12-1-2021

Frequent small group interactions improve student learning gains in physics: Results from a nationally representative pre-post study of four-year colleges

Miguel Rodriguez
Florida International University

Geoff Potvin
Florida International University

Recommended Citation

Rodriguez, Miguel and Potvin, Geoff, "Frequent small group interactions improve student learning gains in physics: Results from a nationally representative pre-post study of four-year colleges" (2021). *All Faculty*. 256.
https://digitalcommons.fiu.edu/all_faculty/256

PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **17,** 020131 (2021)

Frequent small group interactions improve student learning gains in physics: Results from a nationally representative pre-post study of four-year colleges

Miguel Rodriguez and Geoff Potvin

Department of Physics and STEM Transformation Institute, Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA

(Received 22 June 2020; revised 17 February 2021; accepted 24 June 2021; published 20 October 2021)

Student-centered learning has been shown to be more effective than traditional instructional methods, but deeper investigation is necessary to identify how specific classroom practices lead to improved conceptual learning and student attitudes. The purpose of this study was to identify classroom practices associated with gains in student conceptual understanding and physics identity using pre-post survey data taken from a nationally representative sample of first-semester physics students in four-year colleges and universities. From this sample, we found that students who reported working in small groups during every class saw greater gains in their conceptual understanding of physics than students who did not. Other classroom practices, such as the frequent use of computer simulations, using equipment, and performing labs, were also found to increase student conceptual gains. This work provides further evidence that certain instructional practices—small group learning, in particular—provide benefits to students in college physics classrooms.

DOI: 10.1103/PhysRevPhysEducRes.17.020131

I. INTRODUCTION

In the past decade, the President's Council of Advisors on Science and Technology has called for a 33% increase in bachelor's degrees in Science Technology Engineering and Mathematics (STEM) disciplines [1]. Physics is an important topic for many science and engineering majors, and can often serve as a gatekeeper for those students [2]. In the past 20 years, physicists and physics educators have found that students taught in traditional lecture physics courses often lack a conceptual understanding of physical principles, even when they can solve many standard problems [3]. There is a clear need to support students' conceptual understanding of physics. Besides the effort to improve student learning, there is also a need to increase the number of physics majors in the United States. Compared to biology and chemistry, for example, physics has seen significantly less growth in terms of the number of students majoring in physics [4]. Though the raw number of majors has rebounded in recent years from an historic low in the late 1990s, the fraction of incoming freshmen who choose physics as a major is less than it was forty years ago [5]. Also, there is a stark dearth of participation of women and students identifying with traditionally marginalized racial or ethnic identities in physics [6,7]. It has been variously argued [8–10] that all of these problems may be solvable through reformed physics instruction such as the use of student-centered approaches that create more inclusive learning environments by giving students more opportunities to co-construct their understanding of physical systems.

In the United States, educators, policy makers, and education researchers have promoted instructional methods that incorporate more active learning approaches in classrooms of all subjects at the K-12 and postsecondary levels since at least the 1980s [11]. Ideas of "active" learning preexist such initiatives and, since the initiation of these efforts, many studies indicate that active learning experiences in STEM lead to improved student outcomes. One recent meta-analysis that considered results from over 225 separate studies (including 29 in physics specifically) found that for STEM courses, "students in classes with traditional lecturing were 1.5 times more likely to fail than were students in classes with active learning" [12] (p. 8410). In another, earlier study, it was reported that students in introductory physics courses taught with "interactive engagement" methods experienced normalized gains two standard deviations higher than students taught with traditional methods [13]. The definitions used in these works for active learning and interactive engagement are broad, somewhat nonspecific, and could include many different classroom practices and teaching methods that are rooted in different theoretical traditions and with different goals in mind. While the earlier work provides strong evidence for the generic benefits of "active

---

learning," it is not clear which elements of these approaches are particularly beneficial. Further investigation is needed on what specific aspects of active learning are effective for introductory physics instruction. Many different approaches that might all be characterized as active learning may not be directly comparable or commensurate.

Collaborative learning, a subset of practices that might be captured under the broad umbrella of active learning, occurs when students work together on a shared learning activity or towards a common objective [14]. As a particular instantiation of collaborative learning, small group learning has been shown to be an effective instructional strategy which leads to greater academic achievement, improved attitudes toward learning, and increased persistence across STEM fields [15]. Of the studies that have explored the effectiveness of small group learning, relatively few have delved into the particular facets of what students do in small groups and what makes them effective. Much work in this space has been conducted in the context of single institutions and/or have focused on specific pedagogies, such as the use of Modeling Instruction [8], Peer Instruction [16], or student-centered activities for large enrollment undergraduate programs (SCALE-UP) [17], etc. Thus, a limitation of the existing literature on small group learning is that the cross-classroom (or cross-contextual) features of these practices have not been well understood or subjected to comparable analyses. Furthermore, the impacts of various methods on students' attitudes, including their physics identities [18–20] have not been studied in as great a detail as students' conceptual learning. Several reformed pedagogies [21] continue to find that students suffer declines in their physics-related attitudes despite improved learning gains. This limits the possibilities of recruiting and retaining a larger, more diverse population of physics students, since attitudinal constructs such as physics identity [22] have been shown to predict future physics-related career choices more than past performance alone.

Thus, the current study sets out to answer the following two questions, in the context of a nationally representative sample of students enrolled in introductory four-year university-level physics (both calculus based and algebra based):

(i) How does the frequency of working in small groups predict students' conceptual learning gains and physics identity gains?

(ii) For students who regularly work in small groups during their physics classes, what instructional strategies are available to them and characterize their class experiences?

II. SURVEY DATA

The data analyzed in this chapter were drawn from the Conceptual Understanding and Physics Identity Development (CUPID) survey (see Supplemental Material [23]). This instrument was developed in 2014-2015, drawing on earlier work (e.g., the Sustainability and Gender in Engineering "SAGE" instrument) [24]. Two survey instruments were developed, one for use as a pre-course instrument and the other for use as a post-course measure. Both instruments contained the same questions to measure conceptual understanding and physics identity so pre-to-post results could be reported but the two surveys varied on the other questions. The pre-survey included a total of 44 multi-part questions and probed students about their career intentions, previous science learning, demographics, family support for science, and prior academic performances. The post-survey also included 44 questions but instead asked students to detail the physics course they had just taken, in addition to the common constructs indicated. Primarily, these latter items were structured to offer frequency response options (e.g., "every class," "once per week," "once per month," "1–2 times a semester," "never"). A preliminary draft of each survey was distributed to a selection of students and faculty, who provided feedback for improvement and evidence in support of content validity of the various items. The complete, final version of both surveys is available in the Supplemental Material [23]. Further details of relevant items are presented here.

The survey data used in this study were collected from students enrolled in first semester, introductory physics (both algebra- and calculus-based) recruited from a stratified random sample of four-year colleges and universities nationwide. Initially, a list of all colleges and universities in the U.S. (maintained by the National Center for Education Statistics) was acquired. It was then stratified to account for two considerations: institution size and course type. For institution size, the sample was divided into small, medium, and large institutions (see Table V in the Appendix) to ensure the sample fairly represented the substantial number of small institutions in comparison to the much sparser, and much larger, state institutions in the U.S. The course type was also considered so that the proportion of algebra- and calculus-based introductory physics was roughly similar to the national proportion. Once this stratification was complete, 32 institutions in total were recruited to deploy two surveys near the beginning and end of their regular stream, introductory physics I (e.g., mechanics) classes. The surveys were distributed to faculty, instructors, and departmental leaders who were encouraged to include their own classes and others who were available for study, resulting in mixed participation. Thus, the sample at each institution might include multiple sections or multiple physics courses (algebra- and calculus-based physics courses). Also, due to the pseudonymity of the IRB protocol, we did not collect the specific course sections student respondents were enrolled in.

In the end, 1704 responses were received for the pre-CUPID survey and 621 students in the post-CUPID survey spread across 22 institutions in the fall of 2015. Students were given the surveys in their classrooms as paper and pencil instruments. Using students' self-reported identifiers (primarily, unique student numbers), the pre- and post-surveys were matched, giving a final sample of 371 individual students enrolled at 19 different institutions. The distribution of students' affiliated institutions in our sample is provided in Table V of the Appendix. Even though the matched data were only a subset of the originally recruited sample, the matched sample had a distribution closer to the national average than the typical demographics in PER of students in algebra- and calculus-based introductory physics courses. In the sample there were 138 students in algebra- and 201 in calculus-based physics courses (with 32 NA's). The percentage of students enrolled in calculus-based physics courses in our sample is 59%, which is much closer to the national average of 55% [6] than the historic average of 84% found in published PER studies [25]. Comparisons of the demographics of our sample to the four-year college populations and the demographics of published PER research is shown in Table IV in the Appendix.

Portions of the Force and Motion Conceptual Evaluation (FMCE) were chosen to measure students' physics conceptual knowledge [26]. Since these items were part of a longer survey, for time, they needed to be trimmed so that the entire CUPID survey took approximately 20–25 min. With this in mind, only 17 questions out of the 43 that appeared in the original FMCE were adopted into the surveys. The items with strongest validity and reliability according to prior work [27] were prioritized for our survey. The questions chosen tested key concepts such as Newton's three laws and momentum. The following problems from the FMCE were selected (in the order that they appear on the CUPID survey): the sled (questions 1–7), office chair (question 39), collision (questions 35–37) and force vs time graph (questions 14,16–19, 21). On the CUPID survey (see Supplemental Material [23]) these appear as questions 9–25 on the presurvey and questions 25–41 on the postsurvey.

The Force Concept Inventory (FCI) and FMCE are both widely used for measuring conceptual understanding in general physics 1, but they target slightly different topics [28]. There are a few reasons the FMCE was used over the FCI. The first is that the FMCE covers kinematics and Newtonian concepts in one dimension, whereas the FCI explores these concepts in more than one dimension. This was important because the FMCE had questions that required less time than the FCI and for practical constraints the CUPID survey could not take more than 20–25 min. This time restriction was important because designing a survey that was too long may have inhibited student participation. Also, the FMCE has been shown to have better validity evidence than the FCI [28].

The impacts of various classroom experiences on measures of physics affect; specifically, on physics identity were also investigated. Physics identity is a theoretical framework that, generally, allows a researcher to understand how students see themselves as a physics person. For this study, physics identity is treated (and measured) as a quasi-trait (meaning that attitudes and beliefs are semistable over time) that is associated to three related subconstructs: physics interests, recognition beliefs, and performance or competence beliefs [22,24,29].

III. ANALYSIS

Linear regression analysis was used to determine which classroom experiences are associated with learning and physics identity gains. The pre- and post-FMCE as well as physics identity scores were standardized before running our analysis (from 0 to 1). We did not use Hake normalized gain because of its statistical bias towards higher pretest scores [30]. Instead, we opted to assess effect sizes using Cohen's d along with regression coefficients and raw histograms. Since the sample includes students who had a matching pre-to-post survey, the change in both FMCE responses and physics identity could be calculated. The items that had "frequency" (of various classroom experiences) responses were treated as categorical variables, using the "never" response option as the basis of comparison to all the other options in our regression models.

Before the model building phase, the R package *Amelia* was first used to perform multiple imputation on the dataset to maximize the statistical power and reduce statistical biasing due to nonresponse, which was about 2% in the items used in the current analysis [31]. Multiple imputation is a "best practices" approach to preserve statistical power and handle missing data, and is much more robust than the common practice of listwise deletion (e.g., deleting any individual with any missing responses), among other approaches. Multiple imputation takes the distribution of the observed data collected to iteratively estimate the missing values. The imputation process is numerical and incorporates the extent to which other observed variables in the imputation model can predict its true values [32]. Thus, a new source of variance, reflecting the uncertainty of the missingness, is introduced into the process, which is the reason to carry out the imputation process multiple times. In this case, twenty imputations were generated. The *Amelia* algorithm uses a bootstrap-based expectation-maximization algorithm that efficiently handles the missingness and random error. A ridge prior was added to account for the collinearity among the variables in the imputed model by setting the empri setting to 10%. (A ridge prior shrinks covariances but keeps the variances and means the same.)

After the imputation stage, the twenty complete imputed datasets were analyzed in parallel using linear regression. Then the results from the analysis on the individual imputed datasets were pooled together using Rubin's rule [33], to give a combined and presentable result. The package *Zelig* was used for these last two stages with the *Amelia*-generated data [34]. *Zelig* is intended to be used with *Amelia*-generated data files and appropriately handles the within- and between-imputation variance in generating estimates.

After identifying significant effects in the linear regressions, the effect size of various factors was calculated to understand the practical importance of the findings using the package *effsize* [35]. Descriptive statistics were also used to explore what students reported happening in their classroom. The figures in this paper were generated by combining the imputed datasets and the graphs represent the fraction of student responses. The descriptive analysis was used to generate a picture of what the students experienced in their courses, specifically to compare between those people who reported working in groups everyday (which, as will be reported below, was found to be a significant predictor of learning gains) versus those that did not.

IV. RESULTS

A. Modeling FMCE gain

In the linear regression model appearing in Table I, students who reported working in small group activities every class had statistically significant ($p < 0.001$) improvements in their standardized FMCE gain. Translated to the 17 FMCE items asked on the survey, students who reported working in small group activities every class had an improved gain of nearly 2.6 items (or 15% of the response scale), compared to those who reporting never working on small group activities. The other frequencies (1-2 times per semester, once per month and once a week) were not significantly different from the students who reported not having worked on small group activities during the semester. Note that the linear regressions reported in Tables I and II were conducted for each imputed dataset and then pooled together to give the presented results, as described previously.

After noting such a strong correlation, we estimated the effect size between students who reported doing small group work every day (hereafter labeled group A) versus those that had small groups less frequently (labeled group B). The effect size of the difference in gain between these two groups was also calculated and found to be "large" according to Cohen's categorization [36], at 0.86 (95% confidence interval: 0.77 to 0.92). This underscores the real-world importance of this result, providing evidence that regular engagement in small group learning may be beneficial to learning physics concepts. This is consistent with other literature that has argued in favor of the benefits of this type of collaborative learning [37,38].

TABLE I. Linear regression of FMCE on frequency of small group activities.

Model	Linear model	Factors	ß	SE	p
1	FMCE Gain~Did small group activities	1–2 times/semester	0.13	0.21	0.62
		Once a month	−0.31	0.27	0.31
		Once a week	0.18	0.13	0.21
		Every class	0.85	0.17	***

$^{***}p < 0.001$. Adjusted $R^2 = 0.07$. N = 371.

To illustrate this difference graphically, we present a histogram of FMCE gain (post minus pre) in Fig. 1 for the group of students who reported small group work every day (group A) versus those that had small groups less frequently (group B). There were 41 students in group A, 318 in group B and 12 students that were not categorized due to leaving the question empty (and thus had responses imputed). To create Figs. 1 and 2, the imputed datasets were combined and the graphs report the average fractions of students in each category. The gain of group A is clearly spread across positive gain (mean gain 3.2, median of 2) while that of group B is clearly centered around zero (mean gain of 0.63, median of 0). Also, there is a notably greater fraction of students in group A whose scores skew particularly high. Details on the sample demographics and institutions of the data including groups A and B are provided in the Appendix.

In addition to the small group work item appearing in Table I, students were also asked details of the nature of their course experiences in order to more deeply understand the opportunities available to learn. Note that these latter items were too highly correlated with the group work item in model 1 to simply add them to the original model; doing so would violate basic model assumptions (e.g., mutual independence of predictors). Instead, the effect of these variables is presented as four independent models in Table II, which summarize significant associations between other classroom experiences and standardized FMCE gain. In model 2, it was found that students who reported using equipment every class had a statistically significant gain (as opposed to those that reported never using equipment), $p < 0.01$ and those who reported using equipment weekly also showed a borderline significant gain, $p = 0.04$. Model 3 shows that students who reported that their groups collaborated with other groups every class had improved gains, $p < 0.01$, while those who reported engaging in this practice monthly also had a borderline significant gain, $p = 0.03$. In model 4, the students who reported using computer simulations once a week ($p < 0.01$) or every class ($p < 0.001$) had improved FMCE gain. Lastly, model 5 shows statistically significant improvement for students who reported working on labs or projects every class, $p < 0.001$.

B. Modeling physics identity gains

One of the primary goals of the broader CUPID study was to investigate how students' physics identities change from the beginning to the end of a first semester introductory physics course. To this end, regression models analogous to those in the previous section were constructed, regressing physics identity gain (post- minus

TABLE II. Linear regression for FMCE and frequency of small group activities.

Model	Linear model	Factors	ß	SE	p
2	FMCE Gain ~ Your group used equipment ($R^2 = 0.02$, $N = 371$)	1–2 times/semester	−0.16	0.43	0.74
		Once a month	0.17	0.34	0.62
		Once a week	0.19	0.12	0.04*
		Every class	0.48	0.16	2×10^{-3}**
3	FMCE Gain ~ Your group collaborated with other groups ($R^2 = 0.03$, $N = 371$)	1–2 times/semester	0.08	0.17	0.70
		Once a month	0.33	0.18	0.03*
		Once a week	0.06	0.14	0.09
		Every class	0.69	0.20	***
4	FMCE Gain ~ Your group used computer simulations ($R^2 = 0.05$, $N = 371$)	1–2 times/semester	0.26	0.19	0.21
		Once a month	0.24	0.19	0.14
		Once a week	0.45	0.15	2×10^{-3}**
		Every class	0.79	0.21	***
5	FMCE Gain ~ You worked on labs or projects ($R^2 = 0.04$, $N = 371$)	1–2 times/semester	0.11	0.47	0.69
		Once a month	−0.11	0.34	0.94
		Once a week	0.17	0.12	0.06
		Every class	0.60	0.16	***

***$p < 0.001$.
**$p < 0.01$.

pre-, standardized on a scale 0 to 1) on the same classroom experiences discussed above. In the end, *no* significant associations between classroom practices and physics identity gains were found. While unfortunate for the overall goal of this study, this finding is not entirely unexpected, since students' physics identities have been argued to be well formed by the end of secondary education [22,24] and university physics curricula have not traditionally paid significant attention to affective outcomes in their construction, even among reformed classes [21].

C. Comparative description of students who did small group activities every class

In addition to these results, we wanted to understand what learning opportunities the students who reported working in small groups every class ("group A") experienced differently than the students who did not report working in small groups every class ("group B"). To avoid the collinearity problems mentioned previously, a descriptive analysis was conducted to compare groups A and B on the factors appearing in Table II as well as several other factors that were not, on their own, predictive of FMCE gain (and, hence, did not appear in separate models in Table II). Figure 2 presents the proportion of responses for these two groups on a total of nine different items characterizing classroom experiences; the responses of group A are significantly different than group B on seven. Several of these results help to characterize the group of students who reported small group work every class. To test whether the apparent differences appearing in the figure are statistically significant, Wilcoxon rank-sum tests (otherwise known as Mann-Whitney U tests) were performed on these items. For example, 61% of group A students reported working at tables every class (Q16a) compared to only 21% in group B ($p < 0.001$). Similarly, group A reported "facing their group mates" (Q16b) much more regularly than group B ($p < 0.001$). More than 75% of students in group A reported using equipment at least once per week (Q16c), whereas only 55% of group B did ($p < 0.001$). Nearly 60% of students in group A responded that they collaborated with other groups (Q16e) at least once per week, which was also statistically significantly different than group B ($p < 0.001$). Students were asked how often they were expected to simultaneously use numbers, formulas, graphs, and words to solve a physics problem (Q16g) which was also found to be significantly different between the two groups ($p < 0.001$). 66% of the

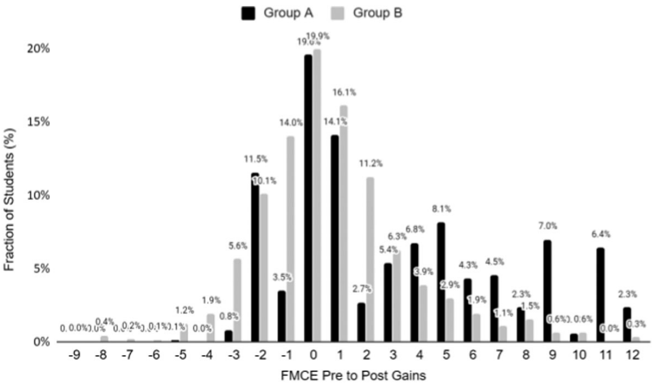

FIG. 1. The pre-to-post FMCE gains for students that reported working in small groups every class (group A) and students that did not (group B).

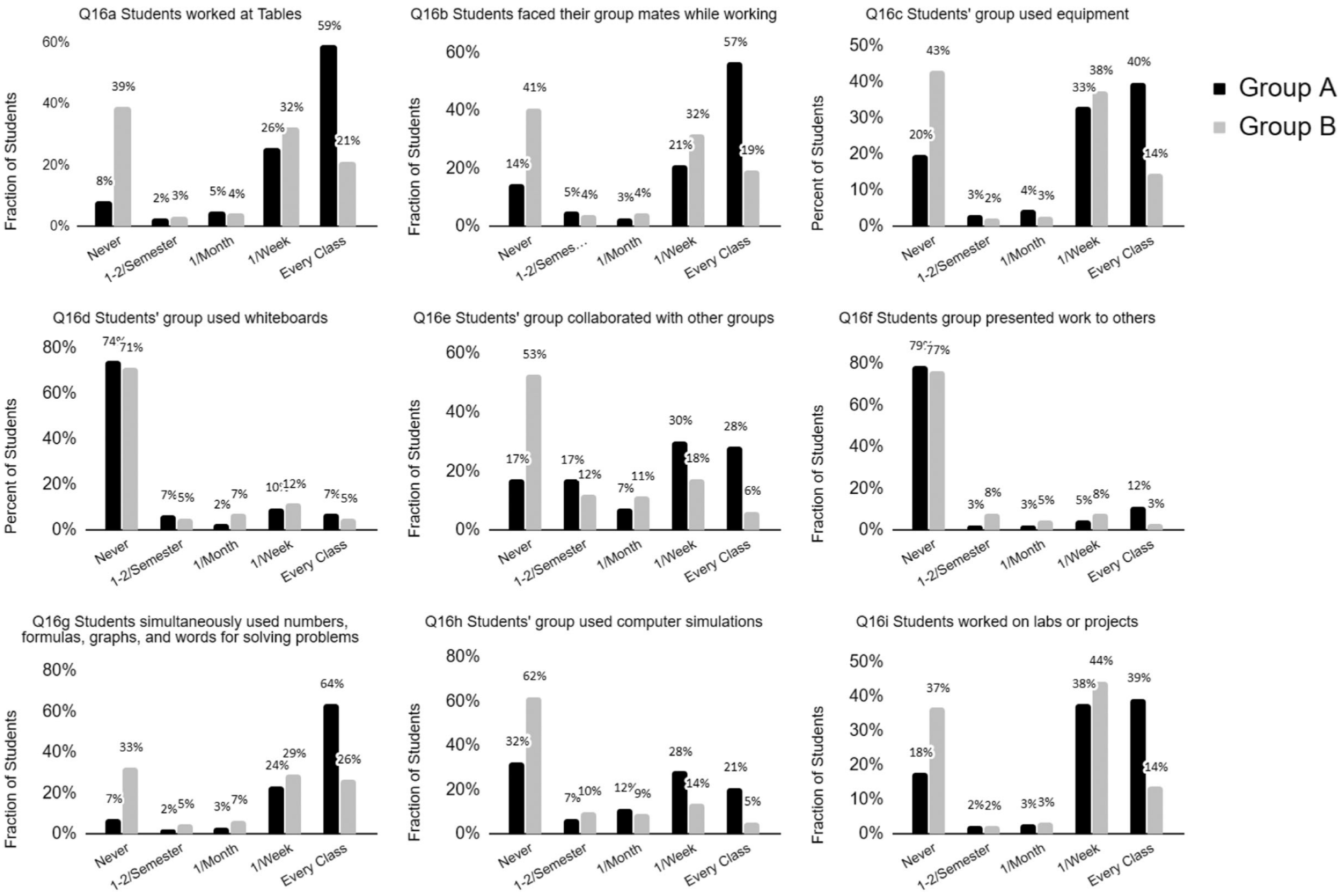

FIG. 2. Histograms of group work details for groups A and B. Question 16 probed students on their group work and asked the frequency of different types of classroom practices throughout their first semester of introductory physics.

students in group A said they had to do this every class and a further 24% said they had to do this weekly, meaning that 90% of this group reported this practice at least weekly. By contrast, only 58% of group B responded at least once per week, with 33% reporting to never have been expected to engage in this type of work in class. The students in group A reported much more exposure to computer simulations (Q16h); only 32% of group A students never had this experience in class compared to 62% of group B students ($p < 0.001$). group A students also reported more regularly working on labs or projects (Q16i, $p < 0.001$). Finally, on two items, Q16d (use of whiteboards) and Q16f (group presenting work to others), no significant differences were seen between groups A and B since both groups reported using these practices much less frequently overall.

As reported above, no correlations between physics identity gains and classroom practices were found but there was one salient difference of interest in the comparison between groups A and B. Students were asked how interested they were in the content and topic. We found a small difference with an effect size of 0.23 (95% confidence interval: −0.035 to 0.50) between groups. Though small, it shows a somewhat higher level of interest among students who worked on small group activities every class, and interest is one of the important subconstructs associated with physics identity. Creating classrooms that generate student interest is important for enhancing student affect towards physics and improving the chances of students pursuing physics-related careers [22,24,39–41]. This is reinforced by other work that suggests interest in science classes is "contagious" [42].

V. DISCUSSION AND CONCLUSION

A. Importance of small group learning; connections with other work

Examining how these results connect to the broader literature, there are several possible reasons to be found that might help explain why our finding that frequent small group learning may make for more effective learning environments. When students work in small groups, complex social interactions take place. Students may want to do well in front of their peers (performance-approach motivation) and may also want to avoid looking bad (performance-avoidance motivation) [43]. This dynamic may make students take tasks more seriously than traditional classroom settings, as there are often only four to six people in a small group and there may be a greater sense of shared responsibility among students. In Fig. 2, we saw that students who reported working in small groups every class also tended to work at tables and face their peers. This puts students who may not normally participate in a traditional course *in situations* where the arrangement encourages cooperation. Working in small groups creates an environment rich for student agency [44,45]. Agency is one's belief of their capacity to act independently and make one's own choices and decisions. This is necessary for building self-efficacy which is an important factor in how students feel they will perform. Agency and self-efficacy are essential elements for student motivation as well as confidence building [46–48].

Working in groups also presents students with the opportunity to learn from each other allowing them to construct and reconstruct their knowledge in tandem with their peers. Group discussions have been found to help students build more confidence [49]. When a student falls behind, they may benefit by having extra, just-in-time help from their peers, whereas during a lecture, students may be primarily dependent on the instructor in order to learn, making it more difficult to catch up. In small group learning, students may be able to keep up with each other since they each have the opportunity to be actively involved in the learning process. Students may also build on each other's ideas while in a group, which makes for effective brainstorming and development of problem-solving skills [50]. When students actively cooperate with each other to achieve learning objectives, it leads to greater effort to achieve and to greater group effectiveness [51].

While working in small groups, students may form support networks with other students due to increased social integration, which then leads to them being resources for each other such as forming study groups outside of class. Research on social networks finds that students with broader networks perform better than their peers [52–54]. When students work collaboratively, they gain the opportunity to build networks within and outside of their classes. This may lead to students having more support overall such as forming study groups, exchanging contact information, sharing notes, all strategies found to lead to more student success [55].

Although the students in our sample were not asked to report if they were in a class that used particular, identifiable pedagogies (e.g., Peer Instruction, etc.), it is worth reflecting on how our results speak to classroom pedagogies that intentionally use small group learning activities daily. Modeling Instruction [8,56], Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) [17], and Peer Instruction [16] are three such highly research-driven pedagogies used in introductory physics that have adopted regular small group activities. Some of the strategies about group work appearing in Fig. 2 are implemented in various ways in these three designs.

In Modeling Instruction, for example, students may spend time every class working on tables, facing their group mates, using equipment, using whiteboards, collaborating with other groups and simultaneously using numbers, formulas, graphs, and words for solving physics problems. Students conduct various labs or experiments frequently (possibly not every class) and may use computer simulations. Modeling Instruction is the only one of the three example pedagogies discussed here that strongly emphasizes summarizing models on whiteboards to compare with other groups after every activity and these classrooms may contain little or no lecturing.

In a SCALE-UP classroom, students regularly work at tables, face their group mates, use equipment, and run computer simulations. There are frequent labs as well but not necessarily the use of whiteboards and groups may not collaborate with other groups on a daily basis. There may a small amount of traditional lecturing in class. Students may have regular opportunities to simultaneously use numbers, formulas, graphs, and words for solving physics problems.

Peer Instruction involves the most reliance on lecturing as compared to the other two examples above. About a third or half of a class is spent solving ConceptTests while the rest may be spent lecturing. During these ConceptTests students may face each other and present their work to each other but students are less likely to work at tables, use equipment, run computer simulations, or conduct labs during class.

In summary, then, MI and SCALE-UP may offer more frequent and varied opportunities to engage in the practices identified as important for conceptual gains in this study. Peer Instruction was designed to apply some student-centered instruction while keeping other features of a lecture course intact.

B. Literature support for other factors found to improve learning gains

From model 2 (see Table II) the use of equipment was associated with more conceptual learning. The regular use of equipment may allow students to be actively involved in modeling phenomena. This result goes hand in hand with model 5, which found that students who worked on labs or projects frequently reported higher learning gains. Labs at the post-secondary level have not been shown to be particularly important for students' conceptual understanding [57]. However, there is evidence for the benefits of the use of physics laboratory equipment in secondary education [58]. In model 4, the positive effects of frequently using computer simulations were seen. This is supported by other research that compares teacher-centered instruction to the use of computer-assisted instruction [59]. The Physics Education Technology Project (PhET) simulations have also been found to be effective as well [60,61]. Some reasons why this strategy may be helpful to students are that it allows them to see phenomena not otherwise visible, learn and repeat virtual experiments at their own pace, conduct experiments (perhaps those not physically possible) and gain instant feedback. Computer simulations may also be used to replace physical labs and equipment which may lead to more access and availability for students [62].

C. Physics identity gains: Need for further research

One of the results from this study was finding no significant associations between students' overall physics identity gain and the selected classroom practices that appeared in the CUPID survey. This was somewhat expected at the university level since physics identity has been found to be fairly stable by this time in a student's life [29]. The reason why so much importance is given to physics identity is because it has been shown to predict students' physics-related career choices [22]. In the current work, the average identity gain over a semester of physics was zero but there was variance with some students gaining while others were dropping. We hypothesized that certain classroom practices may significantly impact physics identity but we did not find evidence of this among the specific classroom practices measured in this study. Even though identity as a whole did not improve for the sample, there was a small increase in interest for students that worked in groups every class, as mentioned above. Interest is one of the subconstructs of physics identity so this was a somewhat promising result [22] and is consistent with other work [42].

D. Low use of physics education reform

Only a small fraction of our sample reported having the experiences that we identified as significantly associated to increased FMCE gain. Many of the respondents were taught in classes that lacked most or all of these instructional strategies. This is important to note, as it indicates a relative lack of the physics' community's knowledge on effective strategies and/or the lack of implementation of these strategies. This provides further evidence that effective practices are not widespread [63]. This underscores a barrier to improving educational outcomes including increasing STEM majors. Other researchers are working towards how to encourage more widespread adoption of evidence-based research [64].

E. Limitations

There are several limitations to this study. One limitation was the inherently discrete scales used to measure classroom practices so students who had experienced practices with frequencies between the ones provided would have to pick the closest choice which could lead to possible measurement error. Another limitation is that direct observations of these classrooms could not be practically performed to compare student responses to what an observer would identify in the classroom practices. Further, students were not asked to identify if they were part of a particularly identifiable pedagogical style (e.g., were they part of a SCALE-UP classroom), so these data do not directly assess the implementation or effectiveness of these approaches. These data only sampled 4-year colleges and universities so these results cannot necessarily be generalized to two-year colleges or high schools.

Another limitation was only being able to match 371 students' pre- and post-surveys. It was only possible to match 371 students because unique identification markers were relied upon and were not always filled out consistently for both surveys. This led to loss of statistical power. For this reason, as well as standard considerations of reproducibility, further work is necessary to increase confidence in the validity of this work. Lastly, quantitative data limits our ability to understand nuance about classroom practices, especially questions of implementation, framing, and outcomes. Thus, there is a need for future work that delves more deeply into the underlying mechanisms of these findings.

ACKNOWLEDGMENTS

We would like to thank the DBER group in the STEM Transformation Institute at FIU as well as the participants in this study for giving their time and experiences. Without them, this research would not have been possible. Also, thanks to the FIU Office of Research & Economic Development for supporting this study.

APPENDIX: COMPARISON OF DEMOGRAPHICS AND SCHOOLS WITHIN THE SAMPLE

In this Appendix, we provide further details of our data sample, which may also help future researchers to compile or incorporate this work into meta-analysis. Table III provides details of the racial and ethnic demographics of all the students in our study. In this table we also provide details of the groups A and B that were identified in the

TABLE III. Demographic comparison between demographics of sample, group A and B.

Race and ethnicity	% All Students ($N = 371$)	% Group A ($N = 41$)	% Group B ($N = 330$)
% Native American	3.50%	0.0%	3.90%
% Middle Eastern	3.20%	2.4%	3.3%
% Black	7.30%	7.3%	7.3%
% Hispanic, Latin	10.00%	4.9%	10.6%
% White	65.50%	58.5%	66.4%
% Asian	14.00%	24.4%	12.7%
% Native Hawaiian	0.50%	2.4%	0.3%
% Other	1.90%	2.4%	1.8%

TABLE IV. Demographic comparison between our sample and national numbers.

Race and ethnicity	Sample data	All students at four-year colleges	Historic PER research population
% Black	7.3%	12.2%	4.8%
% Hispanic, Latin	10.0%	18.5%	10.3%
% White	65.5%	54.3%	62.9%
% Asian	14.0%	6.4%	15.1%

TABLE V. Comparison of schools distributed within sample.

Four year school	Raw FMCE gain	Variance	School size	All students	Group A	Group B
A	0.50	0.50	Small	2	2	0
B	3.62	16.85	Small	34	10	24
C	−1.50	8.94	Medium	10	0	10
D	2.27	6.22	Small	11	0	11
E	0.15	6.56	Medium	20	2	18
F	3.18	11.16	Medium	11	0	11
G	5.41	14.25	Small	22	13	9
H	0.11	3.99	Small	18	1	17
I	1.00	7.33	Large	4	0	4
J	0.60	1.30	Small	5	0	5
K	0.19	7.52	Small	48	1	47
L	1.00	7.00	Small	3	0	3
M	3.33	8.33	Small	3	2	1
N	0.52	6.84	Small	25	0	25
O	−0.20	5.17	Large	15	4	11
P	−0.18	1.36	Small	11	2	9
Q	−0.06	4.73	Large	64	3	61
R	0.49	5.87	Medium	63	1	62
S	−.050	4.50	Small	2	0	2
Total	0.96	9.83		371	41	330

study. In our sample 50% of the participants identified as female. This is less than the percentage of women who attend four-year colleges in 2015 which was 55% [4].

Table IV compares our sample's demographics to the population of college students aged 18–24 at four-year colleges and universities, and the demographics of PER research populations, respectively. The latter comes from a recent analysis [25] while the former comes from the Digest of Education Statistics 2016 [4]. Note that Table III has more distinct factors than Table IV due to the fact that our survey data provided more (nonexclusive) options for students than the national sample data.

Table V shows more details about the school-wide FMCE gains, school size, and the distribution of participants at the (anonymized) four-year colleges and universities that were a part of our sample. The mean and variance of the pre-to-post raw FMCE gain, out of 17 questions. Note that the overall mean and variance in FMCE gain of our sample were 0.96 and 9.83, respectively. We also included the school sizes, where small is less than 5000 students, medium is between 5000 and 15 000 and large is greater than 15 000. Again, we provide breakdowns for subgroups A and B.

In Table IV, we can see broad similarities between our sample data and the general demographics of PER studies, though our sample has less black-identified students and more white-identified students than national college enrollment. In Table V it can be seen that there was a distribution of sample participants between schools and significant variance across school for both groups A and B.

[1] S. Olson and D. G. Riordan, *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineer, and Mathematics*, Science. Report to the President (Executive Office of the President, Washington, DC, 2012).

[2] J. Gainen, Barriers to success in quantitative gatekeeper courses, New Dir. Teach. Learn. **1995**, 5 (1995).

[3] T. Byun and G. Lee, Why students still can't solve physics problems after solving over 2000 problems, Am. J. Phys. **82**, 906 (2014).

[4] T. D. Snyder, C. de Brey, and S. A. Dillow, Digest of Education Statistics 2015, 51st ed. NCES 2016-014 National Center for Education Statistics (2016), Vol. **1**.

[5] S. Nicholson and P. J. Mulvey, *Roster of Astronomy Departments with Enrollment and Degree Data*, https://eric.ed.gov/?id=ED577003.

[6] V. Kistiakowsky, Women in physics and astronomy, Ann. N.Y. Acad. Sci. **323**, 35 (1979).

[7] L. Merner and J. Tyler, African American, Hispanic, and Native American women earning bachelor's degrees in physical sciences, Phys. Teach. **56**, 164 (2018).

[8] E. Brewe, L. Kramer, and G. O'Brien, Modeling theory applied: Modeling instruction in introductory physics, Am. J. Phys. **76**, 1155 (2008).

[9] M. A. Moriarty, Inclusive pedagogy: Teaching methodologies to reach diverse learners in science instruction, Equity Excell. Educ. **40**, 252 (2007).

[10] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[11] P. K. Cross and K. P. Cross, Teaching for learning, AAHE Bull. **39**, 6 (1987), https://files.eric.ed.gov/fulltext/ED283446.pdf.

[12] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).

[13] 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).

[14] P. Dillenbourg, Introduction: What do you mean by "collaborative learning"?, Collaborative Learning: Cognitive and Computational Approaches, https://telearn.archives-ouvertes.fr/hal-00190240.

[15] M. E. Stanne, L. Springer, S. S. Donovan, M. E. Stanne, and S. S. Donovan, Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis, Rev. Educ. Res. Spring **69**, 21 (1999).

[16] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[17] R. J. Beichner and J. M. Saul, Introduction to the SCALE-UP (student-centered activities for large enrollment undergraduate programs) project, Proceedings of the International School of Physics "Enrico Fermi" **1**, 1 (2003), https://files.eric.ed.gov/fulltext/ED459062.pdf.

[18] Z. Hazari, C. Cass, and C. Beattie, Obscuring power structures in the physics classroom: Linking teacher positioning, student engagement, and physics identity development, J. Res. Sci. Teach. **52**, 735 (2015).

[19] R. M. Lock, J. Castillo, Z. Hazari, and G. Potvin, Determining strategies that predict physics identity: Emphasizing recognition and interest, in *Proceedings of the 2015 Physics Education Research Conference* (AIP, New York, 2015), pp. 199–202.

[20] J. Wang and Z. Hazari, Promoting high school students' physics identity through explicit and implicit recognition, Phys. Rev. Phys. Educ. Res. **14**, 020111 (2018).

[21] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer Instruction: Engaging students one-on-one, all at once, Res. Reform Univ. Phys. **1**, 1 (2007), https://www.researchgate.net/file.PostFileLoader.html?id=580ad3bfcbd5c2737400cac1&assetKey=AS%3A419821171822593%401477104575391.

[22] Z. Hazari, G. Sonnert, P. M. Sadler, and M.-C. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47**, 978 (2010).

[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.17.020131 for pre and post conceptual understanding and physics identity development (CUPID) survey.

[24] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice, J. Eng. Educ. **105**, 312 (2016).

[25] S. Kanim and X. C. Cid, Demographics of physics education research, Phys. Rev. Phys. Educ. Res. **16**, 0020106 (2020).

[26] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton' s laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, Am. J. Phys. **66**, 338 (1998).

[27] S. Ramlo, Validity and reliability of the Force and Motion Conceptual Evaluation, Am. J. Phys. **76**, 882 (2008).

[28] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Comparing the Force and Motion Conceptual Evaluation and the Force Concept Inventory, Phys. Rev. ST Phys. Educ. Res. **5**, 010105 (2009).

[29] G. Potvin and Z. Hazari, The development and measurement of identity across the physical sciences, in *Proceedings of the 2013 Physics Education Research Conference* (AIP, New York, 2014), pp. 281–284.

[30] J. M. Nissen, R. M. Talbot, A. Nasim Thompson, and B. Van Dusen, Comparison of normalized gain and Cohen's d for analyzing gains on concept inventories, Phys. Rev. Phys. Educ. Res. **14**, 010115 (2018).

[31] J. Honaker, G. King, and M. Blackwell, Package Amelia. A Program for Missing Data.

[32] D. R. Johnson and R. Young, Toward best practices in analyzing datasets with missing data: Comparisons and recommendations, J. Marriage Family **73**, 926 (2011).

[33] P. Royston, Multiple imputation of missing values: Update of ice, Stata J. **5**, 527 (2005).

[34] A. C. Choirat, J. Honaker, K. Imai, G. King, O. Lau, C. Choirat, J. Honaker, K. Imai, G. King, and O. Lau, Zelig: Everyone's Statistical Software.

[35] M. Torchiano, *Efficient Effect Size Computation* (2016).

[36] R. Rosenthal, Parametric measures of effect size, Handb. Res. Synth. **621**, 231 (1994), https://psycnet.apa.org/record/1993-99100-015.

[37] D. H. J. M. Dolmans and H. G. Schmidt, What do we know about cognitive and motivational effects of small group tutorials in problem-based learning?, Adv. Health Sci. Educ. **11**, 321 (2006).

[38] L. K. Michaelsen, A. B. Knight, and L. D. Fink, Team-based learning: A transformative use of small groups in college teaching, Centers for Teaching and Technology - Book Library. 199 (2004), https://digitalcommons.georgiasouthern.edu/ct2-library/199.

[39] N. A. Fouad, P. L. Smith, and K. E. Zao, Across academic domains: Extensions of the social-cognitive career model, J. Counsel. Psychol. **49**, 164 (2002).

[40] K. A. Miller, G. Sonnert, and P. M. Sadler, The influence of student enrollment in pre-college engineering courses on their interest in engineering careers, J. Precoll. Eng. Educ. Res. **10**, 90 (2020).

[41] R. Lent, S. D. Brown, and G. Hackett, Toward a unifying scct and academic interest, choice and performance, J. Vocat. Behav. **45**, 79 (1994).

[42] Z. Hazari, G. Potvin, J. D. Cribbs, A. Godwin, T. D. Scott, and L. Klotz, Interest in STEM Is contagious for students in biology, chemistry, and physics classes, Sci. Adv. **3**, e1700046 (2017).

[43] K. Murayama, A. J. Elliot, and S. Yamagata, Separation of performance-approach and performance-avoidance achievement goals: A broader analysis, J. Educ. Psychol. **103**, 238 (2011).

[44] J. Jackson, L. Dukerich, and D. Hestenes, Modeling Instruction: An effective model for science education, Sci. Educ. **17**, 10 (2005), https://eric.ed.gov/?id=EJ851867.

[45] M. Klemenčič, From student engagement to student agency: Conceptual considerations of European policies on student-centered learning in higher education, High. Educ. Policy **30**, 69 (2017).

[46] C.-L. L. Dennis, Theoretical underpinnings of breastfeeding confidence: A self-efficacy framework, Journal of human lactation : official Journal of International Lactation Consultant Association **15**, 195 (1999).

[47] D. H. Schunk, Self-efficacy and academic motivation, Educ. Psychol. **26**, 207 (1991).

[48] B. L. Pollack and B. Lilly, Gaining confidence and competence through experiential assignments: An exploration of student self-efficacy and spectrum of inquiry, Mark. Educ. Rev. **18**, 55 (2008).

[49] B. J. Brooks and M. D. Koretsky, The influence of group discussion on students' responses and confidence during Peer Instruction, J. Chem. Educ. **88**, 1477 (2011).

[50] E. Harskamp and N. Ding, Structured collaboration versus individual learning in solving physics problems, Int. J. Sci. Educ. **28**, 1669 (2006).

[51] R. Wageman, Interdependence and group effectiveness, Administrative science quarterly **40**, 145 (1995).

[52] E. Williams, E. Brewe, J. P. Zwolak, and R. Dou, Understanding centrality: Investigating student outcomes within a classroom social network, in *Proceedings of the 2015 Physics Education Research Conference* (AIP, New York, 2015), pp. 375–378.

[53] E. Brewe, L. Kramer, and V. Sawtelle, Investigating student communities with network analysis of interactions in a physics learning center, Phys. Rev. ST Phys. Educ. Res. **8**, 010101 (2012).

[54] J. Bruun and E. Brewe, Talking and learning physics: Predicting future grades from network measures and Force Concept Inventory pretest scores, Phys. Rev. ST Phys. Educ. Res. **9**, 020109 (2013).

[55] A. Yazedjian, M. L. Toews, T. Sevin, and K. E. Purswell, It's a whole new world: A qualitative exploration of college students' definitions of and strategies for college success, J. Coll. Student Dev. **49**, 141 (2008).

[56] I. A. Halloun and D. Hestenes, Modeling Instruction in mechanics, Am. J. Phys. **55**, 455 (1987).

[57] C. E. Wieman and N. G. Holmes, Measuring the impact of introductory physics labs on learning and critical thinking, Am. J. Phys. **83**, 972 (2015).

[58] B. T. Olufunke, Effect of availability and utilization of physics laboratory equipment on students' academic achievement in senior secondary school physics, World J. Educ. **2**, 1 (2012).

[59] C. Bayrak, Effects of computer simulations programs on university students' achievements in physics, Turk. Online J. Dist. Educ. **9**, 53 (2008), https://dergipark.org.tr/en/pub/tojde/issue/16918/176534.

[60] C. J. Keller, N. D. Finkelstein, K. K. Perkins, and S. J. Pollock, Assessing the effectiveness of a computer simulation in introductory undergraduate environments, AIP Conf. Proc. **883**, 121 (2007).

[61] C. E. Wieman, W. K. Adams, and K. K. Perkins, Physics. PhET: Simulations that enhance learning, Science **322**, 682 (2008).

[62] N. D. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. Lemaster, When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment, Phys. Rev. ST Phys. Educ. Res. **1**, 010103 (2005).

[63] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, Phys. Rev. ST Phys. Educ. Res. **3**, 1 (2007).

[64] C. Henderson, A. Beach, and N. Finkelstein, Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature, J. Res. Sci. Teach. **48**, 952 (2011).