. Kalman and S. Rafaeli, Online pauses and silence: Chronemic expectancy violations in written computer-mediated communication, *Communication research* **38**, 54 (2011).
- [14] M. L. Houser, Expectancy violations of instructor communication as predictors of motivation and learning: A comparison of traditional and nontraditional students, *Commun. Quart.* **54**, 331 (2006).
- [15] B. N. Frisby and R. J. Sidelinger, Violating student expectations: Student disclosures and student reactions in the college classroom, *Comm. Stud.* **64**, 241 (2013).
- [16] C. Obermiller, B. Ruppert, and A. Atwood, Instructor credibility across disciplines: Identifying students' differentiated expectations of instructor behaviors, *Business Commun. Quart.* **75**, 153 (2012).
- [17] J. D. H. Gaffney and A. L. Housley Gaffney, How personal effort, student interactions, and instructor support relate to physics student satisfaction, in *2015 PERC Proceedings*, edited by A. D. Churukian, D. L. Jones, and L. Ding (College Park, MD, 2015), pp. 119–122, <http://www.compadre.org/per/items/detail.cfm?ID=13851>.
- [18] J. C. McCroskey and J. J. Teven, Goodwill: A reexamination of the construct and its measurement, *Commun. Monogr.* **66**, 90 (1999).
- [19] A. B. Frymier and C. A. Thompson, *Commun. Educ.* **41**, 388 (1992).
- [20] J. J. Teven and J. C. McCroskey, The relationship of perceived teacher caring with student learning and teacher evaluation, *Commun. Educ.* **46**, 1 (1997).
- [21] A. D. Martinez-Egger and W. G. Powers, Student respect for a teacher: Measurement and relationships to teacher credibility and classroom behavior perceptions, *Hum. Commun. Res.* **10**, 145 (2007).
- [22] A. N. Finn, P. Schrodt, P. L. Witt, N. Elledge, K. A. Jernberg, and L. M. Larson, A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes, *Commun. Educ.* **58**, 516 (2009).
- [23] E. Goffman, On face-work: An analysis of ritual elements in social interaction, *Psychiatry* **18**, 213 (1955).
- [24] J. Kerssen-Griep, Teacher communication activities relevant to student motivation: Classroom facework and instructional communication competence, *Commun. Educ.* **50**, 256 (2001).
- [25] J. Kerssen-Griep, J. A. Hess, and A. R. Trees, Sustaining the desire to learn: Dimensions of perceived instructional facework related to student involvement and motivation to learn, *West. J. Comm.* **67**, 357 (2003).
- [26] J. Kerssen-Griep and P. L. Witt, Instructional feedback ii: How do instructor immediacy cues and facework tactics interact to predict student motivation and fairness perceptions?, *Comm. Stud.* **63**, 498 (2012).
- [27] L. Trad, J. Katt, and A. N. Miller, The effect of face threat mitigation on instructor credibility and student motivation in the absence of instructor nonverbal immediacy, *Commun. Educ.* **63**, 136 (2014).
- [28] N. Hall and D. Webb, Instructors' support of student autonomy in an introductory physics course, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020116 (2014).
- [29] J. D. H. Gaffney, E. Richards, M. B. Kustusch, L. Ding, and R. J. Beichner, Scaling up education reform, *J. Coll. Sci. Teach.* **37**, 48 (2008).
- [30] J. D. H. Gaffney, A. L. Housley Gaffney, and R. J. Beichner, Do they see it coming? using expectancy violation to gauge the success of pedagogical reforms, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010102 (2010).
- [31] J. D. H. Gaffney, A. L. Housley Gaffney, and J. J. Chini, Using expectancy violation to investigate student dissatisfaction in studio physics, in *2013 PERC Proceedings*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (Portland, OR, 2013), pp. 153–156, <http://www.compadre.org/per/items/detail.cfm?ID=13130>.
- [32] J. D. H. Gaffney, Education majors' expectations and reported experiences with inquiry-based physics: Implications for student affect, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010112 (2013).
- [33] J. D. H. Gaffney and A. L. Housley Gaffney, Student satisfaction and perceptions of instructor support in studio physics, in *2014 PERC Proceedings*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (Minneapolis, MN, 2014), pp. 95–98, <http://www.compadre.org/per/items/detail.cfm?ID=13457>.
- [34] C. DiStefano, M. Zhu, and D. Mindrila, Understanding and using factor scores: Considerations for the applied researcher, *Pract. Assess. Res. Eval.* **14** (2009).
- [35] It is possible that similar effects were not seen in previous studies because such studies were either done in controlled environments or where students were asked to reflect on a specific instance, such as a piece of feedback on a speech, rather than on the instructor as a whole. In the case of this study, students spend a great amount of time with their instructors, often forming more casual relationships with them; indeed, in anecdotal conversations with students, we would often hear a phrase like, “He’s a nice guy, and I like him, but I really don’t like the class.” Thus, it is reasonable that students would do facework (here, tact) to protect their instructor’s face.
- [36] For a discussion about the effects of “orientation” in terms of bringing students’ expectations in line with their future experiences; see Refs. [30] and [32]. Additionally, an analysis of first-day activities for physics classes to encourage such an orientation is found in Ref. [3], and

- a collection of user-submitted resources for first-day activities in active-learning science classes is located on the Science Education Initiative webpage, <http://www.colorado.edu/sei/fac-resources/framing.html>.
- [37] 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).
- [38] Indeed, while expected course grade did correlate with course satisfaction ($r = 0.37$, $p < 0.001$), we found that for the sections where we had final grades (four of the eight sections), students' *actual* final grade did not correlate with student satisfaction as measured at week 14 ($N = 81$, $r = 0.057$, $p = 0.522$).
- [39] All names are pseudonyms that reflect gender; all interviewees were white non-Hispanic students.
- [40] In this class, instructor *B* created "mixed-ability" groups [41] with one high-performing, one middle-performing, and one low-performing student assigned to each group. While *B* never told the students that this process was followed, the students quickly figured it out by comparing exam scores.
- [41] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).