2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #50654

Scaffolding for success: How interteaching pedagogy may level the playing field in a combined engineering statics and mechanics course

Dr. Steven Robert Woodruff, James Madison University

Dr. Steven Woodruff is an Assistant Professor of Engineering at James Madison University, where he teaches courses in statics, mechanics, and structural analysis. His research explores how pedagogical strategies can promote student success in foundational engineering courses that often serve as gateways to persistence in the discipline. Drawing from behavioral and active-learning frameworks, his recent work examines the application of interteaching pedagogy to engineering mechanics education, with an emphasis on reducing performance disparities and fostering deeper conceptual understanding.

Benjamin T Blankenship, James Madison University

Ben Blankenship, Ph.D., is an Associate Professor of Psychology at James Madison University. One of his lines of research centers on inclusive excellence in STEM fields, including engineering. In addition to publishing on and editing special issues about how to foster diversity and belonging within STEM education, he also brings expertise in personality and social psychology, where his work explores how social factors and personality shape engagement and a sense of belonging in key societal areas, including in educational contexts.

Scaffolding for success: How interteaching pedagogy levels the playing field in a combined engineering statics and mechanics course

Introduction

Engineering statics and solid mechanics sit at a critical juncture in many engineering curricula. These core mechanics courses commonly function as "gatekeepers," because students must earn a minimum grade to progress to engineering science and design coursework. In many programs, these courses also carry structural features that can intensify risk of failure, including large enrollments and limited individualized feedback. Outcomes in foundational classes, such as statics and introductory solid mechanics, can shape whether students persist in engineering pathways.

These course-level challenges occur within a broader national pattern, where many students who initially intend to pursue STEM do not complete a STEM degree. Using longitudinal transcript data, Chen and Soldner's National Center of Education Statics report [1] documents substantial STEM attrition over the college years and characterizes early undergraduate STEM coursework as a key point where students depart from STEM pathways. In engineering specifically, gateway mechanics courses can, therefore, represent a high-impact target for interventions aimed at improving persistence.

Persistence and performance in gateway STEM courses are also not evenly distributed. National reports have repeatedly noted that students from groups historically underrepresented in STEM face additional barriers and are underrepresented in science and engineering fields, motivating instructional and structural reforms that improve early-course success [2, 3]. Qualitative work on why capable students leave science and engineering majors further highlights the central role of early course experiences, including unwelcoming climates and discouraging introductory instruction, in decisions to switch out of STEM majors [4, 5]. Together, these lines of evidence demonstrate the necessity of not only raising average performance but also reducing early performance gaps that can cascade into withdrawal, repeating courses, and major switching.

Despite these stakes, lecture-centered instruction remains common across undergraduate STEM. Large-scale observational evidence indicates that lecturing is still prominent throughout the undergraduate STEM curriculum, even in classrooms designed to support active learning [6]. This matters because a substantial body of research shows that active learning (i.e., approaches that increase student engagement and practice with feedback) improves achievement and reduces failure rates, compared to traditional lecturing across STEM disciplines [7]. Engineering educators have responded by adopting a wide range of active-learning models (e.g., flipped structures) in mechanics contexts, often emphasizing the need to improve engagement and outcomes in high-enrollment, high-stakes courses [8-15].

Interteaching offers a promising, highly structured alternative to traditional lecture-based instruction that may be especially well-suited to gateway mechanics courses. Rather than centering class time on instructor-delivered exposition, interteaching is designed to increase students' active engagement with the course content and to make instructor time focused on being responsive to what students identify as the most difficult topics. Saville and colleagues reported that unit test scores were typically higher following interteaching-based instruction, compared to following lecture-based instruction, with students scoring meaningfully higher for the same material taught using interteaching [16]. Related summaries of the empirical literature similarly demonstrate interteaching producing higher quiz scores than lecture, as well as being preferred by many students [17-19]. Most recently, a systematic review and quantitative synthesis concluded that interteaching is more effective than traditional lecture (or other control conditions) overall, with a large average effect, while also noting substantial variation in implementation across studies [20].

While interteaching shows promise as an effective alternative to lecture, studies have mostly been limited to disciplines outside of engineering, most notably in psychology. Few studies on the effectiveness of interteaching in engineering have been reported [21], and the studies that do exist often use interteaching in a limited way (e.g., only during recitation sessions [22]). The authors' aims for this study were to extend the findings from other fields to engineering education, specifically in the context of engineering mechanics courses, comparing its effectiveness to traditional lecture-based instruction taught by the same instructor.

For engineering statics and solid mechanics, the key features of interteaching are particularly relevant because misunderstandings that happen early in a course can compound quickly and widen performance gaps by the end of the term. A pedagogy that systematically increases engagement with core representations and problem-solving processes, as well as targets instruction to documented student difficulties, is likely to improve outcomes relative to lecture-based instruction. Furthermore, interteaching may be especially beneficial for students who begin the course at a disadvantage. In this paper, equity refers to reducing academic performance disparities between initially lower- and higher-performing students, not demographic equity gaps. While demographic equity gaps are widely documented in gateway STEM courses, the present study focuses on equity in terms of academic performance trajectories.

This present study evaluates interteaching in a 16-week, combined engineering statics and introductory solid mechanics course. Three sections (n = 62) were randomly assigned to either an interteaching format or a traditional lecture format. We examined (a) changes in exam performance over the semester, including whether effects are concentrated among students who perform below the median early in the course, and (b) student-reported engagement indicators and perceptions related to the learning environment.

The authors generated the following research questions:

> **RQ1:** Does interteaching increase students' levels of engagement in-class, as well as outside of class?

RQ2: Does the interteaching method improve student performance on learning objectives? How do initially high and low performing students see differential benefits for their learning, if any?

Based on the existing literature on interteaching, as well as the intuitions of the authors, theory, and prior knowledge, we conducted the current study to test the following hypotheses:

H1: Participants assigned to the interteaching condition will self-report greater engagement with the course both (a) in-class and (b) outside of class, compared to those assigned to the control (lecture) condition.

H2: Participants assigned to the interteaching condition will show (a) better performance on the cumulative final examination, as well as (b) demonstrate greater improvement in performance over time, compared to those assigned to the control condition.

H3: Initially low-performing participants (those who scored below the median on the first exam) will show significantly greater improvement in their exam scores over the course of the semester, when assigned to the interteaching condition, compared to those assigned to the control condition.

We begin the article by introducing the readers to interteaching pedagogy and give details on how the courses were taught both in the interteaching and lecture formats. Next, we explain the methodological details of the study and the statistical analyses conducted. We finally show the results of the study and discuss the implications, limitations, and future directions of the research.

Interteaching background

Interteaching is a structured instructional approach originally developed by Boyce and Hineline [23] as an alternative to traditional lecture that systematically increases student engagement while maintaining a manageable instructional workload. The method was initially grounded in principles from classical conditioning and learning theory, emphasizing repeated exposure to content through multiple modalities. In its original formulation, students encounter course material through pre-class preparation, instructor-led clarification, and in-class practice, providing several opportunities for engagement with the same concepts. Although the theoretical justification for interteaching emerged from psychology, its core design principles align closely with contemporary evidence-based instructional practices that emphasize active engagement, practice with feedback, and responsiveness to student understanding.

Interteaching was explicitly developed as a pragmatic compromise between lecture-centered instruction and more resource-intensive active-learning models, such as just-in-time teaching and fully flipped classrooms. In a fully flipped structure, students are typically required to engage with instructor-produced content prior to class, with class time devoted almost entirely to activities. While effective when implemented well, flipped models can place substantial demands on instructors to develop and maintain high-quality instructional materials and can be ineffective when students do not complete pre-class work. In contrast, interteaching retains a focused lecture component, while embedding active, structured peer interaction and practice within class time as well. This hybrid structure reduces reliance on perfect student compliance and allows instructors to leverage existing lecture notes, problem sets, and assessments with relatively modest modifications.

A defining feature of interteaching is its use of “muddiest points” to guide instruction. Rather than delivering a comprehensive lecture covering all aspects of a topic, instructors review students’ self-reported areas of confusion prior to class and design a brief clarifying lecture that targets these difficulties. This approach reallocates instructional time towards concepts students actually find challenging and away from material they already understand. In this sense, interteaching functions as a closed-loop feedback system: student preparation informs instruction, instruction supports practice, and practice generates new information about student understanding. For instructors accustomed to lecturing, this structure allows lecture to remain central while making it more targeted and responsive.

A typical interteaching cycle includes four recurring components, summarized in **Figure 1**. First, students complete assigned readings or instructional videos before class, representing their initial exposure to the material. Second, the instructor reviews students’ previously submitted muddiest points and delivers a short lecture (less than half of the class period) focused on clarifying those topics. Third, students work in pairs during class on a structured preparation guide consisting of questions or problems aligned with the assigned material. Finally, students submit new muddiest points reflecting on unresolved difficulties, which inform subsequent instruction. While implementations vary, this cycle is designed to repeat consistently across class meetings, creating predictable expectations for both students and instructors. Importantly, interteaching does not require specialized software or instructional technologies, although learning management systems can streamline logistics.

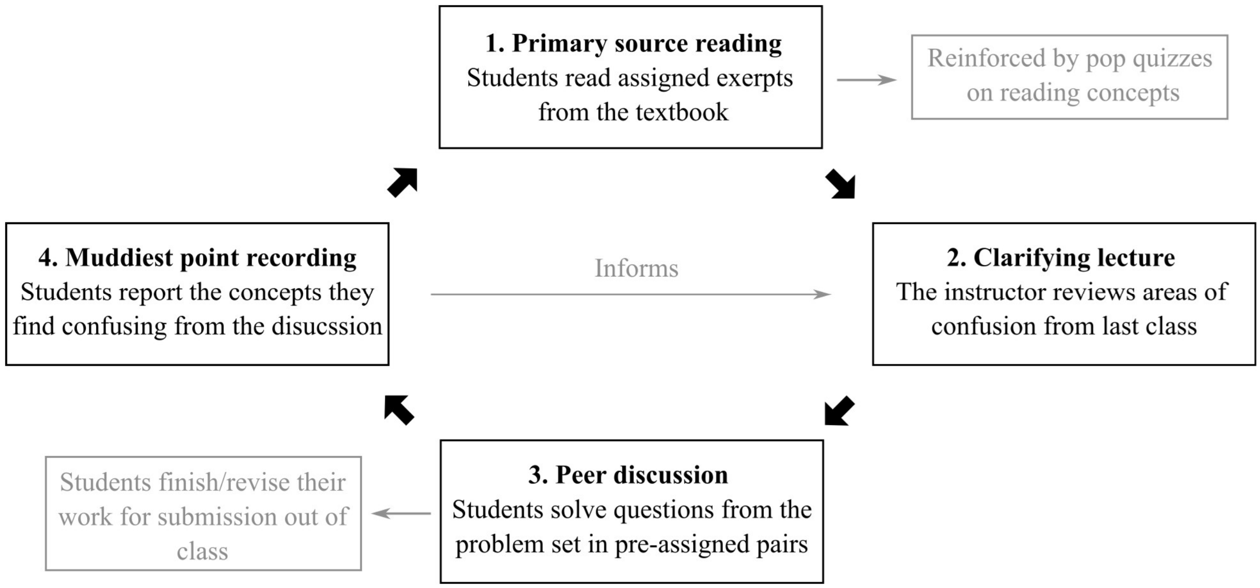


Figure 1 - The daily interteaching cycle (class period workflow).

Practical approaches to applying interteaching can be found in the interteaching literature cited earlier. Furthermore, Saville has written ten tips for effective implementation of the pedagogy which the first author used to help develop the course [24]. Examples of best practices for teaching using an interteaching pedagogy include: (1) explaining to students how interteaching works and why it was selected for the course; (2) explaining to students the best practices for completing their work in class; and (3) explicitly connecting course discussions with exam problems.

Engineering mechanics courses may be particularly well-suited for interteaching because misconceptions and procedural gaps can compound across sequential topics, making early setbacks difficult to recover from without structured feedback and targeted practice. While interteaching has demonstrated learning and attitudinal benefits across multiple disciplines, its use in calculation-intensive engineering courses remains underexplored. This study, therefore, evaluates the implementation of interteaching in a combined statics and introductory solid mechanics course and examines its effects on student perceptions, engagement, and performance trajectories across the semester.

Method

Participants

Participants included 79 undergraduate students enrolled in three sections of a Statics and Introductory Mechanics (ENGR 212) course at a large, public, R-2 institution in the mid-Atlantic region of the United States. This introductory class is restricted to engineering majors and has three prerequisites: An introductory engineering course, second-level differential and integral calculus, and introductory Newtonian physics.

Of the original 79, 17 were excluded from analyses for not completing key survey measures or for inputting the incorrect data key, meaning that their responses to the survey could not be paired with their grades in the course. This produced a final sample size of 62. Demographics were not collected in the survey, in order to prevent possible accidental disclosure of identifiable information through inference (i.e., if there is only one person of color in a class, it would be easy looking at the data to infer how they answered the survey), especially given the relatively small class sizes.

Procedure

Using a random method (coin flip), three sections of Statics and Introductory Mechanics were randomly assigned to one of two conditions: traditional lecture-based instruction or interteaching-based instructions. One of the sections ($n = 26$) was assigned to the traditional lecture instruction, with the final sections ($n = 53$) assigned to the interteaching condition[1].

All classes were taught by the first author. Participants received instruction throughout the semester in accordance with their assigned condition. The primary method of evaluating learning, for the purpose of this study, consisted of five tests administered at equal intervals throughout the semester.

In the interteaching condition, classes closely followed the four-step cycle given in **Figure 1**. Students were assigned sections of the textbook to read before class (the assignment was posted on the class's learning management system, listed in the course schedule, and announced at the beginning and end of each class meeting). Typically, these readings covered a single concept, such as representing forces as Cartesian vectors or calculating the area moment of inertia for a composite cross section. When students arrived in class, the instructor would review the learning objectives from the reading, explain the muddiest points from the last class, and give a roughly 25-minute clarifying lecture. Roughly once per week, students would take a three-question, multiple-choice, open-note quiz to incentivize reading at the start of the class. After the clarifying lecture, students were assigned to rotating pairs using an online round-robin tournament generator to ensure that students interacted with each of their peers at least once. In these pairs, students worked on selected problem set questions relevant to the reading they had just completed. Students were allowed to use the textbook or online resources, along with each other, to aid them in solving the problems. The instructor walked around the classroom during this problem-solving process to answer questions and prod students who seem unengaged. Near the end of the class session, the pairs were disbanded, the students were reminded to individually complete their muddiest point submissions on the learning management system before the end of

[1] To handle any potential issues resulting from unequal sample sizes, as well as account for any preexisting differences between the groups in key variables, analyses were conducted using propensity score matching (e.g. [25]) to weight the participants' scores from the interteaching condition so that they were equivalent to the control (lecture) condition, in terms of key variables, specifically college GPA, course attendance, and grade in University Physics I).

the day. The instructor reviewed these submissions prior to the next class to synthesize the main areas of confusion and to design the clarifying lecture.

In the lecture condition, students were assigned the same readings as the interteaching condition students, but there were no mechanisms for assessing compliance. At the start of the class, the instructor outlined the learning objectives for the lecture. After which, the instructor gave a 50-minute presentation on topics from the reading, including solving problems on the board. The only in-class formative checks used were informal comprehension prompts where the instructor periodically asked students if they had any questions.

In both conditions, students were given the same lecture notes (in the learning management software), problem set assignments, and lab assignments. Students in either condition were also assigned the same five exams (four non-cumulative mid-term exams and one cumulative final exam). Furthermore, attendance was recorded in all sections of the course for a small portion of the students' final grades (5%).

On the last day of classes for the semester, participants completed an in-class survey to measure their perspectives on the course, their self-reported engagement, and information about their performance in previous courses. At the conclusion of the survey, participants were randomly assigned a unique "participant ID," which they were then instructed to enter into an assignment submission in Canvas. These participant IDs were then used to pair their course performance data from Canvas (handled by the first author) with their anonymous survey data. Once the data was successfully combined in R, the participant IDs were removed, making the final dataset unidentifiable.

Participants signed informed consent forms upon entering the course. Participants were fully debriefed in the final survey at the end of the semester. The research protocol was deemed exempt from full oversight by the author's Institutional Review Board (IRB).

Manipulations and measures

Participants in this study were assigned to conditions, based on their course enrollment and random assignment to conditions. Performance was measured at multiple timepoints throughout the semester through five examinations. Key measures of interest were taken in the final survey at the end of the semester. Variables were measured/manipulated in the order they appear below.

Self-reported engagement

Participants' levels of engagement were measured using multiple-choice questions in the final course survey. Out-of-class (i.e., reading) engagement was measured using a single item, "Since the beginning of the semester, how often did you read the textbook before coming to class?," using a 5-point scale from "always" to "never." Self-reported in-class engagement was measured using a single item "Do you feel the class format encouraged your participation?," using a 5-point scale from "strongly agree" to "strongly disagree."

Examination performance

Learning performance was measured through five examinations that were completed during class time by both condition groups. Each exam was graded by the first author and given a percentage score (0-100%). Every exam included calculation-based problems similar to what the students were assigned in their problem sets. Students were graded for correctness and understanding of the concept with partial credit given to those who made minor calculation errors. Along with calculation-based problems, students were assessed using concept-based multiple-choice questions in the fifth exam.

Statistical control variables

In addition to the key measures of interest, the survey contained other measures that allowed for statistical control across conditions. This helps ensure that reported differences between conditions were likely *not* a result of preexisting differences between the participants who happened to enroll in each of the courses. Specifically, performance in University Physics I is a strong indicator of success in engineering classes, so participants were asked "What letter grade did you receive in PHYS 240 - University Physics I or an equivalent course?," which they rated from A to F, in standard +/- increments. Additionally, participants were asked to provide their current college GPA (0.00-4.00). Finally, participants' attendance grades were used as an additional control variable and were scored from 0 to 100 percent. Unless otherwise indicated, all results that allowed for the inclusion of multiple simultaneous effects (i.e. regression) included these as statistical control variables. Additionally, propensity score weighting was used to ensure both groups were statistically equivalent in terms of these variables.

Results

Bivariate Correlations

Most of the study variables were correlated with each other in expected ways. See **Table 1** in the **Appendix** for the full correlation table.

Notably, participants in the interteaching method condition self-reported reading their textbook more frequently before class. Surprisingly, they did not rate their levels of in-class participation significantly higher. Additionally, only Exam 4 showed a significant correlation with instructional methods, such that students in the interteaching condition achieved higher scores. The instructional method did not correlate significantly with any other variables.

Students who reported reading before class also had higher attendance grades. Engagement was inconsistently related to exam performance, though the effects were consistently in the positive direction.

Finally, exam grades were generally correlated positively with each other. Additionally, self-reported college physics grades were also positively correlated with most of their exam scores. Surprisingly, attendance was only correlated with Exam 4 grades. Additionally, GPA was generally correlated positively with scores for exam grades.

Propensity score weighting

Because this study was conducted in a real educational setting, it was not possible to randomly assign students to course sections. In contrast to a true experiment with experimenter-generated condition groups via random assignment, students self-enrolled in sections based on scheduling constraints, preferences, and seat availability. Although the instructor, course content, and assessments were constant across sections, the classes differed in meeting time and day of the week, as well as method of instruction. These structural differences introduce the possibility of systematic pre-existing differences between sections that are unrelated to the instructional method.

To take steps toward addressing this limitation, we employed propensity score weighting, a common and well-established approach in educational research for strengthening causal inference in observational designs by balancing treatment and comparison groups on potentially meaningful covariates that could relate to both the experimental condition and the dependent variable(s) of interest [25, 26]. Propensity score weights were estimated using variables that were likely to relate both to section enrollment and academic outcomes (tests). First, we included self-reported college GPA, as high-achieving students may be more proactive in course planning and thus more likely to secure sections at more desirable times. Second, we included students' self-reported college physics grades, since this is an important prerequisite for this introductory engineering course. Any systematic differences in prior preparation could bias estimates of instructional effects. Finally, we included course attendance as a final weighting variable, given that class meeting times and days could influence students' ability or willingness to attend consistently.

Propensity score weighting was conducted using energy balancing with an Average Treatment effect on the Treated (ATT) estimate, and the optimization procedure converged successfully, indicating that a feasible and stable set of weights was obtained. Before weighting, there were modest imbalances between treatment and control groups across the selected covariates. Standardized mean differences between the interteaching and control groups prior to weighting were not extreme (.9-.14), but they exceeded or approached commonly used thresholds for acceptable balance (.10). After propensity score weighting, balance improved substantially across all covariates (<.03), with an average absolute standardized mean difference across covariates of .01 after weighting. See **Figure 2** for a visual representation of the balances between the groups on these covariates before and after propensity score weighting.

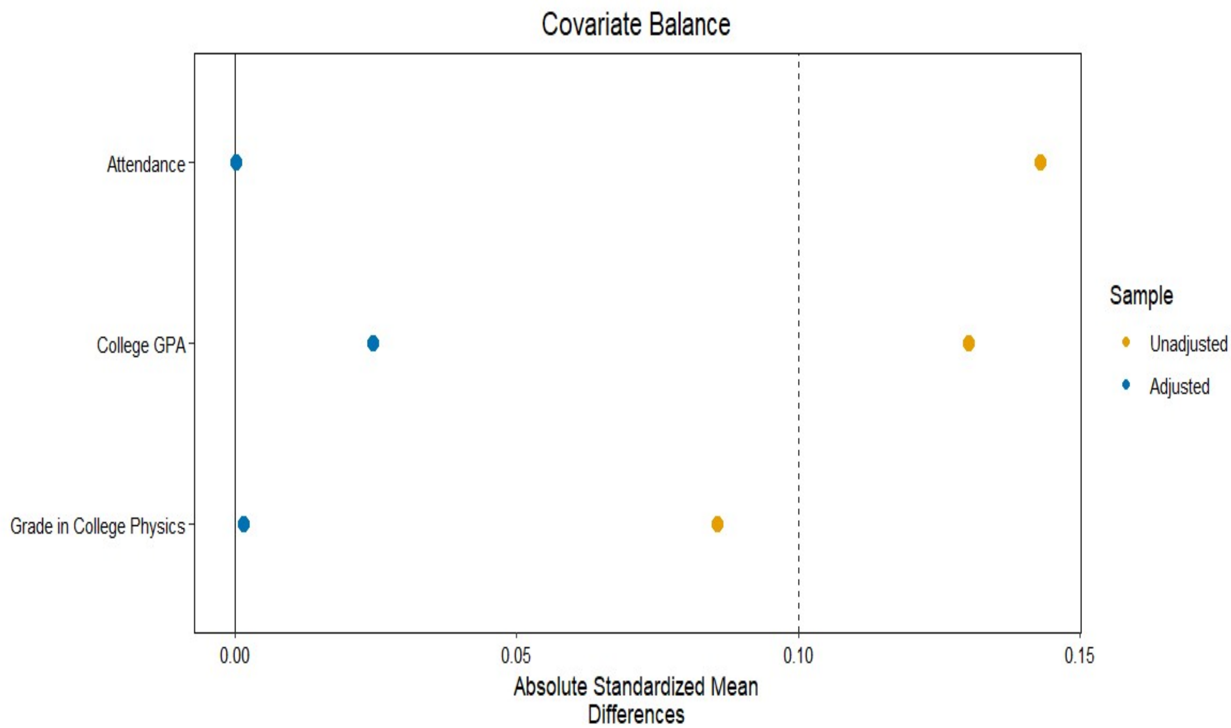


Figure 2 – Balance of covariates before and after propensity score weighting.

Perceived instructional effectiveness and self-reported engagement (H1)

To test our first hypothesis, we performed a series of propensity-score-weighted, regression analyses, such that the course engagement variables were entered as the main outcome variables and type of instruction was entered as the key predictor variable. Additionally, performance on Exam 1, course attendance, self-reported college GPA, and self-reported grade in University Physics I were entered as covariates. Self-reported levels of reading the course textbook also differed significantly across pedagogical conditions. Participants who received interteaching-based course instruction reported significantly higher rates of reading, compared to those who were in the control condition, $B(61) = 2.02$, $p < .001$. The overall model accounted for a significant proportion of the variance in self-reported reading, showing an especially large effect size $R^2 = .55$, $F(5, 61) = 14.8$, $p < .0001$.

In terms of self-reported in-class participation, there was not a significant difference by condition. Participants in the control and interteaching conditions reported similar levels of feeling that the course format encouraged their participation, $B(61) = .17$, $p = .49$.

Course performance

To test our second and third hypotheses, we conducted a series of regression-based analyses to examine differences between the conditions in terms of course performance (exam grades). First, we tested to see whether the two conditions were significantly different in their (cumulative) final exam grades. Next, we modeled a growth formula to predict improvement across exams.

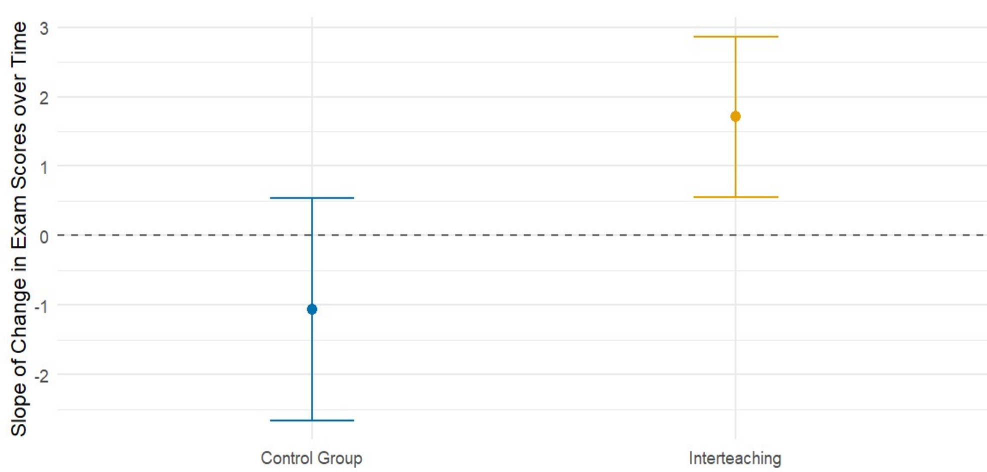
Finally, we performed a median-split of the two groups, in terms of their Exam 1 grades, to examine the differential growth/improvement rates for “high performers” versus “low performers.” All data and analysis files are available by request to the first author.

Testing for differences by condition (H2)

We tested the first part of our second hypothesis (H2a) using multiple regression analyses with propensity score weighting. Our main outcome variable was the participants’ scores on their final cumulative exam. The treatment effect (interteaching versus control) was entered as the main predictor, with scores for the first exam, attendance grades, self-reported grades in college physics, and self-reported GPA entered as statistical control variables. Our results showed a modest positive effect of treatment that trended toward statistical significance, $B(61) = 4.63$, $p = .07$. The overall model fit the data well and accounted for a sizable portion of the variance in final exam scores, $R^2 = .26$, $F(5, 61) = 5.60$, $p < .0001$.

Therefore, students who were taught using the interteaching method scored higher on their final examinations, after accounting for other variables that may have also affected their final exam scores. However, this positive effect did not meet the conventional benchmark for statistical significance and should be considered with some healthy skepticism.

To test the second part of our second hypothesis (H2b), we compared linear time-series functions between the interteaching and control groups, such that the five exam scores were entered as a sequential time-like variable (Exam 1 = T1, Exam 2 = T2, etc.). The model showed a significant difference between the two conditions, in terms of improvement over time, $t(65) = -2.80$, $p = .007$. Specifically, while the control group did not show significant improvement in exam scores over time, B = -1.06, 95% CI [-2.66, .54], the interteaching group did show significant improvement over time, $B = 1.71$, 95% CI [.55, 2.86]. See **Figure 3** for a visual representation of these effects.



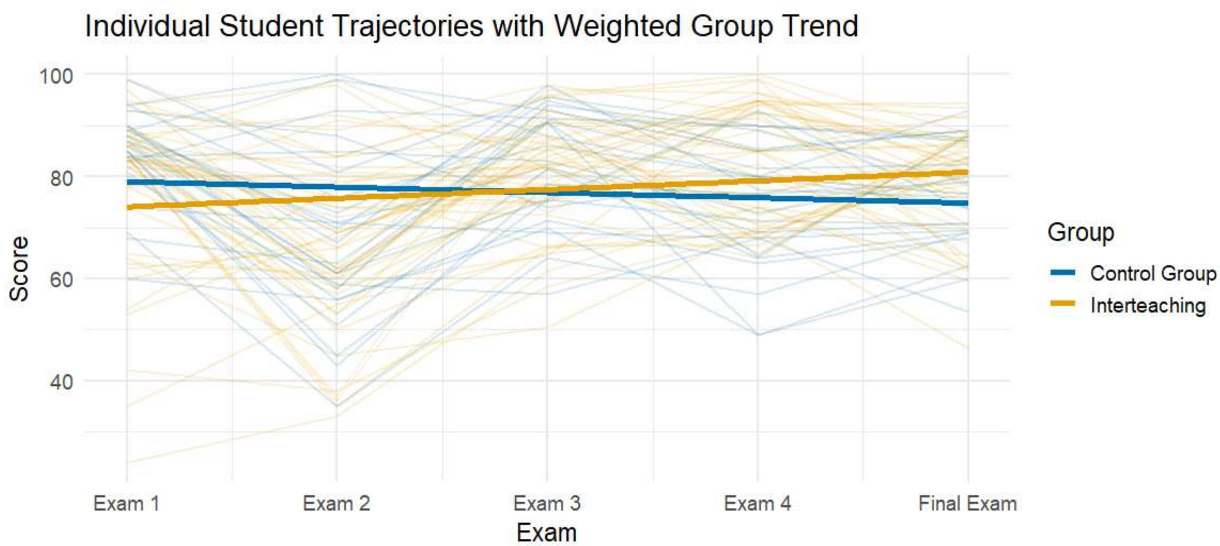


Figure 3 – Slope of change in exam performance over time by group (top). Exam scores across time (bottom).

Practically speaking, this means that, on average, students who were taught using the interteaching method saw an approximately 2% increase in their exam scores from one exam to another. Although 2% growth has questionable practical significance, it is important to note that this effect was the average across *all* participants. Therefore, participants who were already performing well (i.e., "high achievers") would likely have had a more difficult time improving their scores, as they approached the grade ceiling. To account for this, we next split the classes into two groups (high versus low achievers) and compare the effects of interteaching across these.

Testing for differences in improvement by condition and initial exam performance (H3)

To test our final hypothesis (H3), we performed a similar times-series analysis measuring changes in performance across the exams from the semester. However, this time we first performed a median split, where each of the classes was split based on the median score for their initial examination (Exam 1). Those who scored at or below the median were assigned to the "low performers" group, while those who scored above the median were assigned to the "high performers" group. We then modeled four separate time-series effects, such that each combination of type of achiever (low versus high) and type of instruction (interteaching versus control) had its own effect.

Starting with the control group, neither of the initial performance groups showed significant improvement in exam scores over the course of the semester, with the effects for both the low, B = -0.42, 95% CI [-2.58, 1.74], and high achievers, B = -1.76, 95% CI [-4.02, .50], not differing significantly from the null (0). Practically speaking, this means that, on average, participants in the control condition generally performed similarly across their five examinations.

The interteaching condition, by contrast, did show significant, positive change over time. However, this was only statistically significant for those participants who had *low* initial performance on the first examination, B = 2.55, 95% CI [1.11, 3.99]. Participants in the interteaching condition who scored above the median on the first exam, by contrast, did not show significant improvement in their exam scores over time, B = .37, 95% CI [-1.45, 2.19]. Taken together, this means that interteaching-based instruction helped to improve students with low initial performance in the introductory engineering course, without providing any particular benefit or burden to those who started the course performing well. See **Figure 4** for a visual representation of these improvement effects.

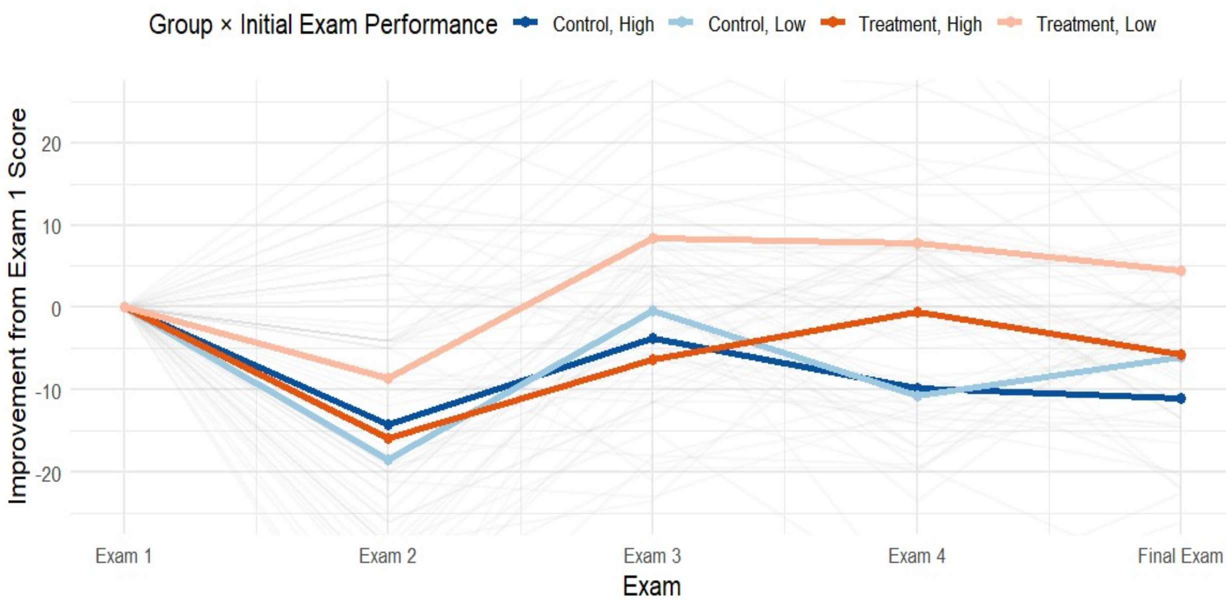


Figure 4 – Improvement in exam scores by initial performance and group.

Discussion

The current study provided a valuable contribution to the extant literature on interteaching by extending prior findings beyond psychology and other concept-focused disciplines into a calculation-intensive engineering mechanics context. Overall, the results provide converging evidence that interteaching can support stronger learning trajectories across a semester, with the most meaningful benefits concentrated among students who begin the course performing below the median on the first exam. This pattern aligns with the broader active learning literature showing that student-centered, practice-oriented instructional approaches tend to improve achievement and reduce failure rates relative to traditional lecture formats, especially in high-stakes gateway STEM courses. It also supports the idea that gateway mechanics courses represent a leverage point for interventions aimed at improving persistence and reducing early performance gaps that may otherwise compound over time.

Student engagement (RQ1)

Consistent with Hypothesis 1, students in the interteaching condition reported substantially higher self-reported reading frequency. This is an important result, given that the interteaching model relies on repeated contact with course content across multiple modalities and is intended to create a structured accountability mechanism for pre-class preparation. From an instructional standpoint, higher rates of textbook use may be particularly valuable in engineering mechanics, where students frequently benefit from revisiting worked examples and problem-solving frameworks outside of class. The finding also suggests that the interteaching may function as a practical "nudge" that increases students' use of core learning resources.

However, contrary to expectations, students did not report significantly higher in-class participation in the interteaching condition. There are several plausible interpretations that warrant consideration. One possibility is measurement limitation: participation encouragement was assessed using a single retrospective self-report item, which may not capture more granular participation behaviors interteaching aims to promote (e.g., time on task during problem solving, peer explanation quality, and number of attempted solutions). Another possibility is that students in the lecture condition still perceived substantial opportunity to participate through question asking and instructor interaction, even if the overall structure remained lecture-centered. Finally, it is possible that interteaching increased out-of-class participation, especially if some students experienced peer discussion as structured but not necessarily more "participatory" in the way they interpret that term. Future work should triangulate participation with behavioral indicators (e.g., discussion record sheets, observation protocols, or peer-evaluation data) rather than relying solely on self-report.

Course performance and learning trajectories (RQ2)

With respect to performance outcomes, the results indicate that interteaching did not produce a large shift in end-of-semester cumulative performance when examined as a simple between-group difference on the final exam, although the estimated effect was in the positive direction and approached conventional significance thresholds. In isolation, the modest final-exam difference could be interpreted conservatively as limited evidence for overall performance improvement. However, the time-series modeling provides a more informative picture of the learning process across the semester: students in the interteaching condition demonstrated statistically significant improvement across exams, while students in the lecture condition did not show evidence of improvement over time. This suggests that interteaching may be less accurately described as producing an immediate "boost" and more accurately described as supporting stronger learning momentum over the long term, consistent with the pedagogical logic of interteaching as a closed-loop feedback system that repeatedly targets student difficulties through guided clarification and structured practice over time.

Critically, the most practically meaningful result emerged when separating students by initial performance. Among students who scored below the median on Exam 1, those in the interteaching condition showed statistically significant improvement across subsequent exams, whereas low-performing students in the lecture condition did not. In contrast, students who began the semester performing above the median did not exhibit meaningful improvement in either condition, a pattern that may reflect ceiling effects rather than a lack of benefit. Taken together, these finding support Hypothesis 3 and suggest that interteaching may function as a form of scaffolding that helps initially struggling students recover, stabilize, and improve over time. This outcome is especially relevant in gateway mechanics courses, where early misunderstandings often compound and students who fall behind may struggle to re-enter the learning trajectory without structured opportunities to revisit core problem types and receive corrective feedback.

Positioning within prior engineering and interteaching scholarship

These findings extend the existing interteaching literature by demonstrating benefits in a problem-solving-intense engineering context where implementation challenges are often perceived as higher than in more discussion-oriented courses. While prior interteaching research documented improved exam performance relative to lecture and generally favorable student perceptions [16; 20], engineering mechanics courses present a different instructional ecology: students must learn not only conceptual principles but also multi-step solution procedures, representational fluency (e.g., free-body diagrams), and computational habits. The present study suggests that the interteaching cycle can be productively adapted to these demands by emphasizing peer problem-solving and targeted clarification. This result complements prior engineering education studies that have attempted to improve mechanics learning through active and flipped structures (e.g., Holdhusen [8], Hu et al. [9], and Lee et al. [12]), while offering an alternative model that may be more scalable for instructors who want structured active learning without fully converting the course into an intensive flipped classroom design.

Limitations

Although the current study goes a long way towards demonstrating the effectiveness of interteaching in introductory engineering coursework, it is not without limitations. Future research exploring the efficacy of interteaching in the context of engineering education should take steps to address these limitations, especially to improve upon the findings.

First, although sections of the class were randomly assigned to conditions, individual students were not able to be randomly assigned, meaning that the study does not reach the level of a true experiment with independent random assignment, in terms of establishing causal inference. Although we used propensity score weighting to account for preexisting differences between the classes in terms of GPA, performance in college physics, and attendance, it is still possible that other variables remain unaccounted for, which could explain the differences detected between our groups.

Second, the study was conducted in a very particular context, which could have implications for external validity. Specifically, it is possible that the type of institution, the instructor, or the types of students enrolled in engineering courses at this large, public institution in the mid-Atlantic region of the U.S. could have implications for the efficacy of this method in this specific context. The efficacy might not translate to other types of schools, students, or instructors. This limitation is confounded by the fact that our study had a very limited sample size.

Third, questions remain about the construct validity of the key measures used in the study. For instance, the measures of student engagement and instructional effectiveness were single items and self-reported. Single-item measures may fail to capture multifaceted aspects of constructs of interest. For instance, the question "How effective do you think the teaching method used in this class was for your learning?" does not take into account that a student might feel equally positive or negative about *any* method of instruction that might be used. Self-reported items also have significant limitations because students may respond falsely out fear of being judged or evaluated negatively, even though they were reassured that their responses were anonymous. Additionally, retrospective reports are also vulnerable to bias because students may have a more or less rosy picture of the instruction at the end of the term than they did during the term, based on their final course grade, stress associated with the last weeks of class, or better report with the instructor over time.

Finally, there are various other limitations that likely remain, but were not addressed explicitly. Research on pedagogical efficacy is often limited by the practical concerns of performing actual instruction with real students. Laboratory studies can be a great way to reduce these limitations, but they also lack the ecological validity of a class-based study, as well as the scale of a full semester of course instruction. Future research in actual classroom settings should aim to address such concerns by leveraging true experimental designs (if possible), replicating the study in other contexts with larger sample sizes, and triangulating the effectiveness of the method by using different operationalizations of key variables (e.g., using observed, archival, or behavioral data, rather than self-report).

Future directions

Although the current study demonstrates significant benefits of the interteaching method for course performance, particularly for students who start the semester in the lower half of the class, there is still much that needs to be investigated about this method. In addition to addressing limitations, future research should be focused on replication, extension, and interrogation of the current findings.

First, as noted previously, gaps in STEM performance can often be attributed to demographic variables like race, gender, and socioeconomic or first-generation status. Given the small sample size of the current study and limited access to diverse demographic groups, we could not evaluate the efficacy of interteaching for addressing such demographics-based gaps in performance. Future research should recruit diverse samples and specifically compare these groups to evaluate whether interteaching can be used as a method to address systemic gaps in performance based on demographic categories.

Additionally, persistence in STEM is not just a matter of performance, but also related to feelings of belonging, mentorship, and career/major identity attachment. Future research should evaluate whether the interteaching method of instruction also produces benefits for other factors that might affect the long-term persistence of STEM majors.

Furthermore, despite the continued persistence of lecture-based instruction in engineering coursework, other types of active learning techniques are also being integrated into engineering coursework. It would be beneficial to assess whether interteaching has benefits above and beyond other types active learning techniques, as well as traditional lecture. Conversely, it would be equally beneficial to evaluate whether interteaching is lacking in important ways that other active methods, like flipped classroom, are not.

Finally, the specific mechanisms and mediators that explain *why* interteaching is effective are still not well understood in the literature. This is especially true for STEM-based interteaching, which is already limited. It will be important for future research to investigate such mechanisms, so that instructors can understand which components of the method are absolutely central to the positive outcomes for engineering-focused educational outcomes and which can be more readily adapted or even cut, without negatively impacting performance.

Conclusion

In high-stakes engineering gateway courses, such as statics and introductory solid mechanics, early performance gaps can quickly become durable barriers to persistence in engineering degrees. The present study extends the literature on the interteaching pedagogical method into a calculation-intensive mechanics context and provides initial evidence that a structured, feedback-responsive pedagogy can support meaningful improvement over time, particularly for students who begin the course with below-average performance. Although average gains in performance in the interteaching condition were modest, the results clearly show that interteaching functioned as a form of scaffolding that helped initially lower-performing students recover across subsequent exams, without negatively affecting students who started strong. Additionally, it was shown to increase students' use of core course resources, which can be a difficult metric to change.

These findings suggest that interteaching is a promising and practically adoptable approach for improving learning in foundational engineering coursework. More broadly, this research highlights the value of shifting instructional time toward targeted clarification and guided peer problem solving, not only to raise achievement, but also to reduce the compounding disadvantages that often emerge in gateway STEM classes. In conclusion, interteaching may be an effective tool to educate, expand, and encourage the engineering workforce of the future.

References

[1] Chen, X. & Soldner, M. (2013). *STEM Attrition: College Students' Paths into and out of STEM Fields* (NCES 2014-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubs2014/2014001rev.pdf

[2] National Academy of Sciences, National Academy of Engineering & Institute of Medicine. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads.* The National Academies Press. https://doi.org/10.17226/12984

[3] National Science Board & National Science Foundation. (2024, May). The STEM labor force: Scientists, engineers, and skilled technical workers (NSB-2024-5). *Science and Engineering Indicators 2024*. https://ncses.nsf.gov/pubs/nsb20245

[4] Hewitt, N. M. (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences*. Avalon Publishing.

[5] Oliveira, A. W., Musah, R. A., Zacharia, Z. C., Pavlou, Y., Sgro, C. M., Lathrop, R., Kim, Y., Espinosa, K., Japashov, N. & Park. J. (2025). The Persistent Problem of Undergraduate Student Attrition in STEM. *International Educational Review, 3*(1), 85-111. https://doi.org/10.58693/ier.316

[6] Stains, M., Harshman, J. Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M. McKay, T. A., Michelotti, N., Palmer, M. S., Plank, K. M., Rodela, T. M., ..., Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470. https://doi.org/10.1126/science.aap8892

[7] Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

[8] Holdhusen, M. H. (2015, June 14-17). *A "flipped" statics classroom* [Paper presentation]. 122nd ASEE Annual Conference & Exposition, Seattle, WA, USA. https://doi.org/10.18260/p.23356

[9] Hu, Y., Montefort, J. M., & Tsang, E. (2015, June 14-17). *An Innovative Redesign of Statics: Approach and Lessons Learned* [Paper presentation]. 122nd ASEE Annual Conference & Exposition, Seattle, WA, USA. https://doi.org/10.18260/p.23528

[10] Komarek, R. & Bielefeldt, A. R. (2015, June 14-17). *Impact of Teaching Style on Student Learning and Satisfaction in Statics Courses* [Paper presentation]. 122nd ASEE Annual Conference & Exposition, Seattle, WA, USA. https://doi.org/10.18260/p.24227

[11] Lee, L. S., Hackett, R. K., & Estrada, H. (2015, June 14-17). *Evaluation of a Flipped Classroom in Mechanics of Materials* [Paper presentation]. 122nd ASEE Annual Conference & Exposition, Seattle, WA, USA. https://doi.org/10.18260/p.24031

[12] Fedesco, H. N. & Troy, C. (2016, June 26-29). *Why This Flip Wasn't a Flop: What the Numbers Don't Tell You About Flipped Classes* [Paper presentation]. 123rd ASEE Annual Conference & Exposition, New Orleans, LA, USA. https://doi.org/10.18260/p.27203

[13] Lee, A., Zhu, H., & Middleton, J. A. (2016, June 26-29). *Effectiveness of Flipped Classroom for Mechanics of Materials* [Paper presentation]. 123rd ASEE Annual Conference & Exposition, New Orleans, LA, USA. https://doi.org/10.18260/p.26907

[14] Cornwell, P., Feier, I., Marco, N. J., & Ozment, D. (2023, June 25-28). *Flipping a Required Mechanics Course with Different Instructors* [Paper presentation]. 130th ASEE Annual Conference & Exposition, Baltimore, MD, USA. https://doi.org/10.18260/1-2--43722

[15] Cornwell, P., Snyder, M., & Anderson, M. (2023, June 25-28). *The Evolution of a Flipped Dynamics Course* [Paper presentation]. 130th ASEE Annual Conference & Exposition, Baltimore, MD, USA. https://doi.org/10.18260/1-2--44450

[16] Saville, B.K., Zinn, T.E., Neef, N.A., Norman, R.V., & Ferreri, S.J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39*(1), 49-61. https://doi.org/10.1901/jaba.2006.42-05

[17] Sturmey, P., Dalfen, S., & Fienup, D. M. (2015). Inter-teaching: A systematic review. *European Journal of Behavior Analysis*, *16*(1), 121-130. https://doi.org/10.1080/15021149.2015.1069655

[18] Krebs, C. A., Kuhn, S. A., Brewer, A. T., & Diller, J. W. (2023). Using interteaching to promote online learning outcomes. *Journal of Behavioral Education*, *32*(1), 76-89. https://doi.org/10.1007/s10864-021-09434-5

[19] Shaw, S. K., Posey, J. L., Zane, T., & Weiss, M. J. (2024). Comparing interteaching and discussion forums in an asynchronous online classroom. *Journal of Applied Behavior Analysis*, *57*(4), 989-998. https://doi.org/10.1002/jaba.2905

[20] Hurtado-Parrado, C., Pfaller-Sadovsky, N., Medina, L., Gayman, C. M., Rost, K. A., & Schofill, D. (2022). A systematic review and quantitative analysis of interteaching. *Journal of Behavioral Education*, *31*(1), 157-185. https://doi.org/10.1007/s10864-021-09452-3

[21] Cezeaux, J.L. & Keyser, T.K. (2018, June 23-27). *Introducing active learning strategies into an undergraduate engineering physiology course* [Paper presentation]. 125th ASEE Annual Conference & Exposition, Salt Lake City, Utah, United States. https://doi.org/10.18260/1-2--30714

[22] Parthasarathy, R. & Jollands, M. (2012, January). *Interteaching in a second year chemical engineering undergraduate project-based course* [Paper presentation]. Proceedings of the 23rd Annual Conference for the Australasian Association for Engineering Education, Melbourne, Victoria, Australia.

[23] Boyce, T.E. & Hineline, P.N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *Behavioral Analysis, 25*(2), 215-226. https://doi.org/10.1007/BF03392059

[24] Saville, B.K. (2013, January 31). *Interteaching: Ten Tips for Effective Implementation*. Association for Psychological Science. https://www.psychologicalscience.org/observer/interteaching-ten-tips-for-effective-implementation

[25] Powell, M. G., Hull, D. M., & Beaujean, A. A. (2020). Propensity Score Matching for Educational Data: Worked Examples. *The Journal of Experimental Education, 88*(1), 145-164. https://doi.org/10.1080/00220973.2018.1541850

[26] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Appendix

Table 1 – Bivariate correlation table.

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Condition	—										
2. Reading	$.70^{***}$	—									
3. Participation	.09	−.19	—								
4. Exam 1	−.20	−.13	.13	—							
5. Exam 2	−.03	−.06	.12	$.42^{*}$	—						
6. Exam 3	−.06	.05	.03	$.46^{***}$	.37	—					
7. Exam 4	$.33^{*}$	$.25^{*}$	.14	$.33^{*}$	.36	$.40^{*}$	—				
8. Final Exam	.12	.00	−.02	$.50^{***}$	$.50^{***}$	$.29^{*}$	$.35^{***}$	—			
9. Attendance	.07	$.31^{*}$	.16	.16	.20	.23	$.34^{*}$	.08	—		
10. College physics grade	−.04	.00	.16	$.26^{*}$	.19	$.25^{*}$	.10	.13	−.12	—	
11. College GPA	.06	.12	.20	$.35^{***}$	$.54^{***}$	$.35^{***}$	.17	$.33^{*}$	.12	$.45^{**}$	—

Notes. Condition: 0 = Control, 1 = Interteaching. Values are Pearson correlation coefficients. Significance indicators are based on *p* values: * $p < .05$. ** $p < .01$. *** $p < .001$.