Evolution of grades and social comparison concern within an introductory physics course

Srividya Suresh and Andrew F. Heckler*

Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Ohio 43210, USA

(Received 5 January 2023; accepted 28 March 2023; published 25 April 2023)

This study investigates the evolution and associations between exam grades and social comparison concern (SCC) among students in an introductory calculus-based physics course. We begin with a descriptive characterization of midterm and final exam scores as well as pre-post SCC scores, including the concurrent evolution of these scores during the course. We hypothesize a feedback loop in which changes in SCC scores are mediated by exam grades, and changes in exam scores are mediated by SCC scores. We employ a structural equation model to determine whether the data are consistent with these hypotheses. Results indicate that there were significant within-student changes in the relative grade standing from exam to exam and that changes in SCC scores depended on both the pre-SCC scores and scores on the first midterm exam. Further, we find evidence that exam scores partially mediate the association between pre- and post-SCC scores, and in turn, post-SCC scores partially mediate associations between midterm and final exam scores, though the mediation effects are somewhat small, comprising 5%–10% of the total effects between exam scores and SCC. We also find that while SCC scores are somewhat correlated with exam scores, they are only very weakly correlated with nonexam grade components, consistent with the idea that exam scores (rather than nonexam scores) are driving changes in SCC and vice versa. Overall, the results provide empirical, correlational evidence to motivate further experimental investigation into a hypothesized dynamic and iterative feedback loop in which student concern about ability or performance compared to others (SCC) can either negatively or positively interfere with student performance on exams.

DOI: 10.1103/PhysRevPhysEducRes.19.010129

I. INTRODUCTION

In this paper, we are interested in the extent to which a student's performance in a university-level introductory physics course is influenced by the student's concern arising from the comparison of performance with other students in the course. There are a number of reasons motivating this investigation. To begin, social comparison theory, and its relation to student performance, has been a topic of study for decades, though the bulk of the investigations were at the K-12 level [1]. This theory is rooted in Festinger's hypotheses that people have a fundamental drive to evaluate their own abilities and that when objective, nonsocial means are not available, they do so by comparing to others' abilities [2]. Such comparisons, at least at the K-12 grade level, have been shown to have both positive and negative consequences in terms of academic performance [1]. On the positive side, "upward" comparisons, namely, comparison to others with higher abilities, can have a positive effect, motivating some students to rise to this higher level [1]. On the other hand, comparison with other students can also heighten feelings of anxiety [1,3,4].

Some empirical findings at the university level also suggest a link between the negative effects of social comparison and academic performance. Specifically, Micari and Drane [5] introduced a construct and scale for social comparison concern (SCC) as a measurement of concern over one's own ability or performance relative to others, and they found that for students in university-level biology, chemistry, organic chemistry, or engineering courses, SCC was a significant though somewhat weak predictor of grade and retention. Recently, Lee *et al.* [6] found that SCC was a predictor of grade ($r \approx -0.3$) in university-level physics courses, further suggesting that SCC may also be an important factor in science, technology, engineering, and mathematics (STEM) courses.

Beyond emerging empirical support, there are also theoretical arguments that provide a mechanism linking SCC to grade performance. For example, the heightened feelings of anxiety resulting from SCC may in turn increase cognitive load and hinder or interfere with cognitive engagement [7] and, ultimately, performance [8]. As pointed out by Micari and Pazos [3], this mechanism of increased cognitive load for SCC is similar to mechanisms proposed to be at play regarding social identity threat [9]. In fact, we propose that SCC may fit within the broader

*heckler.6@osu.edu

perspective of the concept of *psychological vulnerability*, expanded upon by Muenks *et al.* [10], which includes multiple facets: lower belonging, concerns about negative evaluation by others, imposter feelings, and greater negative affect (such as feeling anxious or distressed). They propose that psychological vulnerability can be considered "a psychological threat to students' self-concepts of being smart, competent, and capable in school" and this in turn suppresses engagement and performance in a course. To the extent that SCC may fit within the construct of psychological vulnerability, this mechanism would provide some further insights into why SCC is observed to correlate with grades. Finally, the physics course we examine employs grade-curving (i.e., norm-based grading), which promotes an explicit aspect of competition among students and potentially further adds to SCC. Such competitive academic environments encourage students to compare themselves to others, which can lead to anxiety and doubt in their own competence [11,12], and Canning *et al.* [13] demonstrate that increased perception of a competitive environment leads to increased levels of imposter feelings ("In class, I felt like people might find out that I am not as capable as they think I am"), which in turn leads to lower grades.

The arguments above describe a mechanism for the potential mutual influence between grades and SCC, and such influence would imply that *changes* in one variable depend on the value of the other. For example, a student's change in SCC would depend on the exam grade received between measurements of SCC. Therefore, it is important to first establish the extent to which both grades and motivation change during a course.

First, let us consider the extent to which grades for a given student change within a course. Perhaps surprisingly, there are relatively few formal studies on the evolution of grades within a STEM course at the university level. Some studies suggest that within-student grades do not change much during a course. For example, in introductory biology and several other disciplines, Jensen and Barron [14] have found that about 60% of students had the same first exam/midterm grade as the final course grade. One could argue that if early course performance predicts final grade, then this is an indication of some stability in a student's grade within a course. For example, controlling for ACT score and prior grade point average, Zabriskie *et al.* [15] found that while homework grade was initially the best predictor, the first midterm grade became the best predictor in introductory physics. On the other hand, there is also some evidence for changes in grades. For example, Russell *et al.* [16] investigated early grades and the progression of grades and found that at-risk students who use designed intervention tools positively adjust their grade trajectory within the course, and Sebesta and Speth [17] found that students who use self-regulated learning strategies tend to either maintain high grades or improve their grades during the course.

Second, let us briefly consider the relatively recent and emerging interest in changes in a broad range of motivational factors within a university-level course and the subsequent relation to grades. Note that there are a number of studies examining semester-by-semester changes in motivational factors and grades for K-12 students [18], and fewer at the university level [19], but there are only a small number of studies on motivational change at the university level on within-semester time scales. For example, Corpus *et al.* [20] have documented significant increases in negative motivational factors from the beginning to the end of the semester in first-year students. Robinson *et al.* [21] included performance outcomes in an investigation for engineering students and found that positive motivational beliefs such as expectancy for success and utility value decreased over time and negative beliefs such as psychological cost increased over time and that the trajectories (changes) in these beliefs were related to retention and grades. Magnus and Peresetsky [22] have found that most students in a second-year statistics course are "overconfident" in their grades (i.e., overestimate their grades), and this overconfidence diminishes from exam to exam during the course. Further, they found that women tend to be less overconfident and quicker to adjust their expectations. Khachikian *et al.* [23] found that in several engineering courses, the students who changed their expectations tended to be low-achieving students, appropriately adjusting their expectations downward. Dai and Cromley [24] studied students enrolled in introductory biology classes and found that a decrease in incremental belief (i.e., growth mindset) was associated with dropout from STEM majors. In contrast, Flanigan *et al.* [25] found that a decrease in incremental beliefs resulted in an increase in grades for engineering students in a computer science course. Li and Singh found significant decreases in self-efficacy over a semester in introductory physics, especially for women [26]. Young *et al.* [27] also found significant within-semester declines in a variety of motivational factors (such as self-efficacy and career motivation) and somewhat small correlations with course grades, with students with high grades experiencing smaller changes, though most of the measured motivational factors rebounded in the following semester.

In summary, there is some evidence that within-student grades appear to be relatively stable for many students, but there is also evidence that within-semester grades can change for many others. There is also evidence that a range of positive and negative motivational factors change over the course of a semester and that these changes are associated with performance outcomes such as grades. These studies provide some support for an emerging picture of dynamic changes in grades and motivational factors at the university level and the potential reciprocal influences between them.

In this study, we begin by first establishing and characterizing the extent to which exam grades and SCC change in our introductory physics course context as the course progresses. We do this first because the study is based on the premise that student grades and motivation change within a course. As discussed in the following sections, we find that this premise is warranted, and this observation of changes in grades and SCC allows us to then test the hypotheses of reciprocal influences between grades and SCC based on the theoretical mechanisms of increased threat which reduces performance, as discussed above. We, therefore, propose two research questions and two hypotheses.

RQ1: To what extent do within-student grades and SCC change during an introductory physics course?

RQ2: To what extent are exam and nonexam scores correlated with SCC scores?

Hypothesis 1: Exam scores mediate changes in SCC. That is, achieving a relatively low or high exam grade in the course results in a relative increase or decrease (respectively) in SCC.

Hypothesis 2: SCC mediate changes in exam scores. This hypothesis stems from the consideration of the evolution of exam grades within a semester. For example, while midterm exam scores are empirically positively associated with final exam scores, we hypothesize that this association is partially mediated by SCC because low midterm grades can result in high (or increased) SCC and thus induce a relatively high psychological threat resulting in a lower final exam grade.

Overall, these two hypotheses together propose a feedback mechanism driven by the mutual influences between exams and SCC. It is important to note that while we neglected to preregister these hypotheses [28], we nonetheless report here that we made these hypotheses before this study, which was designed to provide more evidence for (or against) the hypotheses. Finally, we would like to stress that this study is based on correlational data and not experimental intervention, which naturally limits the nature of the conclusions. Of course, the best evidence for causal influences between SCC and grades would include controlled intervention studies, but in this study, we take one step closer to supporting the hypothesis of causation, and thus motivating further study, by examining the longitudinal evolution of exam grades and SCC within a course. To add to this investigation of the hypothesized causal mechanism of psychological vulnerability, we pose RQ2 to also examine the association between SCC and nonexam grades which, because nonexam assignments are in a lower stakes and potentially lower stress context, might be expected to be less associated with SCC than are exam grades.

II. PARTICIPANTS AND METHOD

The data for this study were collected at the Ohio State University, a large public reach university, from a subset of the 1238 students enrolled in the autumn 2019 semester of the first of a two-semester sequence in introductory calculus-based physics. The course structure involves a traditional lecture section with three lectures per week, a weekly recitation section in which there were both individual and group quizzes administered, and a weekly traditional lab section. There were 7 lecturer sections with approximately 200 students enrolled in each section. Most enrolled students (>75%) were engineering majors, and the remaining were mostly made up of STEM majors such as physics, math, biology and chemistry. Each lecture section was taught by an experienced lecturer or experienced tenure-track professor. The lecture sections were also divided into eight recitation and lab sections of about 25 students each and were taught by graduate student teaching assistants. In the lab and recitation sections, students were assigned in groups of four, and the groups were switched after the first midterm. The groups are chosen based on standard exam math scores, and efforts are made so that underrepresented groups (females or ethnic minorities) are not singled out in a group. In addition, the groups are heterogeneous ability groups, therefore efforts were made so that all high- or low-scoring students are not in one group.

The data were collected from three sources. The first source is the university registrar data, which included demographic information, lecture section, class grade, and overall GPA. The second source is the physics course grade book data, which provided all course scores including exams, homework, quizzes, and lab. Note that students were shown the section means and standard deviations for each exam, so they were in principle aware of their relative performance on each exam. This study will focus on exam grades: Midterm 1, midterm 2, and the final exam. The two midterms covered separate content, but the final exam was cumulative. The two midterm exams were worth 15% of the grade and the final exam was worth 25% of the grade.

The third source of data is from the SCC survey data and was collected through *Essential Skills*, an online platform where students are assigned weekly mastery practice in basic math and physics [29]. Students received 1% grade credit for completing all 14 weekly *Essential Skills* assignments, which also included the survey items, and students received full credit for completing the survey items. Table I provides a timeline of when students take the exams and the surveys.

TABLE I. Schedule for social comparison survey and exams.

Week	Task
5	SCC survey
7	Midterm 1
11	Midterm 2
13	SCC survey
16	Final exam

TABLE II. Original social comparison concern survey items and an indication of items included in this study.

Item	Included
I feel different from other people in this physics class.	Yes
I often leave this physics class feeling like I am not as smart as others.	Yes
I often feel intimidated to participate in this physics class.	Yes
I often leave this physics class feeling like I am the only who doesn't understand the material well.	Yes
I worry about getting things wrong in front of my peers in this physics class.	Yes
I have generally understood the material as well as the others understand in this physics class.	No

The original SCC scale from Micari and Drane [5] is composed of six items as shown in Table II adapted for a physics class. The SCC scale ranges from 1 to 7 (Almost never true of myself = 1 to Almost always true of myself = 7), thus for the first five items of the original scale, 1 represents a low social comparison concern and 7 represents a high social comparison concern. However, for this study, we removed the sixth item for two main reasons. First, we found that it was a poor fit in our dataset. For example, when item 6 was removed, the Cronbach's alpha reliability increased from 0.84 to 0.88. Second, there is ample research showing that reverse-coded items tend to decrease the reliability of a scale, and even in cases where it is argued that reverse-coded items may be useful, the original SCC does not align with the recommendations for use of such items, such as including balancing the number and spacing of regular and reversed items [30]. We note here that for the remaining five items used, the composite reliability [31] was 0.88 for the pre-SCC and 0.91 for the post-SCC, in the context of the full structural equation model analysis in Sec. III B.

The subset investigated for most of the results in this study comprises 392 students who completed both the pre- and post-SCC, completed the two midterms and final exam, and consented to participate in the study (about 10% did not). The relatively low percentage (32%) of students in this study compared to the entire enrollment of the class may at least be partially due to the relatively low amount of points awarded for completing an individual assignment, which was less than 0.1% of the grade per assignment. This selection of students presents some concern of bias in the sample since the students were not randomly selected. One reasonable assumption about the (self)-selected population is that our sample is weighted toward students who tend to complete most if not all of the assignments. To understand more about the population of participating students, we present in Table III a comparative breakdown of percentages of students receiving specific grades in both the sample and the entire class, as well as the percentages of two demographic group labels, as indicated in the university database. For gender, we counted students as female who were explicitly labeled as such in the database, though we recognize that there are limitations to such labeling and that gender identity may be a potentially relevant issue for further study. Gender was not indicated for eight of all students enrolled. URMs was defined as underrepresented minorities and includes students who identify as Black, Hispanic, Native American, and Pacific Islander, and only domestic students were included in this designation.

A chi-squared test shows significant differences in grade ($p < 0.001$), gender ($p < 0.001$), and URM status ($p = 0.015$) between the sample and the entire class, indicating that the students in the sample tended to have higher grades, the female students were slightly overrepresented, and URMs were slightly underrepresented compared to the whole class. However, it should be noted that there was still substantial representation of students with grades of A, B, or C, and male and URM students in the study sample. Nonetheless, when interpreting our results, we must keep in mind that the sample does have some biases at least over-weighting students who tend to complete most assignments, and of course, there may be other selection bias effects.

TABLE III. Comparison of the study sample to the entire class.

	Study sample	Entire class
Grade = A	38.8%	25.8%
Grade = B	37.5%	35.7%
Grade = C	19.9%	24.7%
Grade = D or E	3.8%	13.8%
Female	32.1%	25.0%
URM	6.9%	10.6%
Total students	392	1238

III. RESULTS

A. Descriptive statistics of exam and SCC scores

1. Within-student changes in exam and SCC scores in a course

Let us first consider the within-student evolution of exam scores throughout the semester. We focus this analysis of exam score evolution on the $N = 1232$ students who took all three exams, and we will characterize and describe this evolution in two ways.

First, let us describe the distributions of the z-score differences between exams. Recall that the z score of a student is the unitless number of standard deviations away from the mean score. The distribution of the z-score differences between midterm 1 and midterm 2 as well as

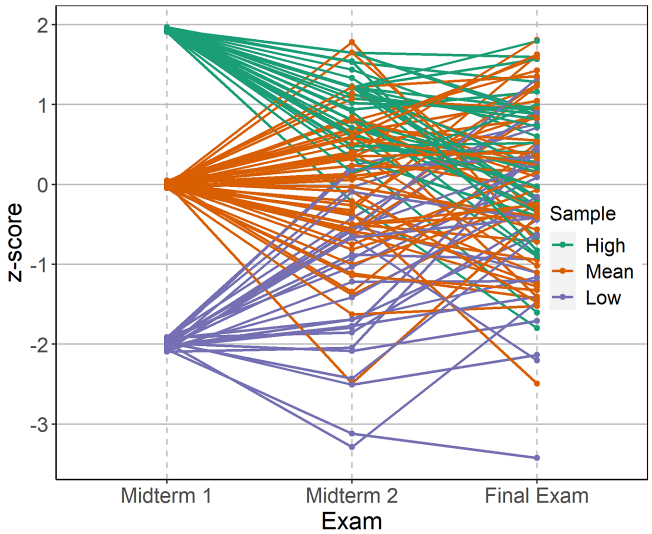

FIG. 1. Evolution of exam z scores for a sample of students: all students with a midterm 1 z score near 2.0 ("High"), near 0.0 ("Mean"), and near −2.0 ("Low").

the differences between midterm 2 and the final exam were both very close to normal distributions with means consistent with zero (as expected) and a standard deviation of 0.99 for the former and 1.16 for the latter. The skewness and kurtosis were consistent with zero for both distributions, except for the midterm 1 to midterm 2 difference kurtosis, which was 0.44, which is still quite small. Examination of the quantile-quantile plots also confirmed that the distributions were very near to normal. Given these standard deviations of the difference distributions, the most important observation is that there was a substantial amount of change in z scores between exams. For example, between midterm 1 and midterm 2, z scores increased or decreased by at least 1 standard deviation for 31% of students and by 0.5 standard deviations for 61% of students. Between midterm 2 and the final exam, the corresponding numbers were similar: 39% and 67%, respectively. Note that this substantial movement of student z scores from exam to exam may provide the context for students to adjust their beliefs in their standing compared to other students.

Second, in order to gain more insight into the substantial longitudinal changes in exam z scores, in Fig. 1, we plotted the evolution of a small sample of students in three groups: all students with a midterm 1 z score near 2.0 (± 0.1) ("high", $N = 37$), near 0.0 (± 0.05) ("mean", $N = 48$), and near -2.0 (± 0.1) ("low", $N = 26$). There are three important features revealed in Fig. 1. First, it is readily apparent that student z scores varied substantially between exams. Second, there was significant regression to the mean, as can be seen by inspection of the high group, which tended to score lower (i.e., toward the mean) on midterm 2 and the final exam, and conversely for the low group. This regression toward the mean is to be expected given the nonperfect correlations between the exam z scores (see Table IV) [32]. Third, even though there was substantial variation and regression to the mean, the correlations were nonzero, and the high group tended to remain relatively high (or at least positive), and the low group tended to remain relatively low (or at least negative).

There were also some substantial longitudinal changes in SCC scores among students. Overall, there was a small mean increase in SCC scores among the students who took both surveys, with the mean pre-SCC of 3.2 and mean post-SCC of 3.4, and a paired t test indicates that the increase is significant [Cohen's $d = 0.10$, [0.004, 0.202] 95% confidence interval (CI)]. To gain more insight into the range of within-student changes, we present a histogram of changes in SCC scores in Fig. 2. The standard deviation of the change in SCC scores was 1.45 points. About 41% of students changed their mean SCC scores by 1 point or more, and 14% changed by 2 points or more. For comparison, the standard deviation of the pre- and post-SCC scores was about 1.6 points. It is important to note that

TABLE IV. Correlations, means, and standard deviations for all students with no missing values on variables ($N = 392$). The midterm, final exam, and nonexam component values are z scores (z scores were calculated using the mean and standard deviation of all 1232 students). All values are significant at $p < 0.05$, except for the cell marked "ns".

	MT 1	MT 2	Final Exam	Pre SCC	Post SCC	Non-exam	Course grade
MT 1							
MT 2	0.650						
Final exam	0.343	0.393					
Pre-SCC	−0.115	−0.124	−0.114				
Post-SCC	−0.222	−0.288	−0.259	0.607			
Nonexam	0.243	0.287	0.312	-0.052^{ns}	−0.113		
Course grade	0.585	0.630	0.748	−0.186	−0.316	0.515	
Mean	0.17	0.23	0.29	3.26	3.41	0.20	3.09
Std. dev.	1.13	1.04	1.00	1.58	1.68	1.01	0.82

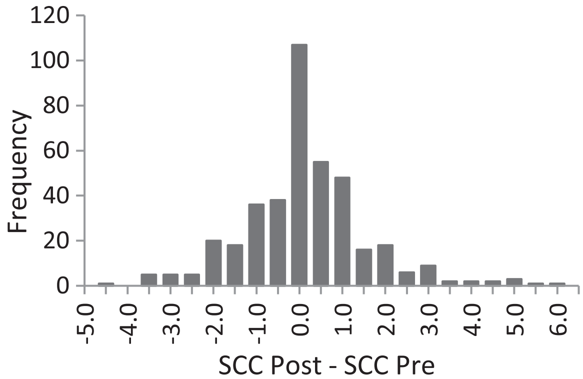

FIG. 2. Histogram of SCC post-SCC prescore differences for all students completing both the pre- and post-SCC. ($N = 392$).

the data in Fig. 2 represent only the 392 students who took both pre- and post-SCC surveys.

2. Correlations between course scores and SCC scores

There were also significant correlations among exam scores, as expected, and somewhat small but significant correlations between SCC and exam scores, with the post-SCC having higher correlations than the pre-SCC with exams, as shown in Table IV. The substantial correlation of post-SCC with final course grades is consistent with the finding from Lee *et al.* [6]. Note that we also included correlations with nonexam components of the course grade, such as homework and lab scores, which do not have the same high-stakes context as exams. It is especially compelling to see that the nonexam components are only very weakly correlated with SCC, much weaker than the exam components. This is consistent with the idea that exam components are associated with social comparison concern and psychological vulnerability.

3. Grouping and analyzing students by "initial state" categories

In order to gain insight into the student evolution of grades and SCC, we grouped students into "initial state" categories based on their pre-SCC and midterm 1 scores, as shown in Fig. 3. The students are grouped into four categories according to whether they are above or below the midterm 1 mean and the pre-SCC mean. There is a fifth category labeled "neutral group" for students within 1 standard deviation of the midterm 1 mean and within 1 point (which is approximately 0.6 SD) of the pre-SCC mean. Therefore, most of quadrant 1 is composed of students with high pre-SCC and high midterm 1 scores we label as the underestimate ability group, and most of quadrant 2 is composed of students with low pre-SCC and high midterm 1 labeled as the high-scoring confident group, most of quadrant 3 is composed of students with low pre and SCC low midterm 1 labeled as the overestimate ability group, and finally, most of quadrant 4 is composed of students with high pre-SCC and low midterm 1 labeled as the low scoring unconfident group. Descriptive statistics of the five groups is provided in Table V.

The behavior of a student in three of the groups, high scoring confident, low scoring unconfident, and neutral group could be considered consistent in the sense that those with high, average, or low midterm 1 grades would tend to have low, average, or high SCC respectively, as also implied by the negative SCC-exam score correlations shown in Table IV. Interestingly, there are a significant number of students who could be considered "inconsistent." The students in the underestimate ability group, have high scores on midterm 1 yet also have high pre-SCC, and students in the overestimate ability group have low scores on midterm 1 yet also have low pre-SCC.

Previous research has found differences in grades [33] and SCC [3,5] between underrepresented groups and their

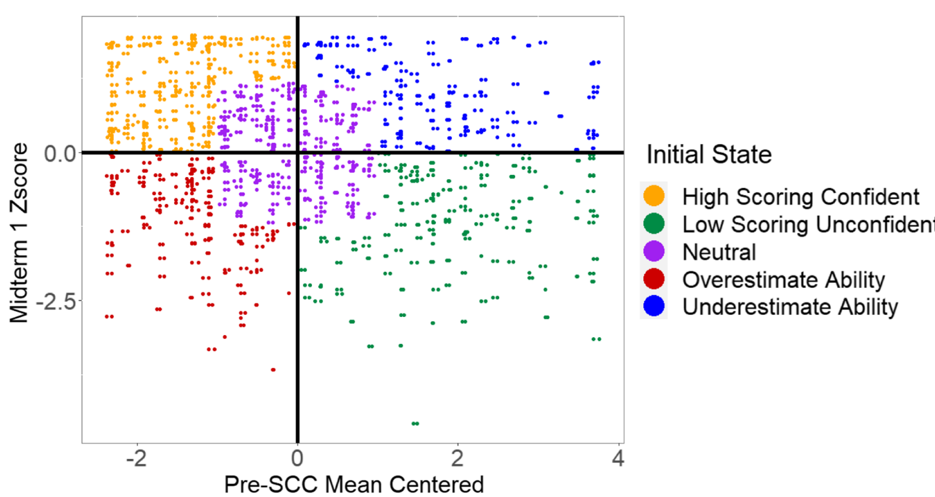

FIG. 3. The five initial-state groups as determined in SCC-midterm 1 score space. Each point represents a student.

TABLE V. Initial-state group summary. Columns 2 and 3 are means for the group.

Group	PreSCC	MT1 Zscore	No. of Students	% Female	% URM
High scoring confident	1.83	1.23	89	17%	4.5%
Overestimate ability	1.83	−1.30	46	41%	8.7%
Neutral	3.10	0.09	148	28%	6.8%
Underestimate ability	5.07	1.12	59	44%	3.4%
Low-scoring unconfident	5.45	−1.24	50	48%	14%
All groups	3.26	0.17	392	32%	6.9%

respective counterparts. Correspondingly, we found demographic differences in our initial-state groups, as shown in Table V. We used a chi-square test of goodness fit to determine whether the observed proportions of demographic factors differ from the proportions of all groups combined. A chi-squared test shows that at least some of the initial-state groups have differing gender proportions χ^2 $[(df = 4) = 21.9,\ p < 0.001]$. The results indicate that, for example, female students are underrepresented in the high-scoring confident group and overrepresented in the low scoring unconfident group. In fact, we found that women had substantially lower pre-SCC scores (Cohen's $d = 0.39$, [0.18, 0.60] 95% CI), yet slightly higher midterm1 scores (Cohen's $d = 0.23$, [0.02, 0.45] 95% CI) We also found evidence of potential differences in the initial state groups for URM students χ^2 $[(df = 4) = 14.8$, $p = 0.06]$, though the numbers of students in this demographic is too small to be meaningfully investigated with these methods.

To provide a sense of the evolution of SCC, Fig. 4 presents a simple picture of the evolution of SSC means within the initial-state groups. Using a paired Wilcoxon signed-rank test, we found significant pre-post changes in SCC, (ΔSCC), for three of the initial state groups. The high-scoring confident group significantly increased in SCC ($\Delta\text{SCC} = 0.5$, $V = 1166$, $p < 0.05$, and $d = 0.45$), the overestimate ability group on average reported a relatively large increase in SCC ($\Delta\text{SCC} = 0.9$, $V = 1531$, $p < 0.001$, and $d = 0.90$), and finally, the underestimate ability group significantly decreased in SCC ($\Delta\text{SCC} = -0.8$, $V = 1877$, $p < 0.001$, and $d = 0.65$). It is interesting to observe that, for example, the low-scoring unconfident group and underestimate ability group started with a similar mean pre-SCC but then diverge for the post-SCC. These results provide a preliminary indication of support for our hypothesis since we would expect, for example, that students in the underestimate ability group, who score relatively high on the first midterm, would decrease their reports of SCC more than the low-scoring unconfident group, who score relatively low on the first midterm. Next, we will more formally investigate the hypothesis that the midterm exam grades are a factor in these differential changes in SCC.

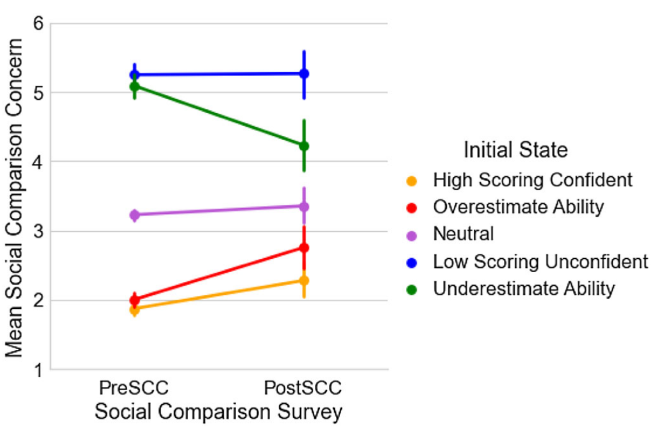

FIG. 4. Pre- and postsocial comparison concern for each of the five initial-state groups.

B. Evidence for mutual influence between grades and SCC

In order to build a model that systematically accounts for correlations among relevant factors, we employed a structural equation model (SEM) analysis to test the time-ordered (thus directional) hypothesized relations between pre- and post-SCC, modeled as latent variables measured by the five SCC items and the exam scores as direct observations. A simplified version of this model is shown diagrammatically in Fig. 5, and the full diagrammatic model is shown in Fig. 8 in the Appendix. Since the items

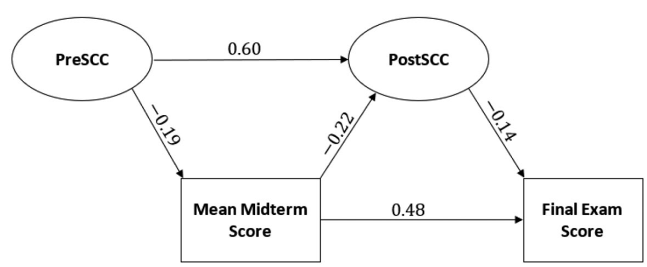

FIG. 5. Structural equation model of SCC and exams and the standardized path estimates (standardized including all variables). Considering the unstandardized estimates and errors in Table VII, all path coefficients are statistically significant (e.g., $p < 0.01$). For simplicity, the indicators of the latent variables and the correlations between residuals of identical indicators are not shown. A diagram of the full model is in Fig. 8 in the Appendix.

in the pre- and post-SCC were identical, correlations between the residuals of identical items were included in the model ([34]). The correlations, means, and standard deviations of the measured variables are reported in Table VI in the Appendix. The analysis was done for the 392 students who completed all exams and SCC surveys, and the absolute values of the skewness and kurtosis of all factors in this model are less than 0.85 and 3.5, respectively, which are well within the range of typically acceptable values to help ensure valid SEM results [35]. We employed *lavaan*, an R analysis package [36,37], and used the maximum likelihood "MLM" estimator which uses robust estimators of standard errors and a Satorra-Bentler scaled test statistic. The analysis ended (converged) normally after 61 iterations.

Results from the SEM analysis are shown in Fig. 5 and Tables VII and VIII in the Appendix. The chi-squared test indicates that the model fails the exact fit test, though this is neither unexpected nor particularly informative since our sample size and the number of degrees of freedom are large [35]. On the other hand, the recommended approximate fit indices for global fit which account for sample size and degrees of freedom in various ways, namely, the root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean squared residual (SRMR), indicate at very least that the model is not a poor fit to the data and is consistent with being a good fit to the data (e.g., for RMSEA < 0.05, $p = 0.16$) [35]. Further, the correlation residuals were all less than 0.10, and the vast majority were less than 0.05, which provides more evidence of a good local fit (i.e., variable by variable) [35]. Considering these global and local fit tests, we retain the model.

Table VII in the Appendix indicates that all of the unstandardized path coefficients and indirect effects are significantly different from zero (ps < 0.001), and Fig. 5 displays the standardized path estimates of the SEM. Let us discuss two important results from this analysis. First, the model estimates support hypothesis 1 that the association of pre-SCC scores with post-SCC scores is partially (albeit weakly) mediated by midterm scores. This result is shown by the nonzero estimates for the indirect effect pre-SCC → midterm mean → post-SCC in Table VII. Specifically, the SEM analysis indicates that while an increase in pre-SCC by 1 standard deviation results in a total effect of an increase of 0.64 standard deviations in post-SCC, this total effect comprises a direct effect increase of 0.60 standard deviations and an indirect effect increasing post-SCC by an additional 0.04 standard deviations, mediated by midterm scores. Thus, this partial mediation result indicates that an increase in pre-SCC is associated with a decrease in mean midterm scores which in turn is associated with an increase in the post-SCC.

Second, the model in Fig. 5 supports hypothesis 2 that the association of the mean midterm scores with the final exam score is partially (though weakly) mediated by post-SCC, again as shown by the nonzero indirect effect midterm mean → post-SCC → final exam in Table VII. That is, an increase of 1 standard deviation in mean midterm scores results in a total effect increasing the final exam score by 0.51 standard deviations. This total effect comprises a direct effect associated with an increase in the final exam score by 0.48 standard deviations, and there is a significant indirect effect increasing the final exams score by an additional 0.03 standard deviations, mediated by post-SCC. Specifically, a greater mean midterm score is associated with a decrease in post-SCC which in turn results in an increase in final exam score.

1. Further analysis and representation of changes in exam and SCC scores

The SEM analysis above provides some formal quantitative support for our hypotheses, but further analyses and representations of the results can provide more insight and support (even if less precise) for the direction and magnitude of the hypothesized mutual influence between SCC and exam scores. To that end, we investigated how within-student *changes* in SCC scores are related to both the midterm and final exam scores. We further investigate this relation for each of the initial state groups to get a sense of how changes in SCC are related to grades for each of these groups. The following analysis is done for the 392 students who completed all exams and SCC surveys.

Let us first consider the change in SCC (i.e., post-pre) in terms of midterm 2 scores for all participants and then broken up by initial state groups. We chose midterm 2 z scores because the initial state groups are defined by the midterm 1 z scores. A simple linear regression with change in SCC as the dependent variable resulted in a significant negative dependence on midterm 2 z score, as expected from hypothesis 1 [coefficient estimate $= -0.28$ (SE $= 0.07$), $p < 0.001$, and $R^2 = 0.04$], indicating that the midterm 2 score accounts for about 4% of the variance in the change in SCC.

In Fig. 6, we present the change in SCC based on midterm 2 scores for the four initial-state groups, showing patterns supporting hypothesis 1. Consistent with the linear regression of change in SCC on midterm 2 scores, we see the trend that the higher the midterm 2 scores tend to result in less positive (or more negative) changes in SCC.

Given the evidence above that midterm scores are related to changes in SCC, let us next more intuitively investigate how changes in SCC are in turn related to final exam z scores. First, we employ a simple linear regression with the final exam z score as the dependent variable and find a significant negative dependence on the change in SCC, consistent with hypothesis 2 [coefficient estimate $= -0.12$ (SE $= 0.03$), $p < 0.001$, $R^2 = 0.03$]. We can gain some further insight into how the final exam score is related to the change in SCC by considering the initial-state groups. Figure 7 shows that for all groups except the low-scoring

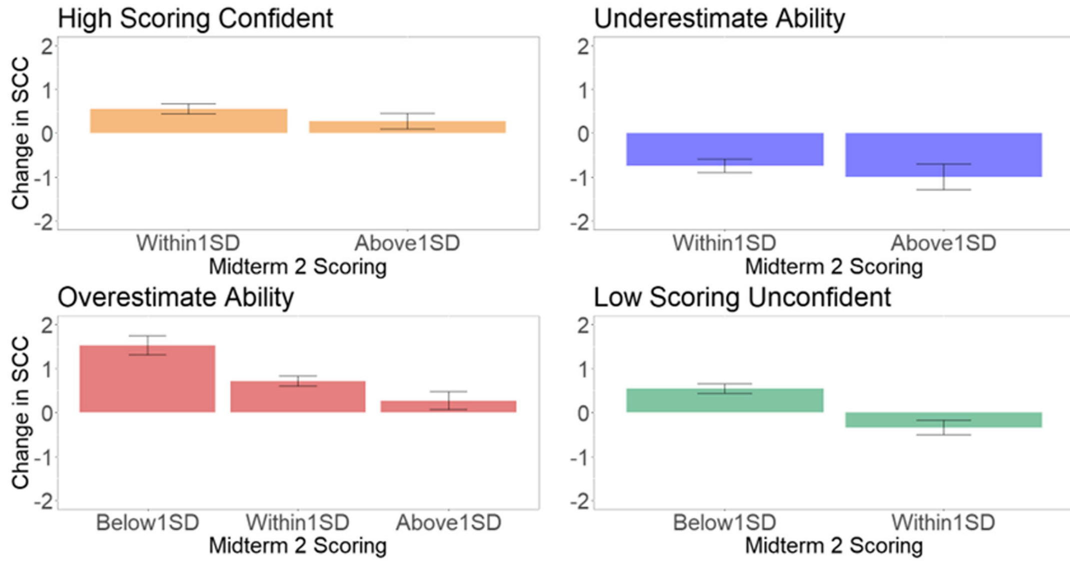

FIG. 6. The mean change in SCC versus midterm 2 scoring above, below, or within 1 standard deviation of the class mean score, for four initial state groups. Error bars are 1 SE.

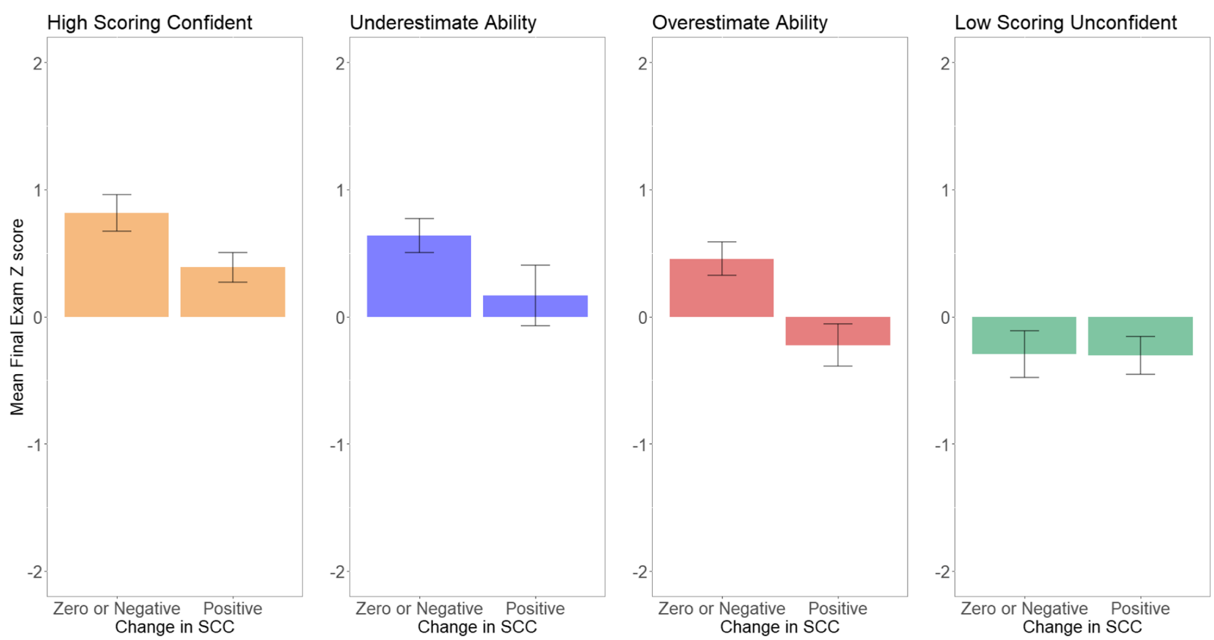

FIG. 7. Mean final exam z score vs the change in SCC and initial-state group. Error bars represent 1 SE.

confident group, students who reported an increase in SCC received lower final exam scores on average compared to students whose SCC did not change or decreased.

IV. CONCLUSION AND DISCUSSION

This study reports a number of findings on the evolution of grades and SCC within a course. With regard to RQ1, the overall picture was one of substantial variation in changes in exam scores and SCC throughout the semester. More specifically, while there were correlations among exam scores, as to be expected, we found that substantial changes in exam z scores throughout the semester were fairly common in this introductory physics course, with about one-third of students changing by at least 1 standard deviation from exam to exam. There were also significant ($\sim$1 standard deviation) changes in within-student pre- to post-SCC scores for many students, and a small overall mean increase of about 0.1 standard deviations in SCC. Further, we confirmed previous work that post-SCC is significantly correlated with grades and exam components of grades, though the correlations were somewhat weak, in the range of 0.25–0.30. The documentation of these changes in exam and SCC scores sets the motivation and context for investigating mechanisms underlying the changes.

With regard to RQ2, interestingly, we found that SCC was only very weakly, perhaps even negligibly, correlated with nonexam components. This is aligned with our theoretical perspective that SCC is likely more related to the high-stakes graded events such as exams that, we suppose, are more likely to invoke psychological vulnerability than lower stakes (and lower time-pressured) nonexam components. Further, the observed difference in correlation of these two grade components with SCC would further support the idea of reducing the weight of exam components (cf. [33]).

A perhaps more fundamentally important finding of this study comes from an SEM analysis that empirically revealed two mediations. The first is that the association between pre- and post-SCC scores was partially mediated by the intervening midterm exams, and the second is that the association between the mean midterm exam and final exam scores was partially mediated by post-SCC scores. While these findings are strictly correlational in nature, they do provide some support for our hypotheses that the evolution of grades and SCC may be causally linked in a feedback loop. For example, we hypothesized that low exam scores provide students with negative feedback which may raise SCC, and this in turn may further lower subsequent exam scores. It is important to note that the estimated effect size is somewhat small. In other words, the mediations are only partial and somewhat weak, at around 5%–10% of the size of the direct effect in each case.

Although the effect sizes may be somewhat small, they may still be both theoretically and practically important for several reasons. First, considering recent discussions, such as by Hofman *et al.* [38], reflecting on the limitations of predicting human behavior, one must acknowledge the context of this study, namely, that exam scores are an outcome of a highly complex environment with many (human) factors, and there are going to be natural upper limits to the prediction by *any* one factor, or even by all reasonably measurable factors combined. For example, ACT score typically accounts for only about 20%–25% of exam score variance [33], and in our study, the exams themselves only predict 12%–42% of the variance in other exams. Second, while the observed mediations may be small, they are consistent with the hypothesis that these meditations may be due to a systematic, causal factor in student performance. Discovering and documenting such a potential causal factor among many in this complex context is important to advance our knowledge of student learning and performance. Further, this hypothesized feedback mechanism may be cumulative both within and between courses, potentially resulting in longer-term, more sizable effects. Put another way, if this mechanism is real, then interventions addressing SCC may have cumulative beneficial effects.

It is also worth noting that this hypothesized SCC feedback effect could have an unintended consequence inevitably arising from the imperfect statistical reliability of exam scores. For example, if a student receives a score that is below their "true score" due to the imperfect reliability of the exam, then this relatively negative score could result in an increased SCC which in turn could result in the student scoring lower on the following exam. Note that this SCC feedback loop, for example, in which relatively low exam 1 scores result in even lower exam 2 scores, is opposite in direction compared to the purely statistical effect of regression to the mean. Further, at least for students scoring relatively low, the SCC feedback loop is opposite to the intended direction of methods of self-regulated learning, in which learners create "self-oriented feedback loops" to monitor their performance and improve [39], increasing the challenge of such efforts. Thus, if the hypothesized SCC effect is eventually found to be real, then understanding ways to counter SCC may be helpful in improving self-regulated learning.

While exam scores and SCC are found to be correlated, we find that students span this two-dimensional space, as shown by our initial state groups. Besides the natural concern with the low-scoring, high-SCC students, of particular interest may be the students who underestimate or overestimate their abilities compared to others. We note that students in these two groups tend to change their SCC the most, aligning it more consistently with their relative exam performance. But it is worth considering for future research and interventions why students in the underestimate ability group, especially women, who are overrepresented in this group, have a high pre-SCC and how to address this.

There are several limitations to this study to keep in mind. One limitation is that the data collected were only during one semester at one institution, thus limiting the generalizability of the findings. Another significant limitation of this study is the fairly low percentage of enrolled students (32%) who took both the pre- and post-SCC and could thus be included in the full (SEM) analysis. We know that the participating students were a biased sample, though it is not clear how the sampled population would bias the results. For example, the students in full analysis tended to have higher grades. While there were students in all grade ranges in the full analysis, we found that 73% of the students in the full analysis received a B grade or above, compared to 58% of the entire class receiving a B or above.

Therefore, it will be important to replicate and continue investigations on this topic to a broader range of student populations. This may be especially important given that we observed that students in different initial state groups evolve differently in the grade-SCC space and a closer examination of these different paths warrants more study. Finally, other significant limitations of this study are in the assumption of the causal agents. Specifically, we assume that grades are the main driving force (feedback) causing the change in SCC, as opposed to some other feedback the student may receive in the course. We also assume the SCC is the main driver of change in grades, when in fact, our measurement of SCC could be a proxy for some other motivational factor such as other aspects of psychological vulnerability.

Finally, let us reiterate that this is an exploratory, correlational study. Such studies are an important part of the scientific discovery process and help to motivate and provide important information for further investigation. As such, we propose that our results warrant further study of this potential causal influence and feedback loop between SCC and student performance. Of course, a controlled intervention study is needed to more rigorously determine whether there are causal influences between grades and SCC. However, we propose that the more prudent next steps are to first determine the replicability and predictability of our findings, including to more contexts and student populations, to design and implement better data collection to more precisely establish the sequential timing of our hypothesized causes and effects, and to probe other potentially viable and related motivational factors to better understand the mechanisms and allow for more informed and better-aimed interventions.

ACKNOWLEDGMENTS

Funding for this research was provided by the Center for Emergent Materials: an NSF MRSEC under Grant No. DMR-2011876. We gratefully acknowledge Mike Lopez who led us to consider the potential importance of SCC in student performance, and we also gratefully acknowledge the cooperation of the introductory physics programs led by Thomas Gramila and especially Thomas Barrett for his valuable assistance in retrieving the registrar data used in this study.

APPENDIX: DETAILED RESULTS OF SEM ANALYSIS

This appendix presents more detailed information on the structure of the SEM used in this paper, the summary statistics of the measured variables in the SEM, and the results of the SEM.

TABLE VI. Correlations and summary statistics of variables in SEM. All values are significant at $p < 0.05$, except for the cells marked "ns".

	Pre-SCC1	Pre-SCC2	Pre-SCC3	Pre-SCC4	Pre-SCC5	Post-SCC1	Post-SCC2	Post-SCC3	Post-SCC4	Post-SCC5	Mean midterm	Final exam
Pre-SC2	0.623											
Pre-SC3	0.582	0.727										
Pre-SC4	0.530	0.777	0.678									
Pre-SC5	0.398	0.547	0.632	0.538								
Post-SC1	0.431	0.373	0.399	0.372	0.323							
Post-SC2	0.367	0.574	0.49–2	0.508	0.387	0.654						
Post-SC3	0.335	0.491	0.532	0.472	0.442	0.643	0.800					
Post-SC4	0.349	0.499	0.513	0.481	0.425	0.635	0.807	0.774				
Post-SC5	0.312	0.416	0.456	0.425	0.499	0.503	0.670	0.736	0.715			
MeanMT	-0.052^{ns}	−0.224	−0.151	−0.137	-0.087^{ns}	−0.198	−0.354	−0.246	−0.303	−0.235		
Final exam	-0.010^{ns}	−0.196	−0.101	−0.153	-0.027^{ns}	−0.257	−0.302	−0.258	−0.228	−0.187	0.529	
Mean	3.44	3.34	3.21	3.04	3.26	3.57	3.50	3.44	3.13	3.41	75.8	66.5
Std. dev.	1.82	1.97	1.90	1.99	1.90	1.86	1.92	1.96	1.98	1.95	16.1	17.8

TABLE VII. SEM estimates of path coefficients. Standardized estimates are standardized for all variables.

	Unstandardized estimates	Standard error	Standardized estimates
Direct effects			
Pre-SCC → post-SCC	0.657	0.067	0.599
Pre-SCC → midterm mean	−2.482	0.743	−0.189
Midterm mean → post-SCC	−0.018	0.003	−0.216
Post-SCC → final exam	−1.782	0.611	−0.135
Midterm mean → final exam	0.534	0.051	0.484
Indirect effects			
Pre-SCC → midterm mean → post-SCC	0.045	0.015	0.041
Midterm mean → post-SCC → final exam	0.032	0.012	0.029
Measurement component			
Pre-SCC → pre-SCC1	1.000	⋯	0.672
Pre-SCC → pre-SCC2	1.432	0.088	0.894
Pre-SCC → pre-SCC3	1.292	0.090	0.835
Pre-SCC → pre-SCC4	1.360	0.098	0.840
Pre-SCC → pre-SCC5	1.013	0.091	0.656
Post-SCC → post-SCC1	1.000	⋯	0.718
Post-SCC → post-SCC2	1.294	0.074	0.904
Post-SCC → post-SCC3	1.284	0.072	0.886
Post-SCC → post-SCC4	1.311	0.077	0.890
Post-SCC → post-SCC5	1.113	0.079	0.776
Model fit parameters			
RMSEA, 90% CI	CFI	SRMR	Chi squared
0.058 [0.045, 0.071]	0.979	0.035	106.1 ($p < 0.001$)df $= 46$

TABLE VIII. SEM estimates of model covariances and variances. Standardized estimates are standardized for all variables.

	Unstandardized estimates	Standard error	Standardized estimates
Covariances			
Pre-SCC1 ↔ post-SCC1	0.546	0.123	0.310
Pre-SCC2 ↔ post-SCC2	0.262	0.074	0.361
Pre-SCC3 ↔ post-SCC3	0.162	0.076	0.172
Pre-SCC4 ↔ post-SCC4	−0.044	0.080	−0.045
Pre-SCC5 ↔ post-SCC5	0.510	0.099	0.294
Variances			
Pre-SCC1	1.824	0.166	0.584
Pre-SCC2	0.777	0.103	0.201
Pre-SCC3	1.089	0.160	0.302
Pre-SCC4	1.163	0.171	0.295
Pre-SCC5	2.042	0.193	0.569
Post-SCC1	1.699	0.179	0.484
Post-SCC2	0.677	0.112	0.183
Post-SCC3	0.816	0.104	0.215
Post-SCC4	0.813	0.122	0.207
Post-SCC5	1.475	0.160	0.397
Pre-SCC	1.505	0.196	0.964
Post-SCC	0.986	0.157	0.704
Midterm mean	250.2	20.6	1.00
Final exam	222.0	18.9	0.545

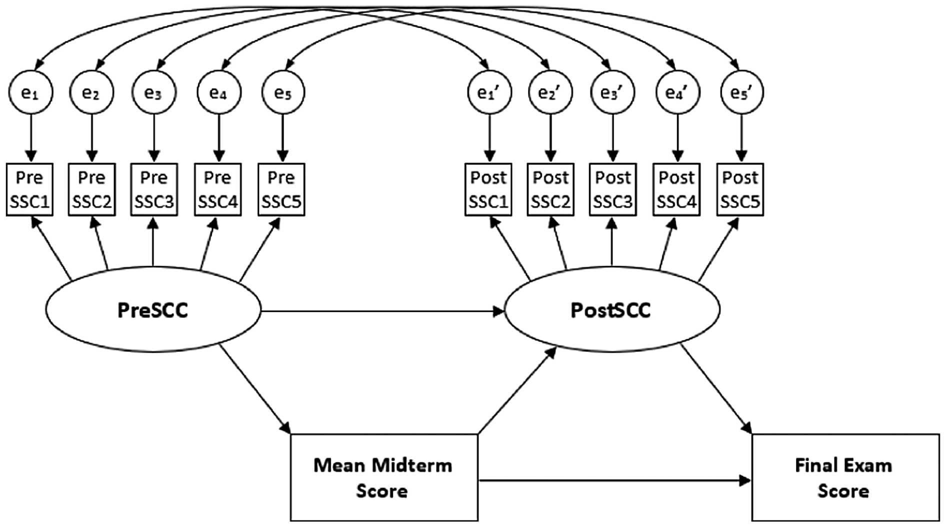

FIG. 8. Diagram of the full SEM model.

[1] P. Dijkstra, H. Kuyper, G. van der Werf, A. P. Buunk, and Y. G. van der Zee, Social comparison in the classroom: A review, Rev. Educ. Res. **78**, 828 (2008).

[2] L. Festinger, A theory of social comparison processes, Hum. Relat. **7**, 117 (1954).

[3] Marina Micari and Pilar Pazos, Worrying about what others think: A social-comparison concern intervention in small learning groups, Active Learn. Higher Educ. **15**, 249 (2014).

[4] B. D. Pulford, B. Woodward, and E. Taylor, Do social comparisons in academic settings relate to gender and academic self-confidence?, Soc. Psychol. Educ. **21**, 677 (2018).

[5] M. Micari and D. L. Drane, Intimidation in small learning groups: The roles of social-comparison concern, comfort, and individual characteristics in student academic outcomes, Active Learn. Higher Educ. **12**, 175 (2011).

[6] H. Lee, L. Y. Shirley, M. Kim, and A. C. Koenka, Concern or comfort with social comparisons matter in undergraduate physics courses: Joint consideration of situated expectancy-value theory, mindsets, and gender, Contemp. Educ. Psychol. **67**, 102023 (2021).

[7] R. Pekrun and L. Linnenbrink-Garcia, Academic emotions and student engagement, in *Handbook of Research on Student Engagement*, edited by S. L. Christenson, A. L. Reschly, and C. Wylie (Springer, Boston, MA, 2012), pp. 259–282.

[8] N. Derakshan and M. W. Eysenck, Anxiety, processing efficiency, and cognitive performance, Eur. Psychol. **14**, 168 (2009).

[9] T. Schmader, M. Johns, and C. Forbes, An integrated process model of stereotype threat effects on performance, Psychol. Rev. **115**, 336 (2008).

[10] K. Muenks, E. A. Canning, J. LaCosse, D. J. Green, S. Zirkel, J. A. Garcia, and M. C. Murphy, Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class, J. Exp. Psychol. **149**, 2119 (2020).

[11] N. Sommet, C. Pulfrey, and F. Butera, Did my MD really go to university to learn? Detrimental effects of numerus clausus on self-efficacy, mastery goals and learning, PLoS One **8**, e84178 (2013).

[12] E. Seymour and A. B. Hunter, *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education* (Springer, Cham, 2019).

[13] E. A. Canning, J. LaCosse, K. M. Kroeper, and M. C. Murphy, Feeling like an imposter: The effect of perceived classroom competition on the daily psychological experiences of first-generation college students, Soc. Psychol. Pers. Sci **11**, 647 (2020).

[14] P. Jensen and J. Barron, Midterm and first-exam grades predict final grades in biology courses, J. Coll. Sci. Teach. **44**, 82 (2014).

[15] C. Zabriskie, J. Yang, S. DeVore, and J. Stewart, Using machine learning to predict physics course outcomes, Phys. Rev. Phys. Educ. Res. **15**, 020120 (2019).

[16] J. E. Russell, A. Smith, and R. Larsen, Elements of success: Supporting at-risk student resilience through learning analytics, Comput. Educ. **152**, 103890 (2020).

[17] A. J. Sebesta and E. B Speth, How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology, CBE Life Sci. Educ. **16**, ar30 (2017).

[18] R. Pekrun, S. Lichtenfeld, H. W. Marsh, K. Murayama, and T. Goetz, Achievement emotions and academic performance: Longitudinal models of reciprocal effects, Child Development **88**, 1653 (2017).

[19] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Damage caused by women's lower self-efficacy on physics learning, Phys. Rev. Phys. Educ. Res. **16**, 010118 (2020).

[20] J. H. Corpus, K. A. Robinson, and S. V. Wormington, Trajectories of motivation and their academic correlates over the first year of college, Contemp. Educ. Psychol. **63**, 101907 (2020).

[21] K. A. Robinson, Y. K. Lee, E. A. Bovee, T. Perez, S. P. Walton, D. Briedis, and L. Linnenbrink-Garcia, Motivation in transition: Development and roles of expectancy, task values, and costs in early college engineering, J. Educ. Psychol. **111**, 1081 (2019).

[22] J. R. Magnus and A. A. Peresetsky, Grade expectations: Rationality and overconfidence, Front. Psychol. **8**, 2346 (2018).

[23] C. Khachikian, D. Guillaume, and T. Pham, Changes in student effort and grade expectation in the course of a term, Eur. J. Eng. Educ. **36**, 595 (2011).

[24] T. Dai and J. G. Cromley, Changes in implicit theories of ability in biology and dropout from STEM majors: A latent growth curve approach, Contemp. Educ. Psychol. **39**, 233 (2014).

[25] A. E. Flanigan, M. S. Peteranetz, D. F. Shell, and L. K. Soh, Implicit intelligence beliefs of computer science students: Exploring change across the semester, Contemp. Educ. Psychol. **48**, 179 (2017).

[26] Y. Li and C. Singh, Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses, Phys. Rev. Phys. Educ. Res. **17**, 010143 (2021).

[27] Anna Young, Paul Wendel, Joan Esson, and Kathryn Plank, Motivational decline, and recovery in higher education STEM courses, Int. J. Sci. Educ. **40**, 1 (2018).

[28] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, The preregistration revolution, Proc. Natl. Acad. Sci. U.S.A. **115**, 2600 (2018).

[29] B. Mikula and A. Heckler, Framework and implementation for improving physics essential skills via computer-based practice: Vector math, Phys. Rev. Phys. Educ. Res. **13**, 010122 (2017).

[30] B. Weijters and H. Baumgartner, Misresponse to reversed and negated items in surveys: A review, J. Mark. Res. **49**, 737 (2012).

[31] T. Raykov, Behavioral scale reliability and measurement invariance evaluation using latent variable modeling, Behav. Ther. **35**, 299 (2004).

[32] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Routledge, London, 2013).

[33] A. Simmons and A. Heckler, Grades, grade component weighting, and demographic disparities in introductory physics, Phys. Rev. Phys. Educ. Res. **16**, 020125 (2020).

[34] H. W. Marsh and K. T. Hau, Assessing goodness of fit: Is parsimony always desirable?, J. Exp. Educ. **64**, 364 (1996).

[35] R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford Publications, New York, NY, 2015).

[36] Y Rosseel, lavaan: An R Package for structural equation modeling, J. Stat. Softw. **48**, 1 (2012).

[37] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2016).

[38] J. M. Hofman, A. Sharma, and D. J. Watts, Prediction and explanation in social systems, Science **355**, 486 (2017).

[39] B. J. Zimmerman and D. H. Schunk, Self-regulated learning and performance: An introduction and an overview, in *Handbook of Self-Regulation of Learning and Performance* (Routledge, New York, NY, 2011), pp. 15–26.