Relationship of prior preparation and conceptual diagnostic measures to success in the upper-level electricity and magnetism course

Fargol Seifollahi*
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

David E. Meltzer
School of Applied Sciences and Arts, College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona, USA

Chandralekha Singh
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

(Received 10 June 2025; accepted 23 March 2026; published 15 May 2026)

The majority of previous studies of physics courses at the upper undergraduate level have been content-related or focused on the format of instruction, while detailed investigation of student performance patterns across various conceptual and quantitative measures in this context has been neglected, relatively speaking, in comparison to the substantial work that has been carried out in introductory courses. In this study, we used a sample of students in a traditional upper-level electricity and magnetism course at a large U.S. university to examine performance patterns and understand the extent to which, and the ways in which, conceptual and quantitative performance were related to each other. To this end, we used various measures of prior preparation (including standardized math scores and background in physics and mathematics in high school and college), in-course quantitative performance (homework, quizzes, and exams), and conceptual performance (scores on the Conceptual Survey of Electricity and Magnetism). Overall, our analysis revealed that prior preparation was a significant predictor of success in this course and that there were strong correlations between students' conceptual and quantitative performance. In particular, grades in the most recent upper-level physics course (upper-level mechanics) were the strongest predictor of upper-level electricity and magnetism course performance, even after controlling for mathematical preparation. Although average normalized conceptual gains were relatively large and independent of preinstruction conceptual scores, threshold effects suggested that students entering with the lowest conceptual understanding were unable to achieve high postinstruction conceptual scores or overall course grades. Our findings suggest that better prior preparation through early development of mathematical skills, as well as instructional strategies that support learners in bridging physics and mathematics effectively, may be crucial for success in upper-level electricity and magnetism.

DOI: 10.1103/b4y3-6d2t

I. INTRODUCTION

Electricity and magnetism is usually regarded as one of the most challenging subjects in the undergraduate physics curriculum [1–6]. This is due to the conceptually complex and mathematically intensive nature of the subject, particularly in upper-level courses where students need to integrate quantitative problem-solving skills with sufficient conceptual understanding of the electromagnetic phenomena [7–10]. In order for instructors to best guide students as they advance from introductory to upper-level electricity and magnetism courses, an understanding of patterns in student learning may become increasingly useful. This calls for identifying factors that contribute to success in these advanced courses, as well as investigating how various measures of preparation and performance are correlated. Understanding such patterns is particularly important in upper-level contexts, where both the student population and course objectives differ substantially from those of introductory level courses [11]. While much of the existing literature in physics education research has been focused on identifying student difficulties in electricity and magnetism and mechanics courses, relatively few research works have focused on patterns of student success in the upper-level context [12–16].

*Contact author: fas102@pitt.edu

The purpose of the current study is to address this gap by assessing the extent to which different conceptual and quantitative measures, including prior academic preparation and conceptual background, relate to student performance and to one another in a traditional upper-level electricity and magnetism course in a large U.S.-based university.

A. Role of academic preparation and math readiness

Students' academic preparation, particularly their earlier mathematics and physics backgrounds, can potentially serve as a significant predictor of success in subsequent college courses. Studies across multiple physics contexts have shown that mathematical and physics skills developed in earlier stages of learning tend to carry forward and predict performance in later coursework [16–18]. For example, standardized mathematics test scores, such as SAT Math, have been found to correlate significantly with achievement in introductory physics [19]. Within the context of electricity and magnetism specifically, while some studies have focused on transforming upper-level courses to go beyond traditional instruction [16,20,21], the majority of studies are centered around identifying topic-specific difficulties and persistent misconceptions related to fields, flux, and electromagnetic induction [12–15,22,23]. Across studies exploring electricity and magnetism difficulties, a commonly emerging theme is students' struggle with the mathematical sophistication of the subject, especially related to the use of vector calculus and differential equations. In introductory electricity and magnetism, for example, students' vector calculus preparation has been shown to correlate significantly with their final exam performance, highlighting the importance of mathematical skills at the introductory level [18].

Still, the extent to which findings from introductory physics apply to upper-level courses remains largely unexplored. Upper-level electricity and magnetism courses are different from introductory courses in terms of both the student population and the learning objectives. Students in upper-level electricity and magnetism have generally taken prerequisite courses and, therefore, are expected to have adequate mathematical skills and baseline familiarity with electromagnetic concepts. These students are also more likely to be majoring in physics, unlike students from a range of academic backgrounds in introductory courses. As a result, they are exposed to substantially more abstract concepts than in introductory electricity and magnetism and are expected to be more at ease with using advanced mathematics for problem solving. However, research shows that students still tend to have persistent difficulties with foundational electricity and magnetism concepts, with little to no measurable improvement from introductory to upper-level coursework [23–25]. Prior preparation appears to play a role here; yet, to our knowledge, no study so far has examined how various measures of prior preparation, including mathematical skills, are associated with success in upper-level electricity and magnetism courses.

B. Connection between physics and math

The relationship between physics and mathematics has been extensively explored, both from historical and educational perspectives [26–30]. In the physics education context, researchers have argued that mathematics should not be viewed merely as a tool for manipulating equations, but rather as a language that helps model and formulate physical ideas into concrete forms. For example, Tzanakis described physics and mathematics as "interwoven in the sense of a bidirectional process" where mathematics has been used as a language for physics, and physics provides the natural framework for operationalizing mathematical theories [26].

Moreover, Redish suggested that the language of math used in physics is not necessarily the same as the one mathematicians use [31,32]. In fact, mathematics can be viewed as the "structural language of physical thought," which is used as a way of logical deductive reasoning that is inseparable from physics and goes beyond something learned exclusively in math [33]. Following a model proposed by Uhden *et al.*, meaningful learning in physics can only occur when the structural role of mathematics, which includes embedding logical structure and defining physical concepts, complements the technical role of executing calculations [27]. This view also aligns with Hestenes' modeling theory [34], which suggests a model-centered strategy for solving problems in physics that allows students to recognize that mathematical models reflect real physical problems. Therefore, expertlike problem solving in physics and specifically in upper-level electricity and magnetism requires the proper integration of physics and mathematics.

C. Conceptual learning in electricity and magnetism

Despite the mathematical intensity in upper-level electricity and magnetism courses, proper conceptual understanding still remains key to developing expertise in the subject. Focusing solely on mathematical skills may lead students toward algorithmic approaches without properly thinking about the underlying physical concepts. To this end, researchers have studied students' conceptual and quantitative problem-solving approaches to make a distinction between their conceptual and quantitative performance. Generally, conceptual understanding involves qualitative reasoning and grasping physical phenomena without explicitly relying on heavy math, while quantitative problem solving involves explicit calculations and mathematical procedures.

Conceptual and quantitative problem solving are not automatically intertwined. Several studies have shown a mismatch between students' overall quantitative course performance and their conceptual understanding, which hints at students potentially using algorithmic approaches in solving physics problems rather than integrating physics and math and engaging meaningfully with the underlying concepts [35,36]. For example, in an early study, Peters found that even high-performing Honors physics students had conceptual difficulties similar to those of students in standard introductory physics courses [37]. Another study similarly found a minimal correlation between quantitative problem solving and conceptual performance on the FCI for beginning physics students [38].

Considering this distinction, several diagnostic tools have been developed to specifically measure conceptual learning in electricity and magnetism courses [23,39–42]. Among them is the Conceptual Survey of Electricity and Magnetism (CSEM), which has become one of the standard tools for investigating conceptual learning gains and identifying student difficulties in electricity and magnetism courses, both at introductory and advanced levels [24,43,44]. Some studies have used the CSEM or similar instruments to examine whether factors such as prior mathematical preparation or instructional approach are associated with conceptual gains in electricity and magnetism courses. For instance, Meltzer [45] found a significant positive correlation between normalized learning gain on the Conceptual Survey of Electricity and students' preinstruction mathematical skills for the majority of the introductory study sample in an algebra-based physics course. Another study used the brief electricity and magnetism assessment (BEMA) to examine the long-term effects of student-centered instruction; this study compared the scores of students who had completed a reformed freshman course that utilized University of Washington tutorials with those of students who had completed a nontutorial freshman course [25]. The results indicated that in the upper-level electricity and magnetism course, the nontutorial group scored significantly lower on the conceptual test, showing long-term positive effects on conceptual understanding from using tutorials.

As previously noted, the vast majority of existing research has focused on introductory-level electricity and magnetism courses, with relatively few studies investigating student performance in upper-level electricity and magnetism [10,25,46–48]. Within this body of research, details about the relationship between conceptual and quantitative performance have been sparse. For instance, in the previously discussed longitudinal study by Pollock, the correlation coefficient of the upper-division students' BEMA post-test score to their course grade was reported as relatively high ($\sim$0.57) [25], and in another study, the correlation of CUE (Colorado Upper Division Electrostatics assessment) to course grades was reported as 0.49 [49]. Beyond these studies, the majority of research in upper-level electricity and magnetism has been focused on preinstruction to postinstruction conceptual gains and their relation to the instructional format, with virtually none examining the relationship between conceptual gains and measures of quantitative performance and prior academic background.

D. Current study

In this study, we seek to do a comprehensive examination of the relationship between students' conceptual understanding and quantitative performance in a traditionally taught upper-level electricity and magnetism course. More specifically, we aim to investigate the extent to which various measures of prior preparation, in-course performance, and conceptual learning (as assessed through the CSEM) relate to one another and to students' overall success in upper-level electricity and magnetism.

Rather than testing causal hypotheses, we have adopted a descriptive and correlational approach to characterize how different measures of student performance and background are associated with success in this advanced context. We use a range of measures, such as homework, quizzes, midterm, and final exam scores, that contribute to the overall quantitative performance of students in the upper-level electricity and magnetism course. Additionally, we use metrics of students' academic preparation prior to and during college, such as SAT math scores, high school GPA, and performance in other math or physics-intensive college courses. Finally, using students' preinstructional and postinstructional CSEM scores, we measure conceptual gains from the beginning to the end of the course. Overall, we aim to understand the following patterns for our study sample:

- Students' quantitative course performance.
- Students' conceptual performance and normalized learning gains.
- Students' prior preparation and its relation to quantitative and conceptual outcomes.
- Correlation patterns across conceptual and quantitative measures.

Additionally, we investigate how some of these patterns may vary by gender and prior exposure to electricity and magnetism concepts through Advanced Placement (AP) or introductory college courses. In the following section, further information about our study sample and measures will be provided.

II. METHODOLOGY

A. Course context and participants

Participants in this study were enrolled in the first-semester mandatory upper-level electricity and magnetism course at a large research university in the United States. The textbook used in this course was Introduction to Electrodynamics by Griffiths [50]. The course structure followed a traditional lecture-based style, with two 75-min sessions of instruction per week. Upon entering the course, students were expected to have some background in terms of prior mathematical skills (as determined by corequisites and prerequisites for the course), including proficiency in vector calculus and differential equations. The course focused on solving problems in electrostatics, magnetostatics, and electromagnetic induction in vacuum and matter. Students were given weekly homework assignments and in-class quizzes, which, in addition to midterm and final exams, contributed to their final course grade.

The data related to course measures and conceptual test performances (measured by the CSEM) were provided by the course instructor. This course was taught by the same instructor in the fall semester of two consecutive academic years, which were used for our study. While the course followed a traditional lecture-based style, we consider the course instructor as PER-friendly in terms of showing interest and engaging with colleagues in the field of PER in the Department of Physics and Astronomy, as well as encouraging students to study in groups and work collaboratively on homework problems. The data from the two semesters were anonymized and combined with data obtained from university records via an honest broker, which included additional grade and demographic information used for analysis. Overall, the final sample consisted of 56 observations from students who had data for the CSEM scores both at the beginning and end of the semester. One of the students in this sample repeated the course, but throughout the paper, we treat the two observations for this student as independent data points, as our results effectively remained unchanged with or without removing the repeated observation. The majority of the students in our sample were Physics/Physics and Astronomy majors, with five students majoring in computer engineering, electrical engineering, and engineering science. The demographics indicated a distribution of 70% men and 30% women students, and in terms of racial distributions, the sample consisted of 71% White, 18% Asian, 7% Black/Hispanic/Indigenous, and 4% unknown.

B. Conceptual measures

The Conceptual Survey of Electricity and Magnetism (CSEM) was used as the tool for measuring students' conceptual understanding of the subject [40]. This survey consists of 32 multiple-choice questions on electrostatics, magnetic fields and forces, and Faraday's law. The survey was given to students both at the beginning of the semester, before any extensive instruction (pre-CSEM), as well as at the end of the semester, before taking the final exam (post-CSEM). The overall scores on these surveys were calculated and reported as the percentage of correct answers given. To examine how much each student gained in terms of their conceptual understanding, we used normalized gain as the absolute gain divided by the maximum possible gain as follows:

$$g = \frac{\text{post CSEM score} - \text{pre CSEM score}}{100 - \text{pre CSEM score}}$$

where pre- and post-CSEM scores are given in percentages. This measure was introduced by Hake [51] and accounts for the variance in pre-CSEM scores. Throughout this paper, we have also included raw gain as simply the difference between pre- and post-CSEM scores for completeness. To maintain consistency across measures, normalized gains are reported as percentages (e.g., 0.4 as 40%).

C. Quantitative measures

1. Course performance

Students' quantitative performance in the upper-level electricity and magnetism course was evaluated through the following components: weekly homework assignments, weekly quizzes, midterm, and final exam. A specific number of points was allocated to each of these components, with the highest being assigned to the midterm and final exams. Relatively fewer points were assigned to each homework and quiz assignment individually, despite their cumulative weight over the semester. Overall, we consider the midterm and final exam as the high-stakes components of this course, where together they accounted for approximately 50% of the total course grade. The remaining portion of the students' grades was determined from weekly quizzes and homework assignments, which were individually low-stakes and each contributed about 2% to the course grade. Points that students obtained from each of these measures were added and rescaled on a percentage basis by the instructor, which determined their overall course performance. Throughout the paper, we use this measure for the overall upper-level electricity and magnetism course quantitative performance (which we refer to as the U-E&M variable). Data from university records also provided each student's final course grade on a 4.0 scale (referred to as U-E&M G. variable). We kept both variables for reporting our results to confirm that they measure the same outcome, despite using different scales and slight differences in distributions.

2. Prior preparation

We used students' grades in some of the prior courses they had taken in college (mainly prerequisite courses) as a way to capture their academic performance in college and the mathematical skills associated with them. These college courses, with their grades being on a 4.0 scale, included calculus 2 (Calc. 2 G.), calculus 3 (Calc. 3 G.), and upper-level mechanics (U-Mech. G.). Some students may have taken Advanced Placement (AP) credits, which exempted them from being required to take some of these courses. Therefore, sample sizes vary for these metrics. (Note that we did not include students' grades in the introductory physics course as one of the metrics because we lacked such data for most of the students in our sample.)

SAT math scores and high school GPA (referred to as HS GPA variable) were also obtained from university records. This information was available for the majority of students in the sample. In the U.S. education system, the SAT is a standardized college admission exam, and its Math section is designed to assess students' quantitative reasoning skills using algebra and advanced math. The SAT Math section is scored on a 200–800 point scale. According to College Board benchmarks, a score of 530 or higher indicates college and career readiness, with a 75% probability that students will earn at least a C in first-semester, credit-bearing college courses in calculus or precalculus. High school GPAs were on a 0–5 scale; this scale reflects the use of weighted grading systems at some U.S. high schools, where advanced coursework may contribute additional grade points. In our sample, one student had a GPA exceeding 5 due to the specific weighting practices of their high school. Additionally, some students had taken the ACT exam instead, and their scores were converted to SAT-equivalent scores using established concordance tables provided by the College Board and ACT [52].

We also investigated students' prior experience with learning electricity and magnetism concepts through data on their AP physics background and introductory physics courses in college. We identified students who had taken AP Physics C: Electricity and Magnetism in high school, which is a calculus-based physics course for students interested in majoring in physical science or engineering. AP exams are scored on a 1–5 scale, where higher scores indicate stronger performance. Many U.S. colleges, including our studied institution, grant college credit for these AP courses, allowing students to skip equivalent college courses for grades of 3 and above, though specific policies may vary by institution. We also checked to see which of the students in the sample had taken the regular versus honors version of introductory physics 2 (focused on electricity and magnetism topics). Within this university, the honors version of the introductory electricity and magnetism course generally follows the same content but spends more time on in-class problem solving than lecture and uses more challenging practice and exam problems. Since taking this version of the course is optional, there is a possibility of self-selection of the student sample in honors classes.

D. Analysis

For descriptive statistics, we used Stata v18 to examine students' overall performance and the distribution of the metrics. We used Cohen's d as a measure for the effect size of differences across groups with varying performances [53]. Unpaired t tests were used to determine the level of significance of effect sizes for normally distributed data and Wilcoxon rank-sum tests for metrics that deviated from normal distribution [54]. Pearson correlations were calculated and plotted as a heatmap in RStudio using the GGPLOT2 package [55]. Regression analyses were carried out in Stata to report standardized regression coefficients (β) and adjusted R^2 values indicating the variance of the outcome explained by the models. For each outcome variable, we initially established baseline models using a single predictor and then explored multiple-predictor models with additional variables to observe changing effects. Throughout, we use $p < 0.001$, $p < 0.01$, and $p < 0.05$ as three different levels of statistical significance. Descriptive statistics and Pearson correlations were also calculated separately for men and women. However, considering the small sample size, the results need to be interpreted with this limitation in mind. Therefore, gender subgroup analyses, along with metric distributions and additional scatter plots, are presented in the Appendix.

III. RESULTS

Table I summarizes the descriptive statistics for student performance metrics. Normalized gains (average of individual normalized gains) and effect sizes in Table I are calculated before rounding off the values. There were no instances of negative normalized gain in our dataset.

In order to examine differences in performance metrics more closely, students were separated into top-half and bottom-half scorers on the CSEM pretest based on whether they were above or below the median pretest score of 55%. (Top half: $N = 27$, mean score = 73%; Bottom half: $N = 29$, mean score = 42%.) Additionally, top-half and bottom-half scorers were analyzed separately based on whether they showed above-average (AA) or below-average (BA) normalized gains with respect to their own group. The comparisons for these groups are presented in Tables II and III. Outcomes of various regression models used to predict course performance (course performance = "Upper E&M," or "U-E&M"), CSEM normalized gain, and CSEM post-test score are shown in Tables IV–VI, respectively. Comparisons between men's and women's performances are shown in Table VII. Performance comparisons for students who had, or had not taken AP courses, as well as those who had taken the honors versus regular version of introductory electricity and magnetism, are shown in Tables VIII and IX.

We investigated how the various precourse, CSEM, and course-performance measures were correlated with each other. Figure 1 shows a heatmap of two-point correlations across different performance metrics used in our study. Significant correlation coefficients are in bold font, and the levels of significance and number of observations are also provided. In the Appendix, cross-comparison of groups by prior electricity and magnetism preparation, as well as individual-measure histograms are provided (Figs. 2 and 3). Additional heatmaps for the various subgroups and scatterplots showing the exact distribution of results for various measure-by-measure comparisons can also be found in the Appendix (Figs. 4–19).

Our primary results are outlined below.

TABLE I. Descriptive statistics for student performance metrics. Sample size (N), mean, and standard deviation (SD) are reported. Effect sizes measure differences between top- and bottom-half scorers on the CSEM pretest; positive values favor the top half ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	All students			Top half			Bottom half			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	Effect size
Pre-CSEM	56	57%	20%	27	73%	13%	29	42%	11%	**2.7*******
Post-CSEM	56	74%	17%	27	86%	10%	29	64%	17%	**1.6*******
Normalized gain	56	42%	30%	27	45%	28%	29	40%	22%	0.2
Raw gain	56	17%	11%	27	13%	9%	29	22%	12%	**−0.9*******
HW	56	89%	11%	27	88%	12%	29	90%	11%	−0.2
Quiz	56	77%	14%	27	80%	14%	29	74%	13%	0.4
Midterm	56	68%	20%	27	77%	13%	29	59%	22%	**1.0*******
Final	56	71%	23%	27	85%	12%	29	57%	22%	**1.5*******
U-E&M	56	77%	11%	27	83%	9%	29	73%	11%	**1.0*******
U-E&M G.	56	3.1	0.7	27	3.4	0.6	29	2.8	0.7	**1.0****
HS GPA	55	4.0	0.6	26	4.1	0.5	29	4.0	0.7	0.3
SAT math	52	730	61	26	764	34.5	26	696	64	**1.3*******
Calc. 2 G.	27	3.1	0.9	13	3.5	0.8	14	2.7	0.8	**1.1****
Calc. 3 G.	50	3.4	0.7	26	3.7	0.4	24	3.1	0.9	**0.9****
U-Mech. G.	49	3.0	0.7	24	3.3	0.6	25	2.8	0.7	**0.9****

A. Overall course performance and prior preparation

1. Students' scores on the conceptual test increased significantly over the course of traditional instruction in upper-level E&M

Students' raw scores on the CSEM increased by an average of 17% from pretest to post-test (mean pretest: 57%, mean post-test: 74%), corresponding to a large effect size ($d = 0.9$; $p < 0.001$). The average normalized gain was 42%, reflecting strong conceptual improvement among students in this traditionally taught course.

2. Students in this sample had average high school GPAs (4.0/5.0) and SAT math scores (730) that were far above national averages

For example, 730 on the SAT Math test corresponds to the 97th percentile on a "nationally representative" sample, according to the College Board [56].

TABLE II. Comparison of student performance metrics for bottom-half scorers with above-average (AA, $N = 13$) or below-average (BA, $N = 16$) normalized gain; positive values favor the AA group. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	Above average normalized gain			Below average normalized gain			
	N	Mean	SD	N	Mean	SD	Effect size
Pre-CSEM	13	47%	8%	16	38%	11%	**1.0***
Post-CSEM	13	79%	9%	16	52%	10%	**2.8*******
Normalized gain	13	60%	14%	16	23%	10%	**3.1*******
Raw gain	13	32%	8%	16	14%	7%	**2.3*******
HW	13	92%	7%	16	88%	13%	0.4
Quiz	13	76%	13%	16	71%	13%	0.5
Midterm	13	69%	14%	16	51%	24%	**0.9***
Final	13	66%	19%	16	50%	23%	**0.7***
U-E&M	13	78%	9%	16	68%	11%	**0.9***
U-E&M G.	13	3.0	0.6	16	2.5	0.7	**0.8***
HS GPA	13	4.2	0.6	16	3.8	0.7	0.6
SAT math	12	732	40	14	666	65	**1.2***
Calc. 2 G.	5	3.3	0.8	9	2.4	0.6	**1.3***
Calc. 3 G.	10	3.5	0.8	14	2.8	0.8	**0.9***
U-Mech. G.	11	3.1	0.4	14	2.5	0.7	**1.0***

TABLE III. Comparison of student performance metrics for top-half scorers with above-average (AA, $N = 11$) or below-average (BA, $N = 16$) normalized gain; positive values favor the AA group. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	Above average normalized gain			Below average normalized gain			
	N	Mean	SD	N	Mean	SD	Effect size
Pre-CSEM	11	73%	11%	16	73%	14%	0.0
Post-CSEM	11	92%	6%	16	81%	9%	**1.4****
Normalized gain	11	72%	17%	16	26%	14%	**3.1*****
Raw gain	11	19%	8%	16	8%	5%	**1.7*****
HW	11	90%	8%	16	87%	15%	0.3
Quiz	11	86%	9%	16	76%	15%	0.8
Midterm	11	80%	15%	16	76%	12%	0.3
Final	11	84%	16%	16	86%	10%	0.1
U-E&M	11	85%	8%	16	81%	10%	0.5
U-E&M G.	11	3.6	0.5	16	3.3	0.6	0.6
HS GPA	11	4.2	0.5	15	4.1	0.5	0.2
SAT math	11	779	26	15	753	36	0.8
Calc. 2 G.	6	3.9	0.3	7	3.3	1.0	0.8
Calc. 3 G.	11	3.9	0.2	15	3.6	0.5	0.8
U-Mech. G.	10	3.5	0.5	14	3.2	0.6	0.6

TABLE IV. Regression models predicting upper E&M total percentage in the course. Standardized regression (β) coefficients are provided. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and n.s. = not statistically significant.

Model	Pre-CSEM	Post-CSEM	SAT math	Calc. 3 G.	U-Mech. G.	HS GPA	Adjusted R^2	N
Baseline 1	0.55^{***}						0.29	56
(a)	0.42^{**}		$0.22^{\text{n.s.}}$				0.30	52
(b)	$0.19^{\text{n.s.}}$				0.66^{***}		0.60	49
(c)	0.31^{*}		$0.16^{\text{n.s.}}$	0.37^{**}			0.44	48
Baseline 2		0.64^{***}					0.40	56
(a)		0.20^{ns}			0.64^{***}		0.60	49
Baseline 3			0.47^{***}				0.21	52
(a)			$0.12^{\text{n.s.}}$		0.69^{***}		0.57	46
Baseline 4				0.60^{***}			0.34	50
(a)				$0.21^{\text{n.s.}}$	0.68^{***}		0.66	44
Baseline 5					0.77^{***}		0.59	49
(a)					0.72^{***}	$0.11^{\text{n.s.}}$	0.59	44
Baseline 6						0.46^{***}	0.19	55

3. Students' average grade in this upper-level E&M course was similar to their previous grades in calculus and upper-level mechanics courses

The average grade of 3.1 in upper-level electricity and magnetism was close to prior course grade averages in calculus 2, calculus 3, and upper-level mechanics.

4. Students' course grades (i.e., total points, "U-E&M") in upper-level E&M were significantly ($p < 0.001$) correlated with precourse measures

Course grades were significantly correlated with:

a. Pretest scores on the CSEM ($r = 0.55$).
b. Grades in calculus 2 ($r = 0.85$), calculus 3 ($r = 0.60$), and upper-level mechanics ($r = 0.77$).
c. SAT math score ($r = 0.47$).
d. High school GPA ($r = 0.46$).

CSEM pretest scores were significantly correlated with SAT math scores ($r = 0.62$), and when both were used in a regression model to predict U-E&M, only the CSEM pretest score was significant. There appears to be a "threshold" effect in that *none* of the 10 students with CSEM pretest scores below 40% achieved an overall course grade (U-E&M) over 82%, while 48% of the other 46 students did so (see Fig. 16). Including the grade in calculus 3 added some predictive power to the model (see Table IV).

TABLE V. Regression models predicting normalized gain. Standardized regression (β) coefficients are provided. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and n.s. = not statistically significant.

Model	Post-CSEM	SAT math	Calc. 3 G.	U-Mech. G.	U-E&M	Pre-CSEM	Adjusted R^2	N
Baseline 1	0.59^{***}						0.34	56
(a)	0.70^{***}	$-0.09^{\text{n.s.}}$					0.38	52
(b)	0.53^{**}		$0.08^{\text{n.s.}}$				0.37	50
(c)	0.53^{**}			$0.02^{\text{n.s.}}$			0.26	49
Baseline 2		0.38^{**}					0.12	52
(a)		0.37^{*}	$0.25^{\text{n.s.}}$				0.23	48
(b)		$0.20^{\text{n.s.}}$		0.36^{*}			0.22	46
(c)		$0.23^{\text{n.s.}}$			0.32^{*}		0.19	52
(d)		0.34^{*}			0.40^{*}	$-0.26^{\text{n.s.}}$	0.21	52
Baseline 3			0.41^{**}				0.15	50
(a)			$0.11^{\text{n.s.}}$	0.42^{*}			0.21	44
(b)			$0.16^{\text{n.s.}}$		0.42^{**}		0.25	50
Baseline 4				0.37^{**}			0.12	49
Baseline 5					0.38^{**}		0.13	56
Baseline 6						$0.13^{\text{n.s.}}$	0.00	56

TABLE VI. Regression models predicting post-CSEM score. Standardized regression (β) coefficients are provided. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and n.s. = not statistically significant.

Model	Pre-CSEM	SAT math	Calc. 3 G.	U-Mech. G.	U-E&M	Adjusted R^2	N
Baseline 1	0.82^{***}					0.67	56
(a)	0.65^{***}	0.26^{*}				0.70	52
(b)	0.61^{***}	0.28^{**}	0.20^{**}			0.81	48
(c)	0.75^{***}		0.25^{**}			0.78	50
(d)	0.68^{***}			0.28^{**}		0.75	49
(e)	0.60^{***}	$0.17^{\text{n.s.}}$		0.26^{**}		0.78	46
(f)	0.67^{***}				0.27^{**}	0.71	56
(g)	0.55^{***}	0.21^{*}			0.24^{**}	0.73	52
Baseline 2		0.67^{***}				0.44	52
(a)		0.60^{***}	0.34^{**}			0.59	48
(b)		0.43^{**}		0.42^{**}		0.56	52
(c)		0.47^{**}			0.42^{***}	0.57	52
(d)		0.48^{***}			$0.17^{\text{n.s.}}$	0.59	48
Baseline 3			0.57^{***}			0.31	50
(a)			$0.14^{\text{n.s.}}$	0.60^{***}		0.45	44
(b)			$0.22^{\text{n.s.}}$		0.59^{***}	0.52	50
Baseline 4				0.66^{***}		0.42	49
(a)				0.44^{*}	$0.28^{\text{n.s.}}$	0.44	49
Baseline 5					0.64^{***}	0.40	56

5. The single most significant predictor of course grades in upper-level E&M was the grade in upper-level mechanics

When the upper-level mechanics grade was included in a regression model, other factors such as CSEM pretest score, SAT math score, and calculus 3 grade were no longer significant predictors. Overall, upper-level mechanics grade independently explained 59% of the variance in upper-level electricity and magnetism course performance (see Table IV).

6. Students with previous exposure to AP physics courses had slightly better course outcomes than those without such exposure

Students with an AP physics background had a mean course grade (U-E&M) of 84% ($N = 13$), while those

TABLE VII. Descriptive statistics for student performance metrics by gender. Sample size (N), mean, and standard deviations (SD) are reported. Effect sizes measure performance differences between men and women. Positive effect sizes favor men ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	Men			Women			
	N	Mean	SD	N	Mean	SD	Effect size
Pre-CSEM	39	62%	18%	17	45%	17%	**1.0****
Post-CSEM	39	79%	16%	17	64%	17%	**0.9****
Normalized gain	39	44%	26%	17	37%	23%	0.3
Raw gain	39	17%	11%	17	19%	12%	−0.2
HW	39	89%	11%	17	90%	13%	−0.1
Quiz	39	78%	13%	17	74%	15%	0.4
Midterm	39	69%	21%	17	66%	20%	0.2
Final	39	72%	23%	17	67%	22%	0.3
U-E&M	39	78%	12%	17	76%	11%	0.2
U-E&M G.	39	3.1	0.7	17	3.0	0.7	0.2
HS GPA	38	4.0	0.6	17	4.2	0.5	−0.4
SAT math	35	743	57	17	704	64	**0.7***
Calc. 2 G.	19	3.2	0.9	8	2.8	0.7	0.5
Calc. 3 G.	33	3.5	0.8	17	3.3	0.7	0.2
U-Mech. G.	33	3.1	0.7	16	2.9	0.7	0.3

TABLE VIII. Comparison of performance among students with and without Advanced Placement physics background. Sample size (N), Mean, and standard deviations (SD) are reported. Positive effect sizes favor those with AP preparation. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	AP score available			No AP information			
	N	Mean (%)	SD (%)	N	Mean (%)	SD (%)	Effect size
Pre-CSEM	13	72	17	43	52	18	**1.1*****
Post-CSEM	13	86	11	43	71	17	**1.0****
Normalized gain	13	48	32	43	40	23	0.3
U-E&M	13	84	11	43	75	11	**0.8***

without such background had a mean grade of 75% ($N = 43$), a statistically significant difference. A similar small outcome advantage was noted for students who had taken the "honors" section of the introductory physics course instead of the "regular" section (see Tables VIII and IX).

TABLE IX. Comparison of performance among students with honors versus regular introductory physics college background. Sample size (N), Mean, and standard deviations (SD) are reported. Positive effect sizes favor those who took the honors version of the introductory physics 2 course ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Bold values indicate statistically significant effect sizes.

	Honors			Regular			
	N	Mean (%)	SD (%)	N	Mean (%)	SD (%)	Effect size
Pre-CSEM	25	59	20	26	55	19	0.2
Post-CSEM	25	78	17	26	72	17	0.3
Normalized gain	25	49	24	26	38	25	0.5
U-E&M	25	82	10	26	73	10	**0.8****

7. The degree of over- or under-performance in upper-level E&M compared to upper-level Mechanics was uncorrelated with pretest scores on the electricity and magnetism conceptual test

Scatterplots of "change in grade" ΔG ($\Delta G =$ U-E&M Grade—U-Mech. Grade) versus pretest scores on the CSEM show no significant correlations; see Fig. 12.

B. Performance on the CSEM conceptual test and prior preparation

1. Students' pretest scores on the CSEM were significantly correlated ($p < 0.001$) with SAT math scores ($r = 0.62$) and grade in upper-level mechanics ($r = 0.55$)

Top-half scorers on the CSEM pretest had significantly higher scores on the SAT math test than did bottom-half

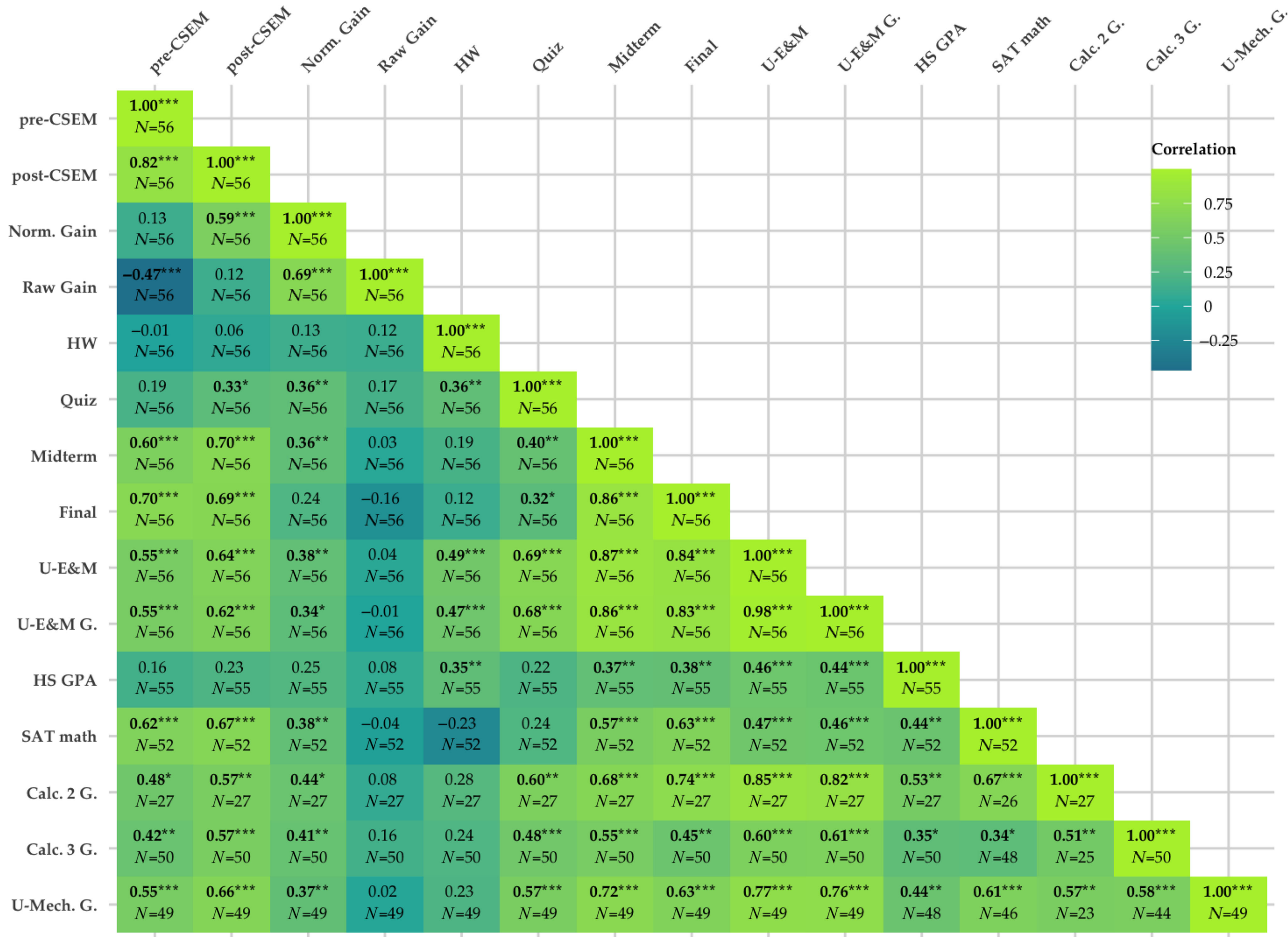

FIG. 1. Correlation matrix of student performance metrics. Significant correlations are shown in bold fonts. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

scorers (764 versus 696; effect size 1.3; $p < 0.001$), as well as higher grades in related courses such as calculus 2, calculus 3, and upper-level mechanics (effect size: ≈ 1.0; $p < 0.01$).

2. Pretest scores on the CSEM were significantly higher for students who had taken Advanced Placement physics, but not for those who had taken the "Honors" instead of the "Regular" section of the introductory electromagnetism course

The relatively small group of students ($N = 13$) who had completed the AP Physics C: Electricity and Magnetism course in high school had higher CSEM pretest scores (72%) than those *without* AP course grades (52%; $N = 43$): effect size $= 1.1$, $p < 0.001$. This significant difference carried over to CSEM post-test scores ($p < 0.01$), but the normalized gains of the two groups did not differ significantly. There were no significant differences in the conceptual test performance of students who had taken the regular or honors version of introductory E&M in college. Detailed comparisons are shown in Tables VIII and IX, and Fig. 2.

3. Students' pre- and post-test scores on the CSEM were highly correlated, suggesting a minimum threshold effect

Students' postinstruction scores on the CSEM were very highly correlated with preinstruction scores ($r = 0.82$, $p < 0.001$). As was the case with the overall course grade (U-E&M), there appears to be a "threshold" effect in that *none* of the 10 students with pretest scores below 40% achieved a post-test score over 63%, while 87% of the other students did so (see Fig. 13).

4. Students' post-test scores on the CSEM were largely predictable based on their pretest scores, but grades in previous calculus and upper-level mechanics courses added predictive power

Regression models predicting post-CSEM scores are shown in Table VI. The predictive power of models including pre-CSEM was substantially improved by adding either grades in calculus 3 or upper-level mechanics. By comparison, models that relied solely on other course grades *without* including pre-CSEM accounted for much less variance in the CSEM post-test scores.

5. Students' normalized gains on the CSEM were significantly correlated ($p < 0.01$) with several precourse measures

Correlations between normalized gain on the CSEM and various precourse measures were in the 0.37–0.41 range. Specifically, normalized gain correlated significantly ($p < 0.01$) with SAT math scores ($r = 0.38$), calculus 3 grades ($r = 0.41$), and upper-level mechanics grades ($r = 0.37$). However, these correlations were *only* significant for the group with below-median CSEM pretest scores ("bottom half" group; $r \approx 0.50$), and not for the above-median ("top-half") group. For example, bottom-half scorers on the CSEM pretest with above-average (AA) normalized gains had much higher SAT math scores in comparison to bottom-half scorers with below-average (BA) normalized gains (732 versus 666: effect size: 1.2; $p < 0.05$). (By contrast, for the top-half group, the difference in SAT scores and previous course grades for those with above- and below-average normalized gains was not significant.) Analogously, correlations between normalized gain and both SAT math scores and upper-level mechanics grades were significant for women ($r = 0.58$ and $r = 0.65$, respectively; $p < 0.05$ and $p < 0.01$) but not for men ($r \approx 0.21$, n.s.).

6. Students' normalized gains on the CSEM were not significantly correlated with pretest scores on the CSEM

Students' normalized gain was completely independent of pre-CSEM scores (r = 0.13, not significant), with pre-CSEM explaining none of the variance in normalized gain. This was further confirmed using regression models, which are presented in Table V. The normalized gain of top-half scorers on the pretest (45%) was not significantly different from that of bottom-half scorers (40%), as one could expect from the fact that top-half scorers on the CSEM pretest had significantly lower raw (absolute) gain on the CSEM than bottom-half scorers (13% versus 22%; effect size: 0.9; $p < 0.001$).

C. Relationship between course performance and performance on the CSEM conceptual test

1. Students' course grades (i.e., total points, "U-E&M") in upper-level E&M were significantly ($p < 0.001$) correlated with their CSEM pretest score ($r = 0.55$)

As mentioned in Sec. III A 4 above, students' course grades in upper-level electricity and magnetism were strongly correlated ($r = 0.55$) with their CSEM pretest scores. Top-half scorers on the CSEM pretest had significantly higher final exam scores (85%) and course grades (83%) than bottom-half scorers (57% and 73%, respectively; effect sizes: 1.0–1.5, $p < 0.001$). Specifically, as can be seen from the scatterplot in the Appendix (Fig. 16, top row), about 70% of the students who had relatively high (>60%) pretest scores on the CSEM also had high (>80%) final course grades, while less than 10% of the students with low (<40%) pretest scores achieved that same course grade performance level.

2. Students' post-test scores on the CSEM were strongly correlated ($p < 0.001$) with measures of course performance

Students' post-test scores on the CSEM were strongly and significantly correlated ($p < 0.001$) with both final exam score ($r = 0.69$) and overall course grade ($r = 0.64$). As can be seen from the scatterplots in the Appendix (Fig. 13), about 70% of the students who had high (>80%) post-test scores on the CSEM also had high (>80%) course grades, while less than 10% of the students with low (<60%) post-test scores achieved that same course grade performance.

3. Students' normalized gains on the CSEM were weakly correlated with measures of course performance

Normalized gain showed a significant correlation ($p < 0.01$) with overall course performance "U-E&M" ($r = 0.38$) and with some other quiz and exam scores, but the correlation with final exam score was nonsignificant. However, the correlation between normalized gain and course performance was *only* significant for the group with below-median pretest scores ("bottom-half" group; $r = 0.46$), and not for the above-median ("top-half") group. Bottom-half scorers with above-average (AA) normalized gains had higher course performance scores in upper-level electricity and magnetism than did those with below-average (BA) normalized gains (effect size = 0.9; $p < 0.05$). Analogously, the correlation between normalized gain and course performance was *only* significant for women ($N = 17$; $r = 0.62$, $p < 0.01$) and not for men ($N = 39$; $r = 0.28$, n.s.). In this regard, it is essential to note that women were heavily overrepresented in the bottom-half group, as their mean CSEM pretest score (45%) was far lower than that of the men (62%); this difference in pretest scores was statistically significant despite the small sample sizes (effect size = 1.0; $p < 0.01$). (71% of women were in the bottom-half group, while only 44% of men fell in that group.)

As mentioned in Sec. III A above, we had also defined a rough measure of change in quantitative problem-solving performance (ΔG = Upper E & M Grade − Upper Mechanics grade) to capture shifts in students' quantitative performance across these two advanced courses. As shown in Fig. 12, this measure was not significantly correlated with either normalized or raw gain on the CSEM, suggesting that improvements in conceptual scores were largely independent of this measure of variation in quantitative performance in upper-level courses.

D. Gender differences on precourse and in-course measures

1. Large, statistically significant differences were observed between men and women on some precourse measures

Comparisons across men's and women's performances are shown in Table VII. Large, statistically significant differences were observed between men ($N = 39$) and women ($N = 17$) on the CSEM pretest: men scored significantly higher than women (means: 62% versus 45%, effect size = 1.0, $p < 0.01$). (As expected, this significant difference carried over to the CSEM post-test; effect size = 0.9, $p < 0.01$). Men also had significantly higher SAT math scores (means: 743 versus 704, effect size = 0.7, $p < 0.05$). However, high school GPA and prior college course grades showed no significant gender differences.

2. There were no statistically significant differences between men and women on course performance measures

Performances of men and women did not differ significantly on homework, quizzes, midterm exam, final exam, or overall course grade.

3. Correlation between normalized gain and both course performance and pre-course measures was higher for women than for men

Figures 4 and 5 in the Appendix provide heatmaps (similar to Fig. 1) for correlations between variables for men and women separately. As noted in Secs. III B and III C above, women's normalized gain correlated significantly with some precourse measures and with overall performance in the course, while this was, for the most part, not true for men.

IV. DISCUSSION

Although there has been an increasing focus on upper-level courses in physics education research, very little attention has been paid to preinstruction characteristics and indicators that might be associated with success in those courses. Various conceptual assessments have been administered to upper-level students, and new assessment instruments have been developed for that purpose; however, there is relatively little research on the relationship between performance on those conceptual assessments and course outcomes as measured by course grades. The present study aimed to fill in both of those gaps by examining in a systematic fashion the relationship between course grades in upper-level electricity and magnetism and both prior preparation and performance on a conceptual test.

A. Role of prior preparation in course outcomes

In general, we find that students with stronger prior preparation on a variety of measures had better course and conceptual outcomes. This is aligned with previous findings that show prior preparation is related to course outcomes in introductory physics [19,57,58]. More specifically:

(1) Significant correlations between upper-level electricity and magnetism course grades and both precollege mathematical preparation (SAT math) and college calculus grades exemplify the positive association between mathematical skills and higher grades in upper-level electricity and magnetism. (This positive correlation is substantially larger for the college calculus courses.) The positive association between course performance and prior mathematical preparation also shows a consistency between our results in the upper-level electricity and magnetism context and other recent work focused on introductory electricity and magnetism courses [57,59]. This association is not surprising since mathematical skills can facilitate success on quantitative problem-solving tasks in both lower-level and upper-level electricity and magnetism.

(2) The similarity of students' course grades in upper-level electricity and magnetism to their grades in previous upper-level math and physics courses suggests that recent college grades may overall be the best measure for predicting student performance in upper-level electricity and magnetism. This consistency in performance across courses, which has been previously studied among physics major students [60], suggests that characteristics that support success in one course, such as problem-solving skills or consistent study habits, can be generalizable to other courses.

(3) Students' grades in upper-level mechanics are significantly predictive of their grades in upper-level electricity and magnetism even *after* controlling for general math preparation. This result, though perhaps unsurprising, does not appear to have been reported previously. This predictive power suggests that the effect of many potentially relevant variables (such as prior preparation, students' attitudes toward learning, and logical reasoning skills) is already captured by their upper-level mechanics performance. Prior studies on students' mental models suggest that there are overlapping skills and concepts (e.g., problem solving and mathematical reasoning) that are carried over by students from their mechanics to their electricity and magnetism courses [61].

(4) Students' performance on the preinstruction conceptual test (pre-CSEM) was positively correlated with their overall course performance, but perhaps even more striking was the fact that only one of the ten students whose pre-CSEM score fell below 40% had a final course grade above the class median.

(5) Students with prior exposure to electricity and magnetism concepts through Advanced Placement courses or the honors version of the introductory electricity and magnetism courses showed a somewhat

better overall performance in upper-level electricity and magnetism than students without AP exposure or than those who had taken the "regular" version of the introductory course.

There are a number of possible explanations for the positive association between course performance and prior preparation, and they are not mutually exclusive. For example, superior performance in upper-level electricity and magnetism may be associated with (a) a reduced cognitive load during learning through prior preparation; (b) a commonality of factors related to success on both precourse measures and in upper-level electricity and magnetism (for example, a factor such as "general intelligence"), even if there is no direct *causal* relationship between prior preparation and performance in the electricity and magnetism course; (c) selection effects, whereby higher-performing students are more likely to enroll in AP or honors courses or to take the upper-level electricity and magnetism course, even if the AP course or their prior math and conceptual preparation did not *directly* impact their success in upper-level electricity and magnetism.

Previous studies on students' difficulties in electricity and magnetism show that students tend to struggle with abstract concepts and mathematical representations in these courses [12,13,15], which lends support to the idea that better prior preparation and greater availability of cognitive resources can be important for success in this course. Based on cognitive load theory, unfamiliar mathematical notations or too many abstract or new concepts can increase cognitive load and thereby hinder learning in upper-level courses [62]. However, it is important to emphasize that nothing in our findings constitutes evidence that improving scores on preinstruction math or conceptual tests would lead *directly* to improved performance in upper-level electricity and magnetism.

B. Patterns in conceptual outcomes

Compared to some of the previously reported conceptual gains in the context of upper-level electricity and magnetism courses, the average normalized gain on the CSEM was relatively high (42%) for our sample. Research using the CSEM with matched upper-level students in traditional courses has found normalized gains of approximately 31%, which is lower than the normalized gain we observed here [24]. Other studies using different conceptual instruments, such as the CUE test, have reported similar normalized gains in upper-level courses (~30%) [10,47]. The relatively greater normalized gains we observe here may reflect positive selection effects (students had exceptionally high SAT math scores at the 97th percentile), potentially better-integrated conceptual and quantitative instruction, or the characteristics of the particular cohort studied. While the average normalized gain exceeded expectations for a traditionally taught course, a closer examination provides further insights. Namely:

(1) As the high correlation between pre- and post-CSEM scores suggested, the overall students' ranking in terms of conceptual knowledge did not change much during the upper-level electricity and magnetism course. In fact, the threshold effect shows that students entering the course with the weakest conceptual preparation were uniformly unable to achieve high scores on the post-test. This may suggest that traditional instruction may not adequately remediate preexisting large knowledge gaps across students. Prior research has shown that upper-level electricity and magnetism students do tend to struggle with the same concepts that introductory level students often answer incorrectly on their post-tests [24].

We did find that the normalized gain on the CSEM for our sample was, on average, independent of students' initial performance on the conceptual survey. (Meltzer reported analogous results for an introductory interactive engagement course [45].) Normalized gain represents the proportion of initial (preinstruction) errors that the student corrects on their post-test; thus, this finding indicates that the majority of students in this course were able to make meaningful conceptual improvements regardless of their initial level, but not to the extent that post-test scores for more weakly prepared students were comparable to those of more strongly prepared students.

(2) Post-test scores and gains on the conceptual test were significantly and positively correlated with SAT math (preinstruction) scores. This result is reminiscent of Meltzer's previous finding that normalized gain on the Conceptual Survey in Electricity for students in an introductory electricity and magnetism course was strongly related to preinstruction scores on a mathematics diagnostic test [45]. It should be emphasized that skills measured on the SAT math test have little direct overlap with questions on the (largely nonquantitative) conceptual test, so there is no obvious explanation for this correlation; certainly, there would not appear to be any direct causal connection. It is, perhaps, conceivable that stronger mathematical preparation served as an aid for reducing the cognitive load required for making progress on the quantitative course tests, thus helping to enable stronger conceptual gains. However, at least as likely is that the same background and mental characteristics that enabled students to score high on the SAT math test also enabled them to make strong gains on the conceptual test. In addition, consistent with frameworks that emphasize deliberate integration of conceptual and quantitative learning, it is possible that a stronger mathematical preparation supported more balanced outcomes across both domains [63].

(3) Post-test scores and gains on the conceptual test were also significantly and positively correlated with course performance in upper-level electricity and magnetism. (The degree of correlation was higher for the group with lower pretest scores; see discussion in Sec. IV C 2 below.) It is not clear whether good performance or high gain on the conceptual test directly influenced (or was influenced by) success in the upper-level electricity and magnetism course. That is, we are unable to say whether the larger normalized gains and/or higher post-test scores were directly (causally) tied to doing better in the course.

On the one hand, it could be that instruction in this upper-level electricity and magnetism course was successful at helping the more successful students effectively integrate quantitative and conceptual aspects of the course content. This hypothesis would align with research findings that focus on the structural role of mathematics in physics and the entanglement of conceptual and quantitative understanding in physics problem solving [26,27,33,64]. However, it is also possible that the same preinstruction background and characteristics that enabled students to make good conceptual learning gains also enabled them to achieve good performance on the quantitative course exams.

(4) Correlations between normalized gain on the CSEM and various precourse measures were significant for bottom-half scorers on the CSEM but not the top-half scorers. More specifically, the degree to which bottom-half scorers on the CSEM pretest were able to make large normalized gains was positively correlated with their prior mathematical preparation. Bottom-half students with higher SAT math scores had significantly larger normalized gains on the CSEM, did better on the high-stakes assessments, and overall, performed better in the course. This pattern was not observed for top-half scorers,

There are several possible explanations for this finding, including the following: (a) The range of SAT math scores for the top-half group was much smaller—about half of that of the lower-half group, and thus the absence of significant correlation might be an artifact of a faulty, insufficiently broad sample (see scatterplots in Figs. 17 and 18); (b) it may be that above a certain minimum threshold of precourse mathematics preparation, improved mathematics preparation does not contribute to improved gains in conceptual performance; (c) it may be that, above a certain minimum threshold of preinstruction conceptual knowledge, improvements in conceptual performance are unrelated to mathematical skill. Specifically, for students entering the course with weaker conceptual foundations, a stronger mathematical background may have facilitated progress by reducing cognitive load in problem-solving tasks, while for students with strong conceptual foundations, any such reductions in cognitive load may have been insignificant.

Research on problem-solving expertise shows that expert physicists opportunistically blend conceptual and formal mathematical reasoning while solving problems, whereas novices treat equations as purely computational tools [64]. For students with a weaker conceptual background, stronger mathematical skills may enable an easier transition toward more expertlike reasoning that integrates mathematics with conceptual understanding. Strong mathematical skills may position these students within what is known as a zone of proximal development for learning, where some scaffolding can effectively bridge their existing knowledge to more abstract concepts, while students without such preparation may fall outside this zone and require more intensive support [65,66]. Conversely, students with a stronger conceptual background may be less sensitive to the effects of varied mathematical skills.

(It is interesting to note that Meltzer and King [57] reported an analogous difference (albeit in the “opposite” direction) in the degree of correlation between course grades and preinstruction scores on a math skills test, such that students who had high scores on a preinstruction test of reasoning skills seemed to have a *higher* correlation between grades and math skills than those with low reasoning scores. It is possible that falling below a certain reasoning skills threshold may have negated any potential advantage of improved math skills in achieving high course grades).

(5) Students who took AP physics (AP Physics C: Electricity and Magnetism) had higher pre- and post-test scores on the conceptual test, compared to students who had not taken AP courses. As previously discussed in Sec. IV A, these differences could be due either to selectivity effects or to actual improved preparation via the AP course (or to a combination of both effects). Our findings are consistent with prior reports that suggest a correlation between taking AP physics and greater conceptual understanding [67]. By contrast, variations in the rigor of college-level introductory electricity and magnetism (honors versus regular sections) showed no such effect on conceptual test scores, although they *were* associated with differences in overall course performance.

C. Conceptual-quantitative interplay and subgroup dynamics

From a broader standpoint, the significant correlations found between conceptual test scores and course performance measures suggest that a better understanding of electricity and magnetism concepts was associated with stronger course grades and/or exam performances. More specifically:

(1) As noted above, we observed strong correlations between pre-post CSEM scores and upper-level electricity and magnetism course grades. Similar and relatively high correlations between conceptual test scores and course grades have been reported in other upper-level electricity and magnetism courses [25]. These correlations may suggest that conceptual mastery is associated with quantitative problem-solving skills. This could arise through different mechanisms. For example, conceptual and quantitative mastery could improve simultaneously through integrated learning, where improved proficiency in one aspect directly aids the other [26,27,63]. It could also be that students coming in with better quantitative preparation can invest more time and effort into enhancing their conceptual skills or vice versa. In the second case, growing conceptual and quantitative skills may not have necessarily happened in an integrated form, but instead, in parallel. At the same time, our exploratory ΔG measure, defined as the difference between upper-level electricity and magnetism and upper-level mechanics course grades, did not show a clear association with either normalized or raw conceptual gains (Fig. 12). This could indicate that changes in quantitative course performance across these two advanced courses were not strongly associated with conceptual level or conceptual improvement in this student sample. Taken together, these patterns are more consistent with the picture that students who begin the sequence with stronger quantitative preparation tend to maintain an advantage, rather than with a uniform, parallel development of conceptual and quantitative skills throughout instruction.

(2) The moderate overall correlation between normalized gain and course performance masked the stronger association between these two variables for specific subgroups. For example, among students with lower pretest scores (bottom half) and women, normalized gains correlated more strongly with course grades. This may be due in part to the somewhat narrower range of course grades of the upper-half group compared to that of the lower half; this could tend to constrain the correlation artificially due to a limited sample. However, it may also (or instead) be the case that gains in conceptual understanding for students with lower initial scores may be tied more closely to improvements in performance on quantitative electricity and magnetism problems, while students with stronger initial preparation may be beyond the point at which measurable conceptual improvement aids (or is aided by) better quantitative performance. (It should be noted that the *absolute* gain in conceptual exam scores for the upper-half group is smaller than that of the lower-half group due to the ceiling effect on post-test scores, thus severely constraining the range of post-test scores for the upper-half group; compare scatterplots in Figs. 14 and 15).

(3) No significant gender differences emerged in quantitative performance in this upper-level electricity and magnetism course. This pattern differs from research in the introductory physics context, where gender gaps in course performance have been observed [68]. On the other hand, the persistence of gender differences on the conceptual survey at the end of the semester follows directly from analogous differences in the conceptual pretest scores and may be related to the preexisting difference in mathematical preparation (SAT math scores). The gender gap on various physics conceptual tests has been previously studied [69–71]. More specifically, Madsen *et al.* found that the weighted average gender difference on electricity and magnetism conceptual tests was 3.7% for pretest scores, 8.5% for post-test scores, and 6% for normalized gain [70]. However, the pattern observed here appears to mainly reflect students' prior conceptual level rather than gender, as men with similarly low pre-CSEM scores tended to show roughly comparable post-CSEM outcomes. Our results are aligned with previous findings in the introductory context that show women's self-assessments of their conceptual problem-solving skills are lower compared to quantitative skills [72]. The absence of gender differences in upper-level electricity and magnetism course grades, despite initial conceptual knowledge disparities, also suggests that women were likely able to compensate for lower initial conceptual preparation, where and when needed, thereby achieving comparable quantitative outcomes.

V. SUMMARY AND CONCLUSION

Our goal for this study was to provide a detailed understanding of student performance patterns in upper-level electricity and magnetism. This was done through a comprehensive examination of several quantitative and conceptual measures for students in a traditional, lecture-based upper-level electricity and magnetism course at a large U.S. university. We examined students' in-course quantitative performance (homework, quizzes, midterms, final, and total course grade) and conceptual performance (pre- and post-instructional scores on the Conceptual Survey of Electricity and Magnetism), as well as prior preparation before beginning college (SAT math scores, high school GPA, and Advanced Placement physics grades) and previous college course grades (both calculus and physics). We then looked carefully at the relationships among these variables, looking for significant correlations, investigating regression models to "predict" students' course grades, and searching for possible interaction effects between and among the variables that potentially could impact students' course performance.

Overall and across multiple measures, students with stronger prior preparation tended to have higher course grades and stronger conceptual outcomes. Precollege math scores, college calculus grades, and grades in recent upper-level physics were all shown to be positively correlated with performance in upper-level electricity and magnetism. In particular, grades in upper-level mechanics were strongly predictive of electricity and magnetism grades even after controlling for mathematical preparation, which suggested that performance in the most recent physics course already captured a broad set of factors related to success for upper-level electricity and magnetism. Additionally, students' scores on the preinstruction conceptual test were significantly correlated with quantitative course performance. In fact, none of the students who started with preconceptual scores below a certain threshold ended up in the higher-end grade range. Moreover, students who had prior exposure to electricity and magnetism through Advanced Placement courses or honors introductory sections showed somewhat better quantitative course outcomes than those without such backgrounds. Many of these results are consistent with previous studies of similar measures in introductory physics courses.

In terms of conceptual learning, we observed a relatively large average normalized gain for this sample, which was substantially higher than previously reported averages at other institutions. However, threshold effects also emerged in conceptual performance. Students who entered the course with the lowest initial conceptual knowledge were uniformly unable to achieve high scores on the postinstruction conceptual assessment, despite having no significant differences in average normalized gains. This threshold effect suggests that traditional instruction, while effective at providing similar normalized learning gains across the sample, may not have adequately bridged the large knowledge gaps present among students at the beginning. In other words, the ranking of students in terms of conceptual knowledge remained relatively stable in this upper-level electricity and magnetism course.

Investigation of the relationship between conceptual test scores and in-course performance suggested a positive association between conceptual mastery of electricity and magnetism topics and better quantitative performance in the course. While this association can emerge through different mechanisms, such as integrated or parallel development of conceptual and quantitative skills, our data seemed to suggest that students maintain their incoming advantage even during this traditional quantitative course. The correlation between conceptual and quantitative performance, i.e., normalized gain and course grade, appeared to be stronger among those with lower pretest scores; this included women as a group who, however, despite initial conceptual disparities, maintained similar course-grade performance as men.

From an instructional standpoint, our findings provide insights for educators teaching upper-level electricity and magnetism courses. First, prior preparation significantly predicting course success argues for intentional scaffolding or prerequisite support for students who start the course with weaker mathematical skills or conceptual backgrounds. Second, the threshold effects observed here suggest that traditional, lecture-based instruction may not be able to close initially large performance gaps. Therefore, within courses such as upper-level electricity and magnetism that require a strong mathematical background, instructors may consider using active-engagement methods or diagnostic tools early in the course to identify the specific needs of students and provide targeted scaffolding, such as brief math refreshers or additional one-on-one support outside of the class as needed. Finally, instructors can foster an integrated teaching approach by incorporating conceptual prompts in their lectures that invite students to go beyond thinking of mathematics as a tool for manipulating equations and solving problems. This integration can also be reinforced through deliberate assessments that include cues for conceptual reasoning, especially on low-stakes assessments such as weekly homework or quizzes. This would provide students with a chance to engage thoughtfully with problems without the pressure of the high-stakes grading that might predispose them to take a "safe" algorithmic approach just to meet minimum performance requirements.

As a final remark, we note that our findings are limited to a small sample of students within a traditional course at a large U.S.-based institution. Therefore, the subgroup analyses presented here are sensitive to any generalization. Additionally, the findings presented here are correlational in nature and do not account for any other possible confounding variables; this precludes any causal interpretation of the results. Future studies may use additional variables to explore similar patterns in performance for larger samples across multiple institutions, especially for students with a more diverse background and preparation.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

APPENDIX

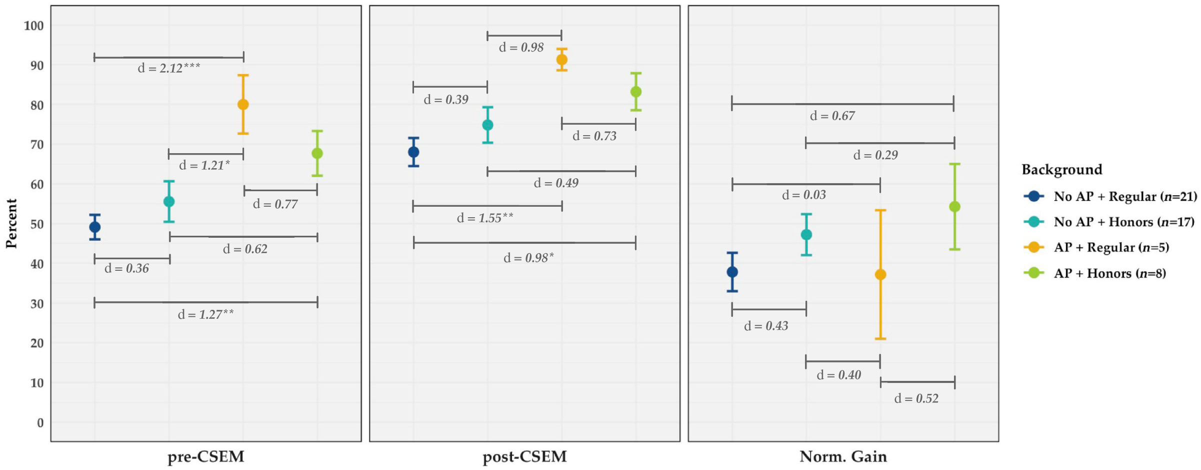

FIG. 2. Comparison of average conceptual performance by type of background in terms of exposure to electricity and magnetism concepts. Pairwise effect sizes and their significance levels are provided. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

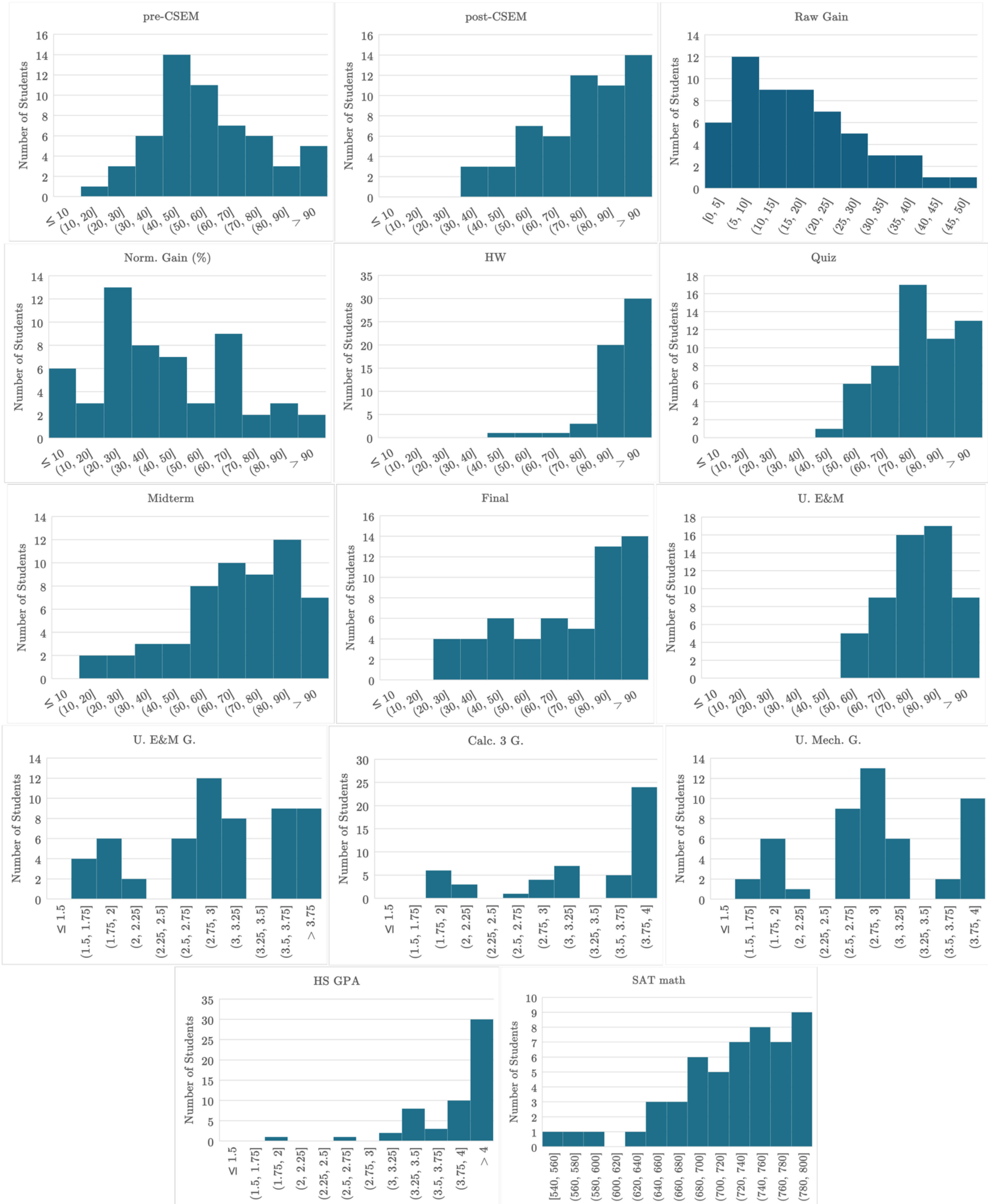

FIG. 3. Histograms of student performance metrics.

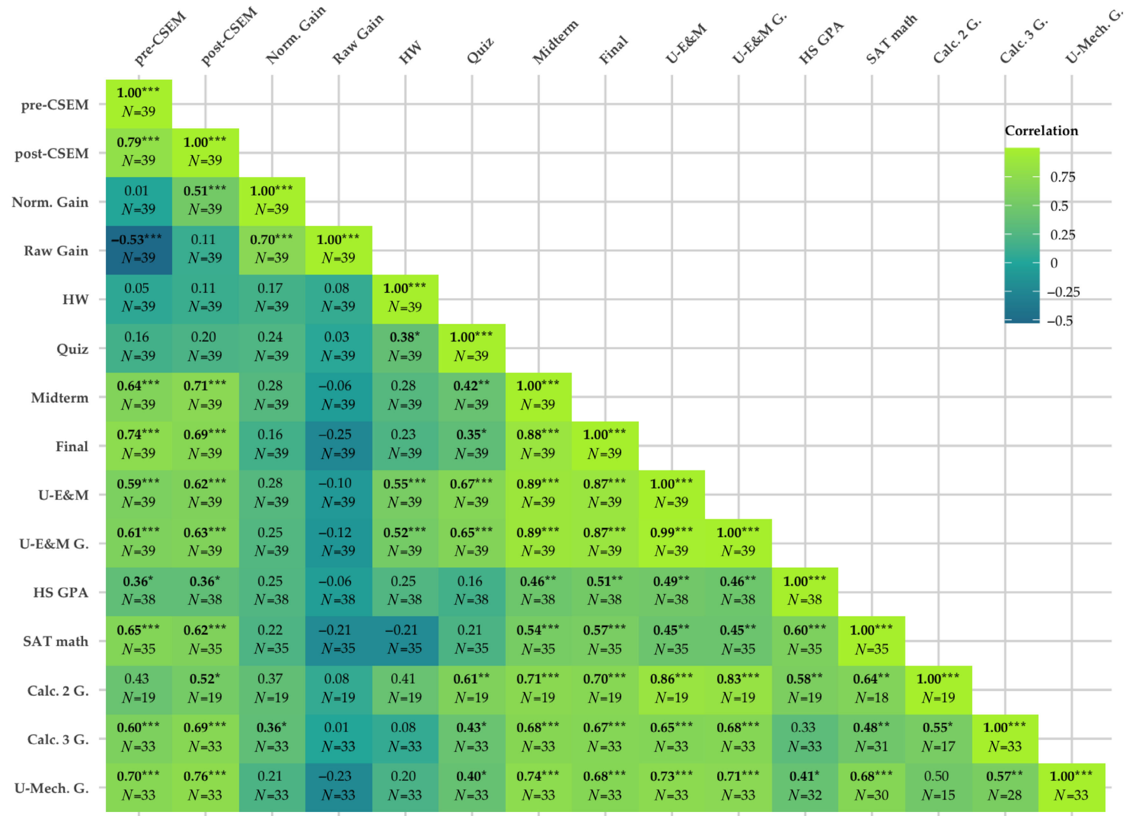

FIG. 4. Correlation matrix of student performance metrics for men. Significant correlations are shown in bold fonts.

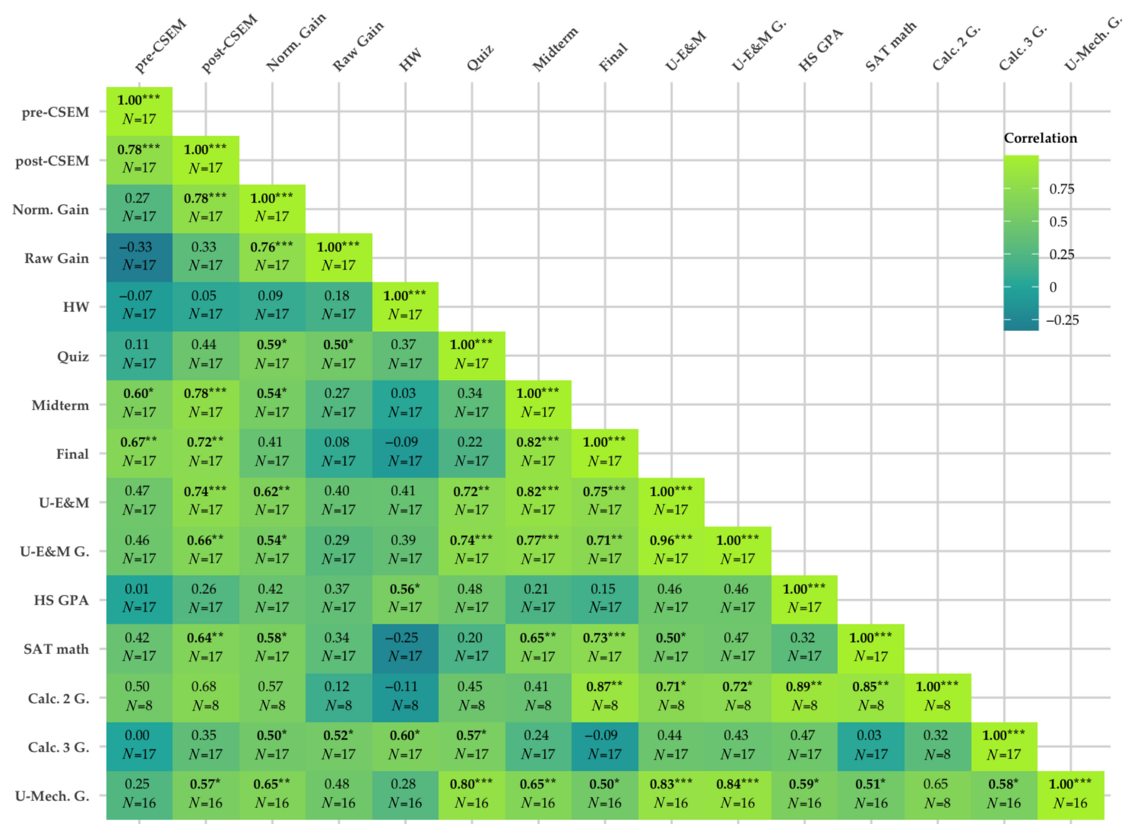

FIG. 5. Correlation matrix of student performance metrics for women. Significant correlations are shown in bold fonts.

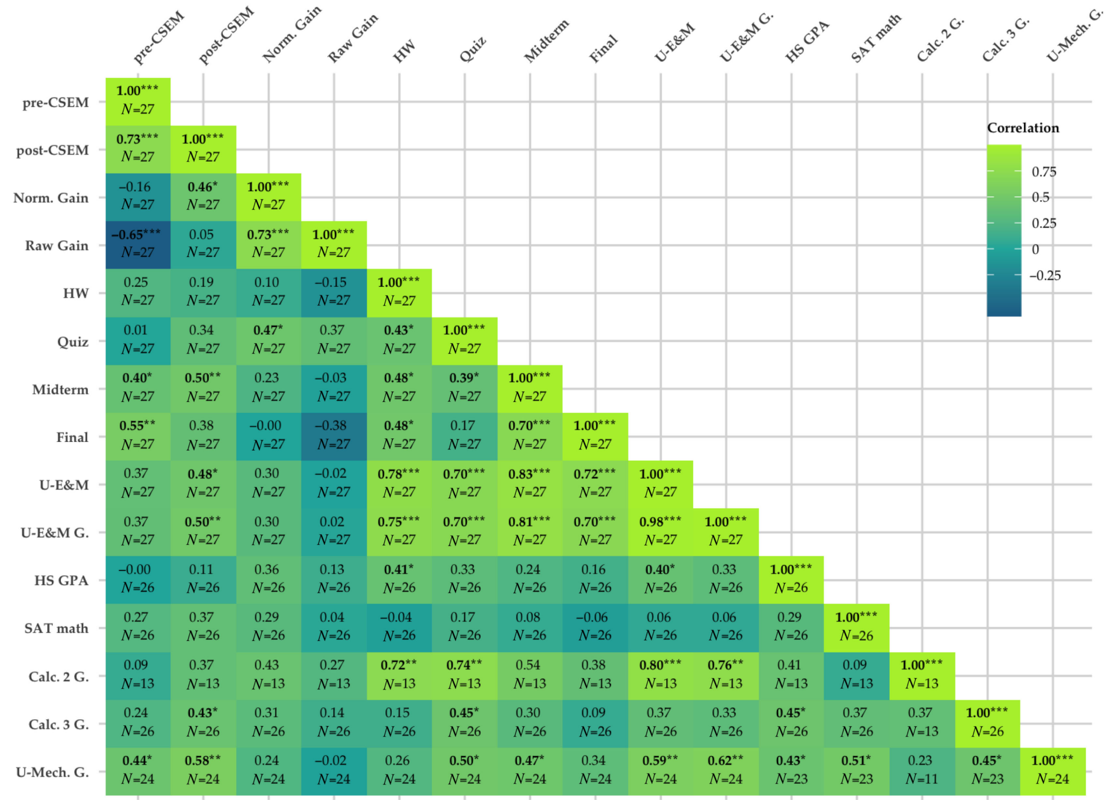

FIG. 6. Correlation matrix of student performance metrics for the top-half group. Significant correlations are shown in bold fonts.

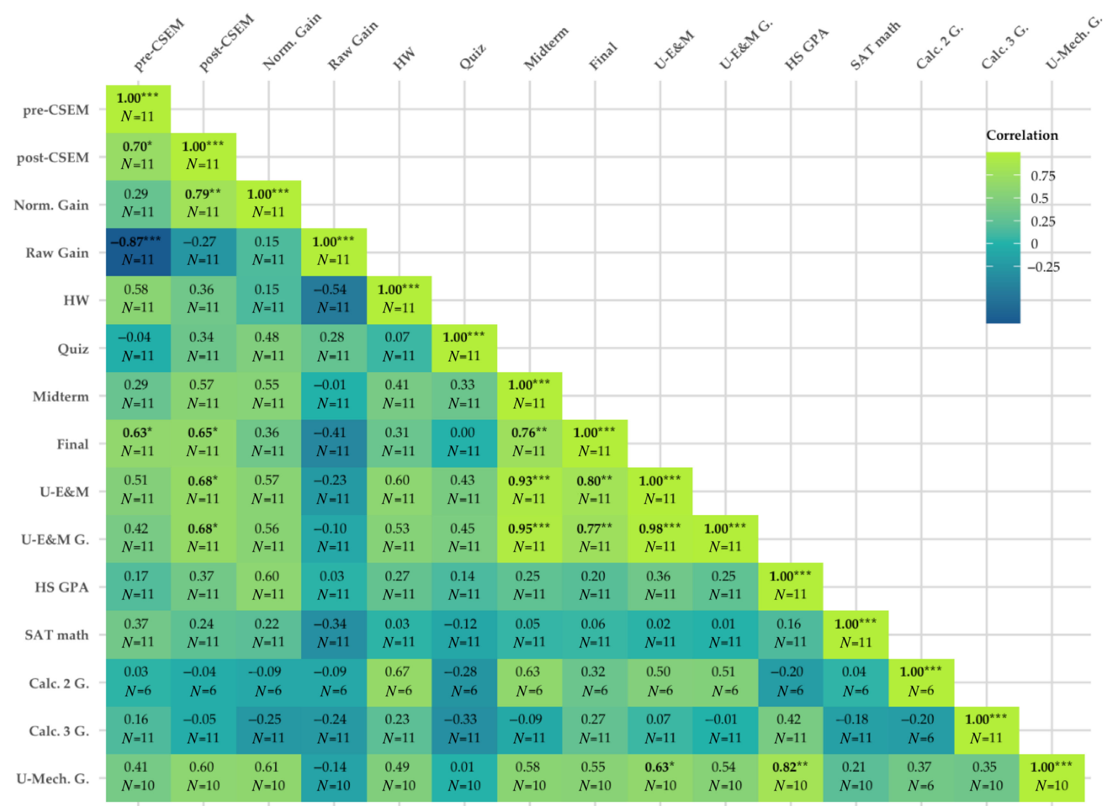

FIG. 7. Correlation matrix of student performance metrics for the top-half group with above average (AA) normalized gain. Significant correlations are shown in bold fonts.

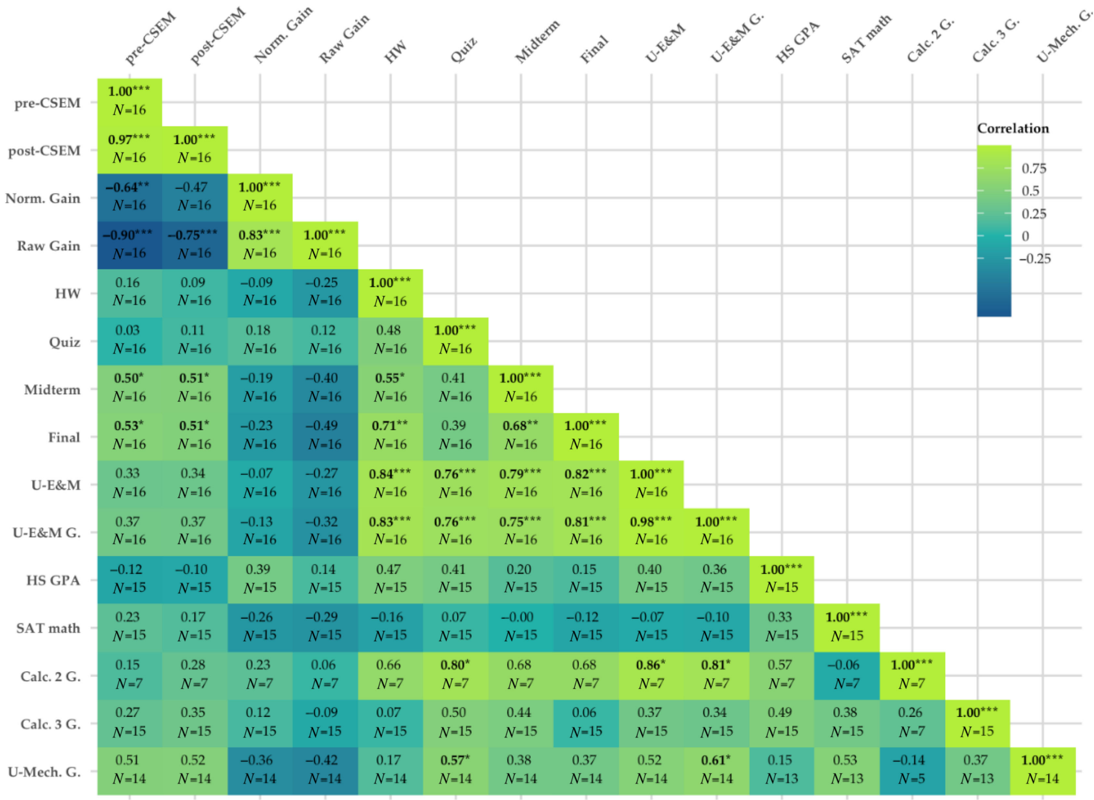

FIG. 8. Correlation matrix of student performance metrics for the top-half group with below average (BA) normalized gain. Significant correlations are shown in bold fonts.

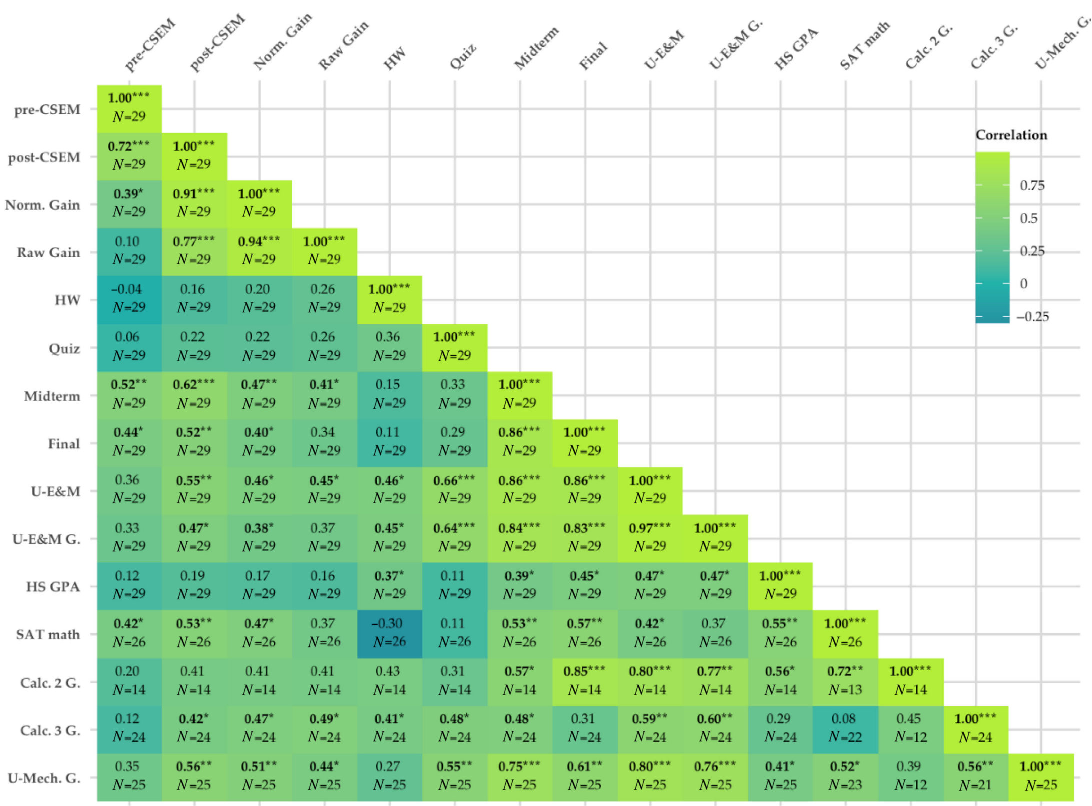

FIG. 9. Correlation matrix of student performance metrics for the bottom-half group. Significant correlations are shown in bold fonts.

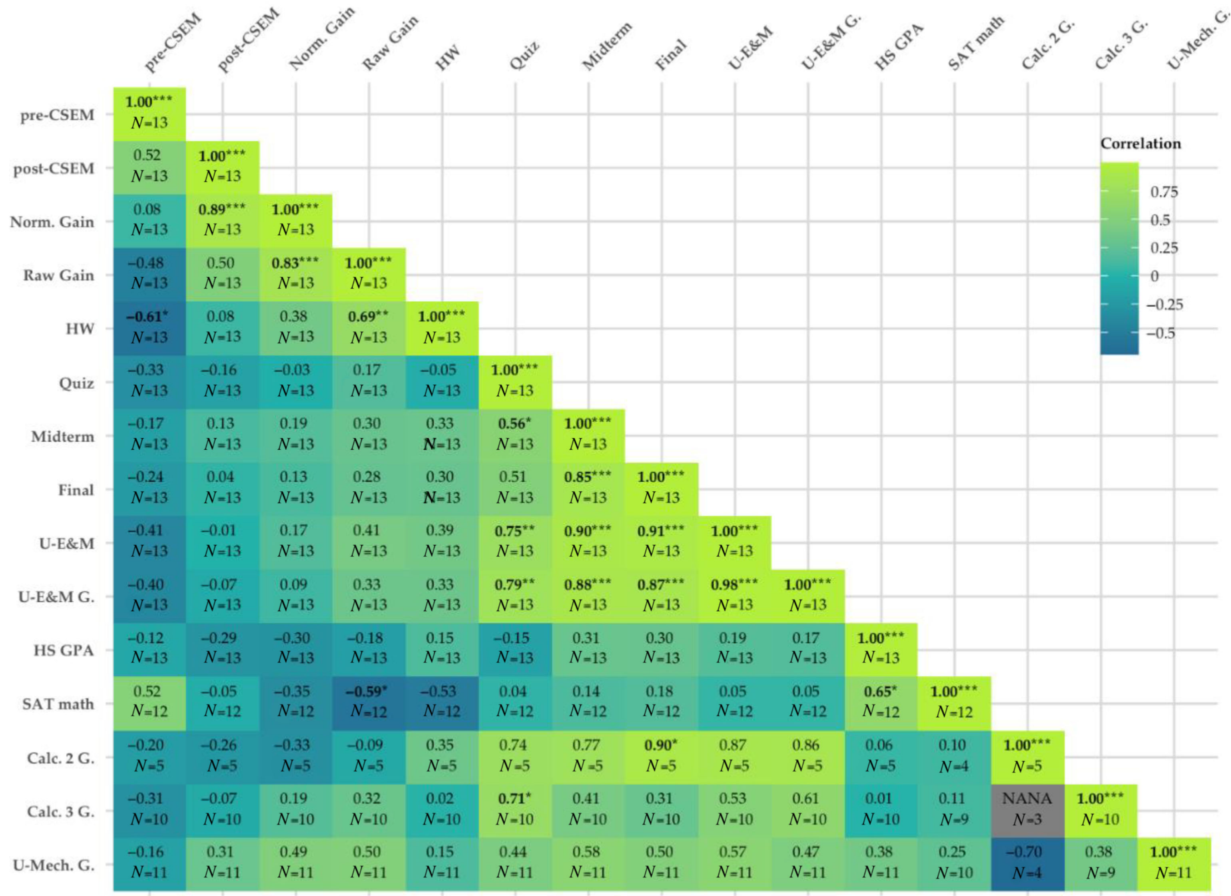

FIG. 10. Correlation matrix of student performance metrics for the bottom-half group with above average (AA) normalized gain. Significant correlations are shown in bold fonts.

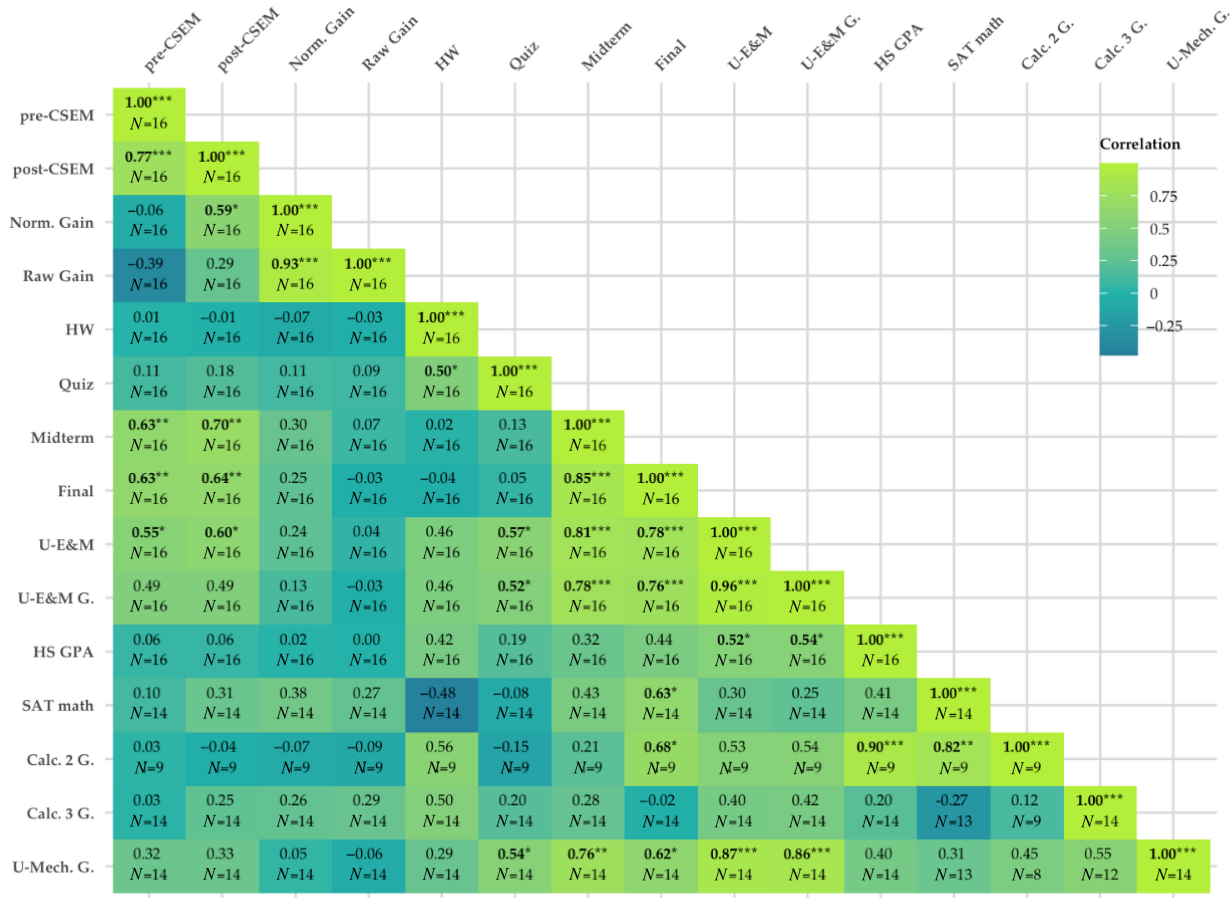

FIG. 11. Correlation matrix of student performance metrics for the bottom-half group with below average (BA) normalized gain. Significant correlations are shown in bold fonts.

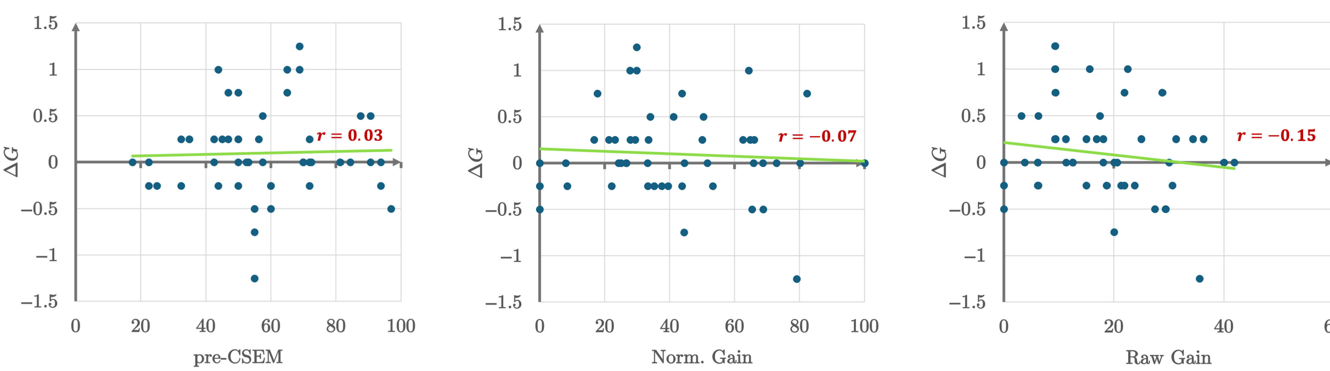

FIG. 12. Scatterplots of change in grade (U-E&M Grade—U-Mech. Grade) versus conceptual learning metrics for all students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

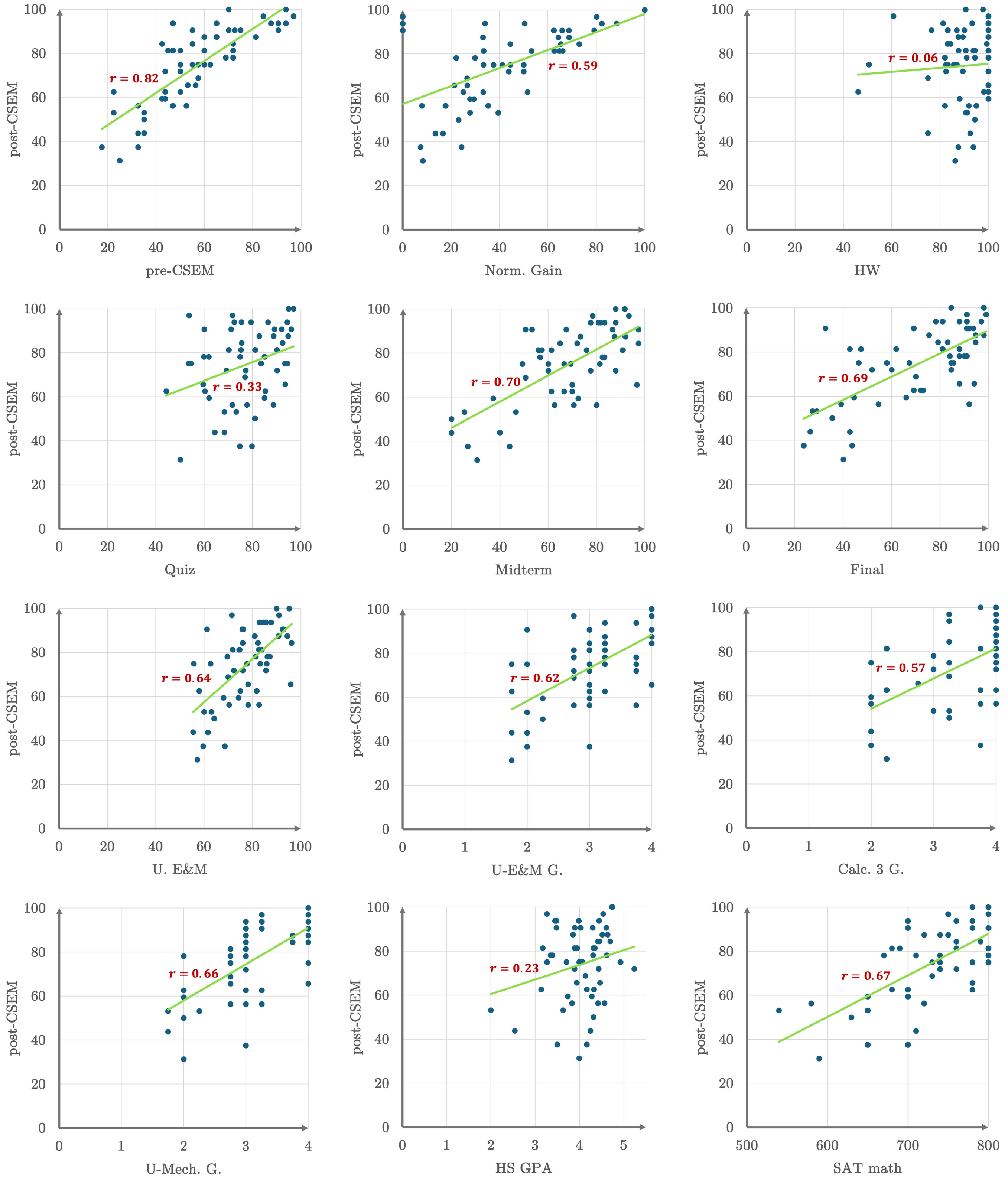

FIG. 13. Scatterplots of post-CSEM scores versus other performance metrics for all students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

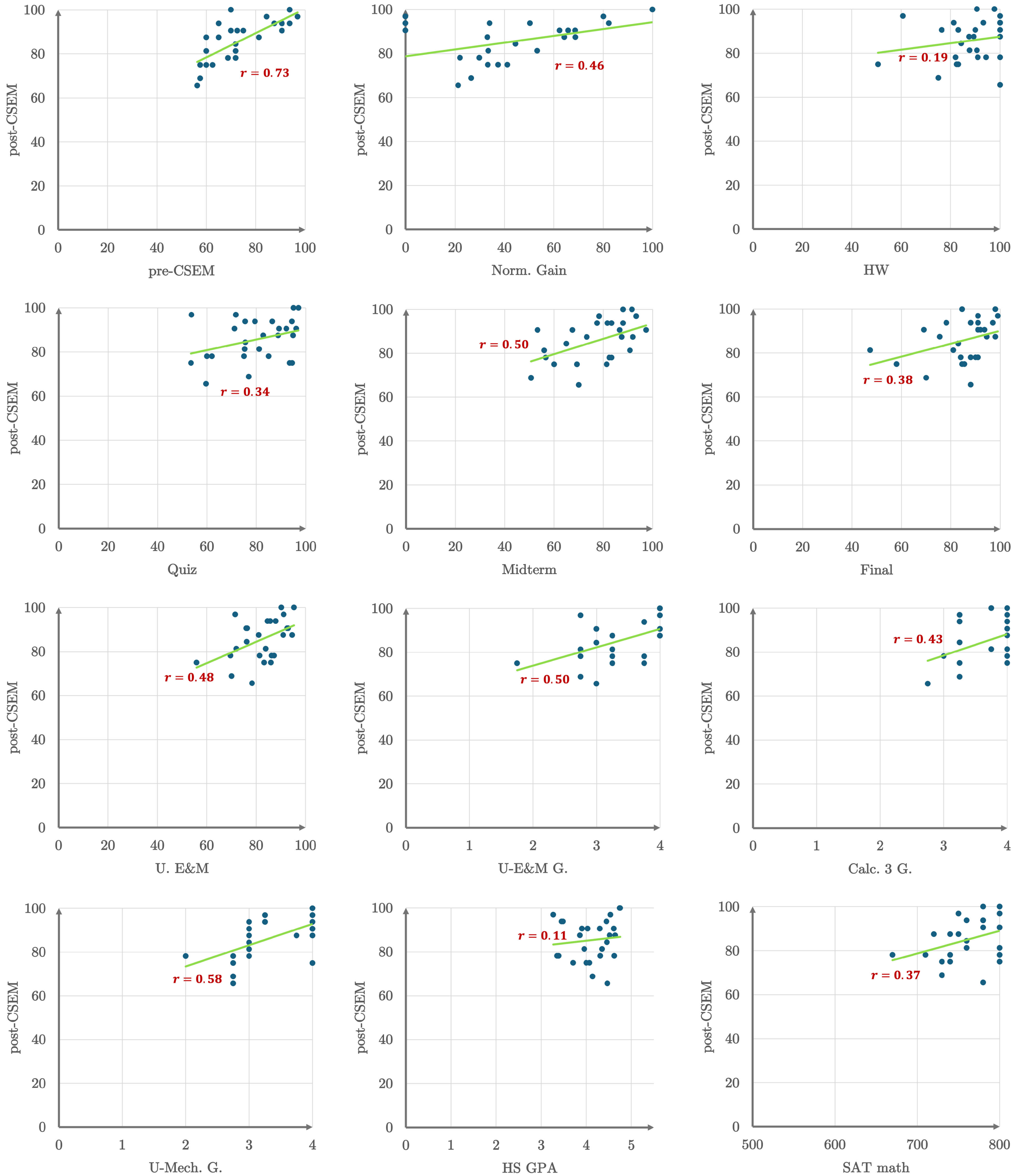

FIG. 14. Scatterplots of post-CSEM scores versus other performance metrics for top-half students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

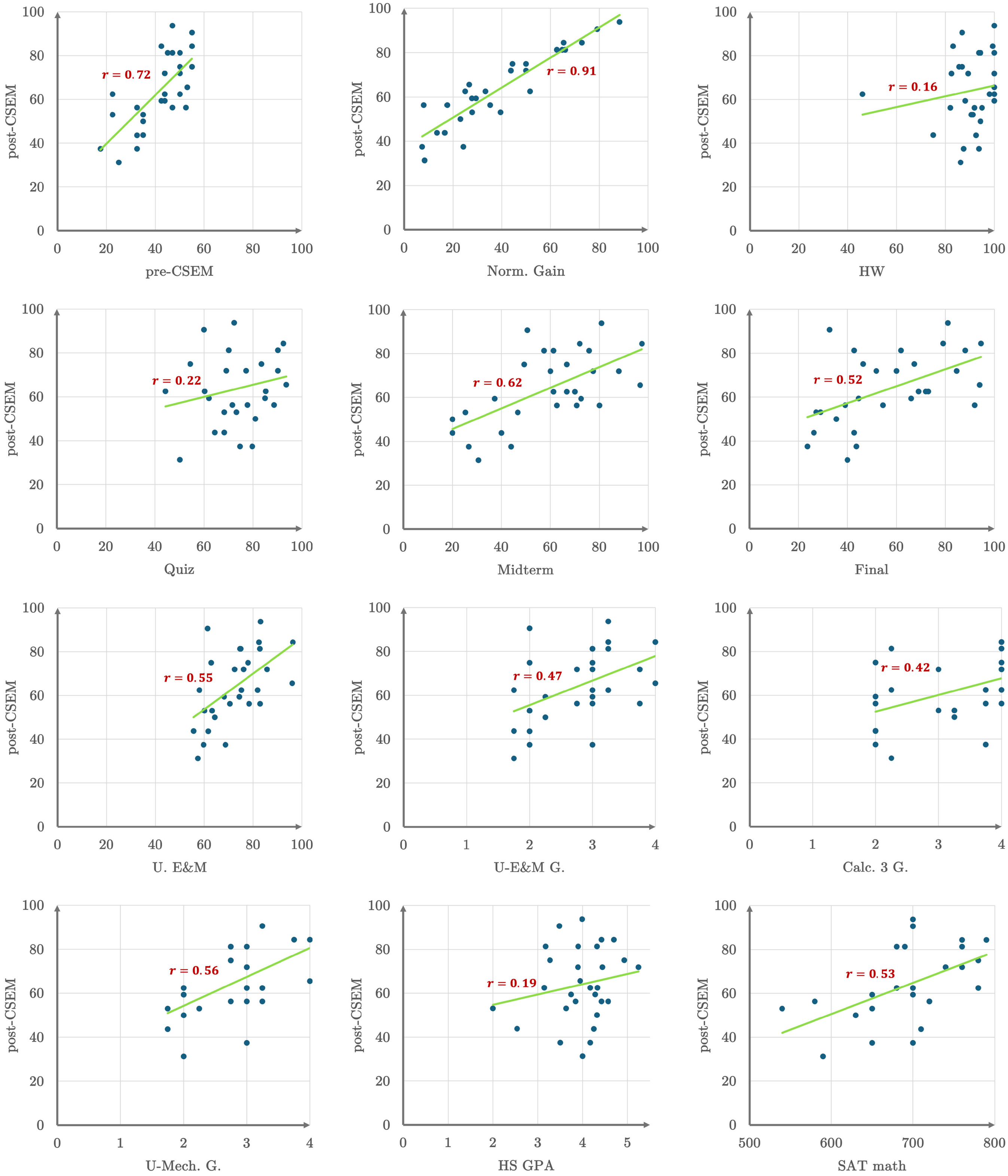

FIG. 15. Scatterplots of post-CSEM scores versus other performance metrics for bottom-half students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

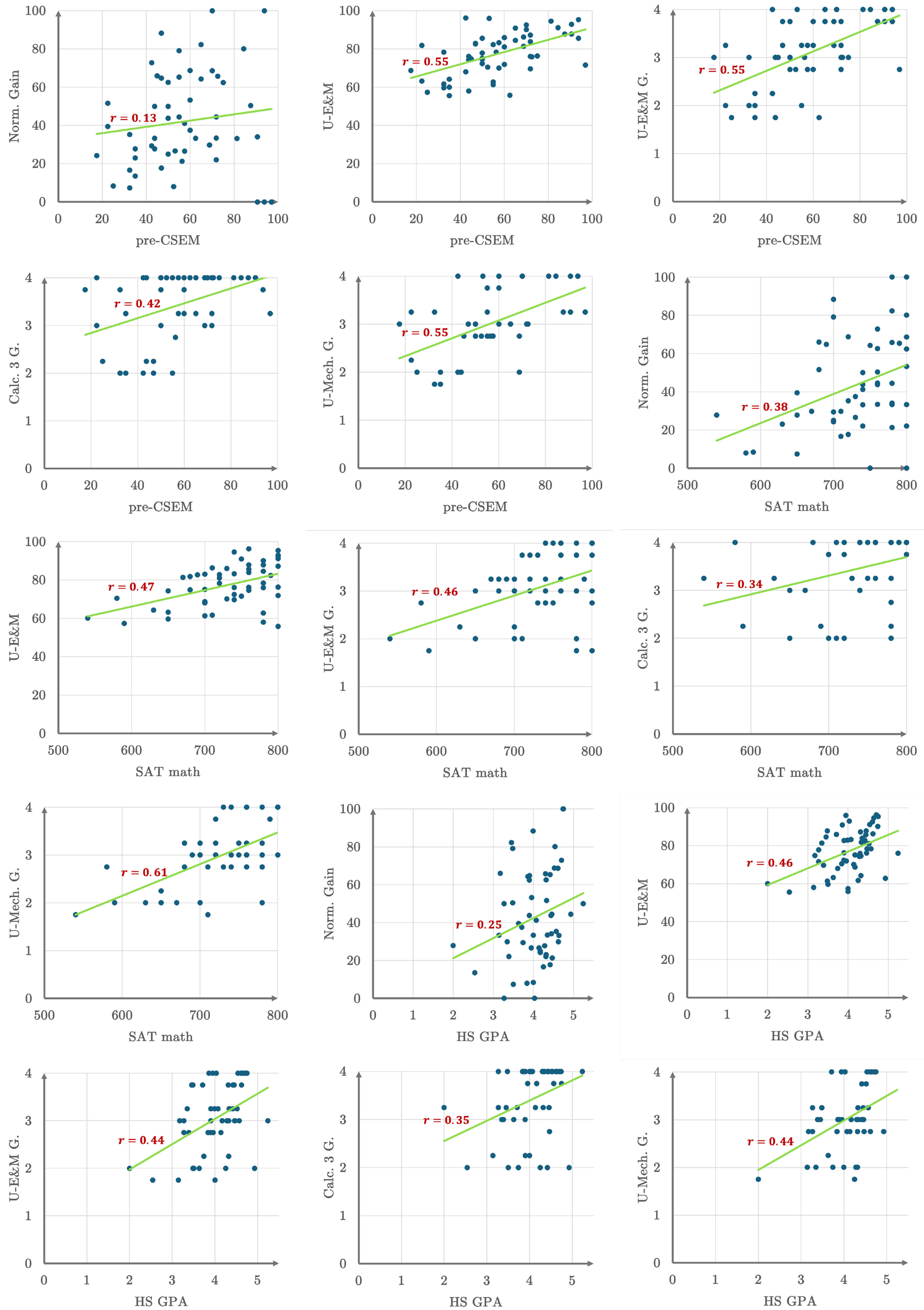

FIG. 16. Scatterplots of normalized gain and course grades versus precourse measures (pre-CSEM, SAT Math, high school GPA) for all students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

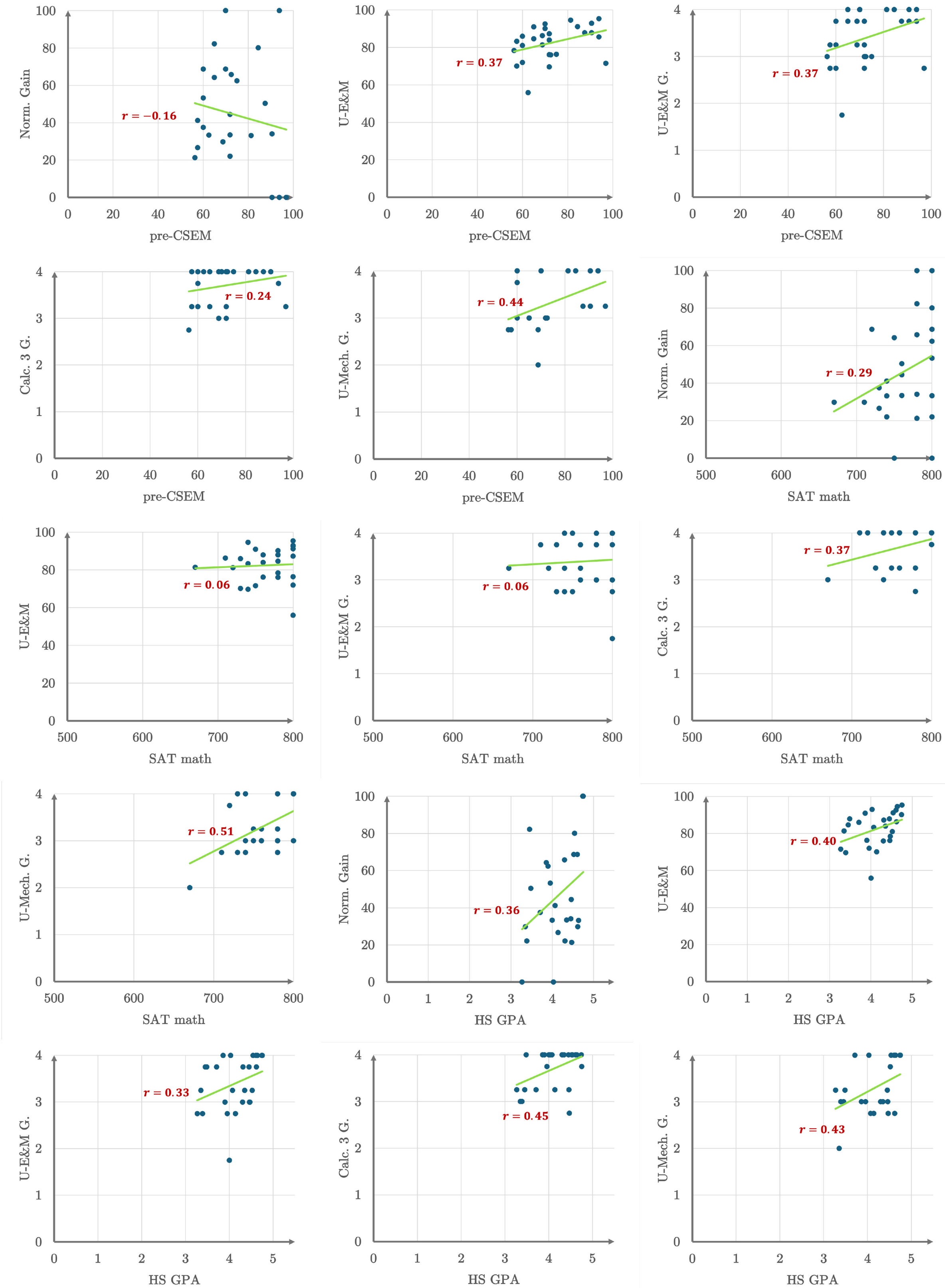

FIG. 17. Scatterplots of normalized gain and course grades versus precourse measures (pre-CSEM, SAT math, high school GPA) for top-half students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

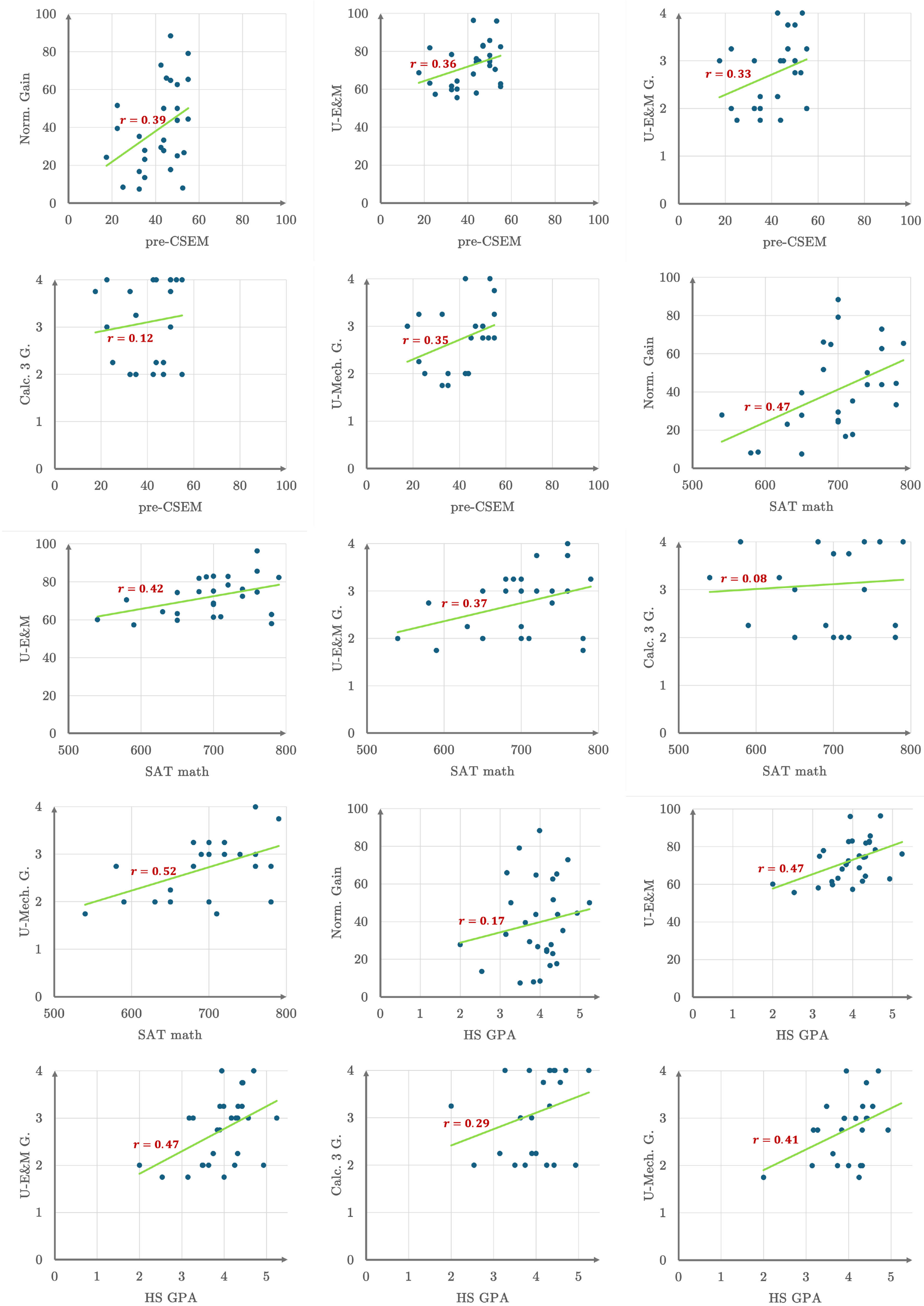

FIG. 18. Scatterplots of normalized gain and course grades versus precourse measures (pre-CSEM, SAT math, and high school GPA) for bottom-half students. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

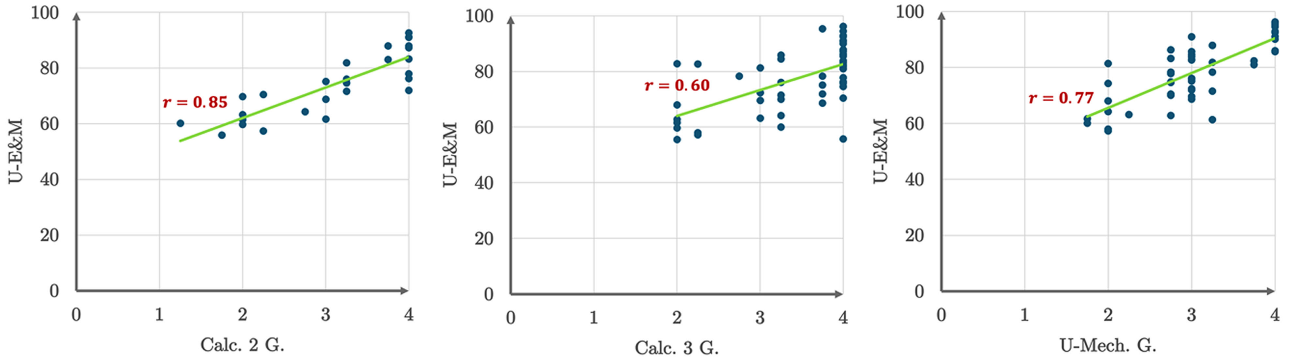

FIG. 19. Scatterplots of U-E&M total percentage versus prior college grades. Linearly fitted trendlines and correlation coefficients are presented in each subfigure.

[1] Á. Suárez, A. C. Marti, K. Zuza, and J. Guisasola, Learning difficulties among students when applying Ampére-Maxwell's law and its implications for teaching, Phys. Rev. Phys. Educ. Res. **20**, 010143 (2024).

[2] E. Campos, G. Zavala, K. Zuza, and J. Guisasola, Students' understanding of the concept of the electric field through conversions of multiple representations, Phys. Rev. Phys. Educ. Res. **16**, 010135 (2020).

[3] E. Campos, G. Zavala, K. Zuza, and J. Guisasola, Electric field lines: The implications of students' interpretation on their understanding of the concept of electric field and of the superposition principle, Am. J. Phys. **87**, 660 (2019).

[4] J. Li and C. Singh, Investigating and improving introductory physics students' understanding of the electric field and superposition principle, Eur. J. Phys. **38**, 055702 (2017).

[5] K. Zuza, M. De Cock, P. van Kampen, L. Bollen, and J. Guisasola, University students' understanding of the electromotive force concept in the context of electromagnetic induction, Eur. J. Phys. **37**, 065709 (2016).

[6] K. Zuza, J.-M. Almudí, A. Leniz, and J. Guisasola, Addressing students' difficulties with Faraday's law: A guided problem solving approach, Phys. Rev. ST Phys. Educ. Res. **10**, 010122 (2014).

[7] R. E. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Observations on student difficulties with mathematics in upper-division electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **8**, 010111 (2012).

[8] D.-H. Nguyen and N. S. Rebello, Students' difficulties with integration in electricity, Phys. Rev. ST Phys. Educ. Res. **7**, 010113 (2011).

[9] E. Campos, E. Hernandez, P. Barniol, and G. Zavala, Phenomenographic analysis and comparison of students' conceptual understanding of electric and magnetic fields and the principle of superposition, Phys. Rev. Phys. Educ. Res. **17**, 020117 (2021).

[10] S. V. Chasteen, R. E. Pepper, M. D. Caballero, S. J. Pollock, and K. K. Perkins, Colorado upper-division electrostatics diagnostic: A conceptual assessment for the junior level, Phys. Rev. ST Phys. Educ. Res. **8**, 020108 (2012).

[11] C. A. Manogue, K. Browne, T. Dray, and B. Edwards, Why is Ampère's law so hard? A look at middle-division physics, Am. J. Phys. **74**, 344 (2006).

[12] L. Bollen, P. van Kampen, and M. De Cock, Students' difficulties with vector calculus in electrodynamics, Phys. Rev. ST Phys. Educ. Res. **11**, 020129 (2015).

[13] B. R. Wilcox and S. J. Pollock, Upper-division student difficulties with separation of variables, Phys. Rev. ST Phys. Educ. Res. **11**, 020131 (2015).

[14] B. R. Wilcox *et al.*, Upper-division student understanding of Coulomb's law: Difficulties with continuous charge distributions, AIP Conf. Proc. **1513**, 418 (2013).

[15] L. Bollen, P. van Kampen, C. Baily, M. Kelly, and M. De Cock, Student difficulties regarding symbolic and graphical representations of vector fields, Phys. Rev. Phys. Educ. Res. **13**, 020109 (2017).

[16] S. V. Chasteen, S. J. Pollock, R. E. Pepper, and K.K. Perkins, Thinking like a physicist: A multi-semester case study of junior-level electricity and magnetism, Am. J. Phys. **80**, 923 (2012).

[17] P. M. Sadler and R. H. Tai, Success in introductory college physics: The role of high school preparation, Sci. Educ. **85**, 111 (2001).

[18] E. W. Burkholder, G. Murillo-Gonzalez, and C. Wieman, Importance of math prerequisites for performance in introductory physics, Phys. Rev. Phys. Educ. Res. **17**, 010108 (2021).

[19] S. Salehi, E. Burkholder, G. P. Lepage, S. Pollock, and C. Wieman, Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics, Phys. Rev. Phys. Educ. Res. **15**, 020114 (2019).

[20] C. A. Manogue, P. J. Siemens, J. Tate, K. Browne, M. L. Niess, and A. J. Wolfer, Paradigms in physics: A new upper-division curriculum, Am. J. Phys. **69**, 978 (2001).

[21] S. V. Chasteen, S. J. Pollock, R. E. Pepper, and K. K. Perkins, Transforming the junior level: Outcomes from instruction and research in E&M, Phys. Rev. ST Phys. Educ. Res. **8**, 020107 (2012).

[22] J. Li and C. Singh, Investigating and improving introductory physics students' understanding of symmetry and Gauss's law, Eur. J. Phys. **39**, 015702 (2018).

[23] C. Singh, Student understanding of symmetry and Gauss's law of electricity, Am. J. Phys. **74**, 923 (2006).

[24] A. Maries, M. J. Brundage, and C. Singh, Using the Conceptual Survey of Electricity and Magnetism to investigate progression in student understanding from introductory to advanced levels, Phys. Rev. Phys. Educ. Res. **18**, 020114 (2022).

[25] S. J. Pollock, Longitudinal study of student conceptual understanding in electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **5**, 020110 (2009).

[26] C. Tzanakis, Mathematics and physics: An innermost relationship. Didactical implications for their teaching and learning, in *Proceedings of the 2016 ICME Satellite Meeting – HPM 2016*, edited by L. Radford, F. Furinghetti, and T. Hausberger (2016), pp. 79–104.

[27] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, Sci. Educ. **21**, 485 (2012).

[28] D. J. Kevles, *The Physicists: The History of a Scientific Community in Modern America* (Harvard University Press, Cambridge, MA, 1995).

[29] T. H. Kjeldsen and J. Lützen, Interactions between mathematics and physics: The history of the concept of function—teaching with and about nature of mathematics, Sci. Educ. **24**, 543 (2015).

[30] G. Pospiech *et al.*, *Mathematics in Physics Education* (Springer, New York, 2019).

[31] E. F. Redish, Problem solving and the use of math in physics courses, in *Proceedings of the World View on Physics Education in 2005: Focusing on Change, Delhi* (2005), arXiv:physics/0608268.

[32] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, Sci. Educ. **24**, 561 (2015).

[33] M. Pietrocola, Mathematics as structural language of physical thought, in *Connecting Research in Physics Education with Teacher Education*, edited by M. Vicentini and E. Sassi (International Commission on Physics Education, 2008), pp. 13–18, https://www.per-central.org/items/detail.cfm?ID=9881.

[34] D. Hestenes, Toward a modeling theory of physics instruction, Am. J. Phys. **55**, 440 (1987).

[35] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, Am. J. Phys. **67**, 755 (1999).

[36] National Research Council, *Adapting to a Changing World: Challenges and Opportunities in Undergraduate Physics Education* (The National Academies Press, Washington, DC, 2013), pp. 60–65, 10.17226/18312.

[37] P. C. Peters, Even honors students have conceptual difficulties with physics, Am. J. Phys. **50**, 501 (1982).

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

[39] J. Li and C. Singh, Developing and validating a conceptual survey to assess introductory physics students' understanding of magnetism, Eur. J. Phys. **38**, 025702 (2017).

[40] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[41] P. V. Engelhardt and R. J. Beichner, Students' understanding of direct current resistive electrical circuits, Am. J. Phys. **72**, 98 (2004).

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

[43] M. Planinic, Assessment of difficulties of some conceptual areas from electricity and magnetism using the Conceptual Survey of Electricity and Magnetism, Am. J. Phys. **74**, 1143 (2006).

[44] N. Karim, A. Maries, and C. Singh, Exploring one aspect of pedagogical content knowledge of teaching assistants using the Conceptual Survey of Electricity and Magnetism, Phys. Rev. Phys. Educ. Res. **14**, 010117 (2018).

[45] D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores, Am. J. Phys. **70**, 1259 (2002).

[46] A. J. Mason, J. M. McCardell, P. A. White, and J. S. Colton, Improving performance in upper-division electricity and magnetism with explicit incentives to correct mistakes, Phys. Rev. Phys. Educ. Res. **19**, 020104 (2023).

[47] J. P. Zwolak and C. A. Manogue, Assessing student reasoning in upper-division electricity and magnetism at Oregon State University, Phys. Rev. ST Phys. Educ. Res. **11**, 020125 (2015).

[48] R. E. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Observations on student difficulties with mathematics in upper-division electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **8**, 010111 (2012).

[49] S. J. Pollock and S. V. Chasteen, Longer term impacts of transformed courses on student conceptual understanding of E&M, AIP Conf. Proc. **1179**, 237 (2009).

[50] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Cambridge University Press, Cambridge, England, 2017).

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

[52] ACT Inc. and College Board, *Guide to the 2018 ACT®/SAT® Concordance* (ACT, Inc., New York, NY and College Board, Iowa City, IA, 2018).

[53] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1988).

[54] F. Wilcoxon, Individual comparisons by ranking methods, in *Breakthroughs in Statistics: Methodology and Distribution* (Springer, New York, 1992), pp. 196–202.

[55] H. Wickham, *GGPLOT2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York 2016).

[56] College Board: Understanding Scores, https://research.collegeboard.org/reports/sat-suite/understanding-scores.

[57] D. E. Meltzer and D. H. King, Pre-instruction diagnostic tests can predict grade probabilities in introductory physics, Am. J. Phys. **93**, 528 (2025).

[58] J. Stewart, G. L. Cochran, R. Henderson, C. Zabriskie, S. DeVore, P. Miller, G. Stewart, and L. Michaluk, Mediational effect of prior preparation on performance differences of students underrepresented in physics, Phys. Rev. Phys. Educ. Res. **17**, 010107 (2021).

[59] J. K. Hirons *et al.*, Click, watch, learn: The impact of student self-study materials on physics E&M course outcomes, arXiv:2508.12143.

[60] K. M. Whitcomb and C. Singh, For physics majors, gender differences in introductory physics do not inform future physics performance, Eur. J. Phys. **41**, 065701 (2020).

[61] S. F. Itza-Ortiz, S. Rebello, and D. Zollman, Students' models of Newton's second law in mechanics and electromagnetism, Eur. J. Phys. **25**, 81 (2004).

[62] J. Sweller, Cognitive load during problem solving: Effects on learning, Cognit. Sci. **12**, 257 (1988).

[63] P. D. Justice *et al.*, A framework for understanding the impact of integrating conceptual and quantitative reasoning in a quantum optics tutorial on students' conceptual understanding, Educ. Sci. **15**, 1314 (2025).

[64] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend conceptual and formal mathematical reasoning in solving physics problems, Sci. Educ. **97**, 32 (2013).

[65] K. Shabani *et al.*, Vygotsky's zone of proximal development: Instructional implications and teachers' professional development, English Lang. Teach. **3**, 237 (2010).

[66] L. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).

[67] E. W. Burkholder and C. E. Wieman, What do AP physics courses teach and the AP physics exam measure?, Phys. Rev. Phys. Educ. Res. **15**, 020117 (2019).

[68] E. Burkholder and S. Salehi, Exploring the pre-instruction gender gap in physics, PLoS One **17**, e0271184 (2022).

[69] S. J. Pollock, Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA, AIP Conf. Proc. **1064**, 171 (2008).

[70] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).

[71] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, Exploring the gender gap in the Conceptual Survey of Electricity and Magnetism, Phys. Rev. Phys. Educ. Res. **13**, 020114 (2017).

[72] A. Ghimire and C. Singh, Perspectives of women and men students and faculty on conceptual and quantitative problem-solving in physics from introductory to graduate levels, Educ. Sci. **15**, 1602 (2025).