ACS central science

Research Article

http://pubs.acs.org/journal/acscii

Group Problem Solving in Class Improves Undergraduate Learning

Brent R. Stockwell,*[,†,‡] Melissa S. Stockwell,[§,∥,⊥] and Elise Jiang[†]

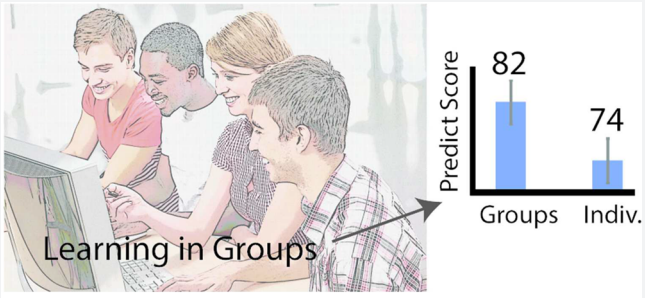
[†]Department of Biological Sciences, Columbia University, Northwest Corner Building, MC 4846, 550 West 120th Street, New York, New York 10027, United States

[‡]Department of Chemistry, Columbia University, New York, New York 10027, United States

[§]Department of Pediatrics and [∥]Department of Population and Family Health, Columbia University, New York, New York 10032, United States

[⊥]NewYork-Presbyterian Hospital, New York, New York 10032, United States

ABSTRACT: Developing methods for improving student learning is a long-standing goal in undergraduate science education. However, the extent to which students working on problems in small groups versus individually results in improved learning among undergraduate science students has not been evaluated in a randomized controlled trial. We have performed such a trial with 80 students in an undergraduate biochemistry class, in which students were randomized to either learning in groups or learning individually. All students participated in the same class, which consisted of a lecture with periodic breaks for students to solve problems using an audience response system. Students in the individual learning condition answered these questions on their own, but students in the group-based learning condition answered these questions in an assigned group of four students. At the end of the class, all students then took the same exam as individuals. The exam had two types of questions—*recall* questions, in which students had to simply recall information provided to them, and *predict* questions, in which students had to apply their new knowledge to a new context. Students in the individual and group-based learning conditions performed similarly well on recall questions. However, students who had been in the group-based learning condition performed significantly better as individuals on the *predict* questions. This suggests that learning in groups may be more effective than individual learning for undergraduate science students, particularly for applying their knowledge to new contexts; this highlights the potential need for pedagogical approaches in undergraduate science courses that incorporate learning in groups.



Adapted from image by Konstantin Chagin/Shutterstock

INTRODUCTION

Science students typically learn individually. In the most common pedagogical approach, students prepare for class on their own, for example, by reading a section of a textbook followed by listening to an instructor lecture on the assigned material during class. This traditional approach to learning is not effective for many students.[1] In recent years, several new approaches to science education have emerged, including online learning.[2,3] Moreover, active learning and scientific teaching strategies have been explored and found to improve student performance.[4−6] For example, we and others have reported that having students individually solve problems during class promotes better learning.[5,7] Similarly, interactive teaching is associated with improved student learning in physics classes.[8] Thus, a number of studies suggest that interactive, problem-based learning is an effective pedagogical strategy.[9]

Despite the importance of these previous studies, they did not evaluate sufficiently the impact of individual versus group-based learning. There are a number of reasons that have been proposed to support the hypothesis that when students work in small groups they learn more effectively.[10] First, in such groups, students have the opportunity to explain to each other gaps in background knowledge necessary to understand and apply class material; indeed, it has been postulated that the benefit of learning in groups is due to both better engagement with the course material and with other learners.[11,12] Second, each student may retain and process different aspects of the class material and the preparation material, and they can share these different perspectives in a small group environment. Third, it has been proposed that novices may be more effective at teaching novices, as experts can be far removed from the initial challenges in learning information-dense material. Fourth, working in groups may encourage students to persist in solving difficult problems beyond the point when they might give up as individuals. Finally, it has been proposed that learning in groups is based in the constructivist theory of learning, in that errors and inconsistencies in learners' current knowledge and schema are identified and corrected through peer discussions.[11,13]

Received: March 27, 2017

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

On the other hand, there are also reasons to think that some students might not learn as well in a group environment compared how they learn as individuals.[14] First, some students may prefer to learn individually and might feel uncomfortable in the learning environment in groups. Second, lower performing students might not be given the opportunity to learn when in a group environment, with higher performing students taking charge of the work for the group. Finally, teams may be dysfunctional, and plagued by fighting and disagreement.[15]

Thus, there are reasons to think that group-based learning might be either more effective or less effective for students compared to learning as individuals. A number of studies have examined empirically the effect of group-based learning in a variety of settings.[16−19] One specific format of learning in groups, referred to as "team-based learning", has been extensively studied and practiced in a number of educational settings.[20−22] However, few studies have employed randomization to directly evaluate the impact of group work in improving learning. For example, one study found that in an undergraduate microbial physiology course, final exam scores improved in the years in which team projects were assigned;[23] however, this study design did not rule out a variety of other explanations for the year-to-year differences in final exam scores. Other studies evaluated the change in knowledge scores before and after employing team-based learning,[23−25] but the impact of team-based learning versus other course factors was not examined, potentially obscuring the impact of this educational method. Other studies have given students the choice of selecting a team-based learning section of a course during medical school in Austria,[26] or compared different courses with and without team-based learning;[27] however, while students in the team-based learning course had a trend toward higher final exam scores compared to students in a nonteam-based-learning course, student motivation and ability could not be disentangled from the team-based learning approach, since no randomization was used. Another study allowed students to volunteer for a small group learning environment and offered additional points toward their course grade as an incentive, which would result in a nonrandom distribution of students in the different learning environments, and different grade distributions.[28] Only one study of team-based learning used randomization of students: Thomas and Bowen compared the highly structured "team-based learning" approach with another small group learning approach in an ambulatory medicine course.[29] They assigned each of 112 students to one of these two learning preferences and then switched the learning approaches for the second half of the study; they found that exam scores were higher in the team-based learning section, but it was not clear if the same exam was used in both courses,[16] potentially raising doubt about the impact of team-based learning. This course was also in the medical setting rather than in an undergraduate science course.

A recent review of active learning methodologies discussed the evidence supporting another type of small group learning, called "peer-led team learning",[30,31] and suggested that this approach can lead to gains in student performance, retention, and attitudes, as well as higher order thinking.[32] However, while supportive of the notion that group learning improves a variety of student outcomes, these cited studies have many of the same limitations noted above. For example, two studies in general chemistry courses, as well as another study in an organic chemistry course, examined the impact of peer-led guided inquiry, but did not use randomization, raising the possibility that other variables explain the results.[33−35] Another study compared students in a traditional organic chemistry course from 1992 to 1994 with those who were in a peer-led team learning environment from 1996 to 1999; while students in the latter courses earned more exam points, the many differences between these groups, including the changing admissions criteria over time noted by the authors, complicates any clear-cut conclusions about the impact of team learning.[36,37]

Indeed, a review on the evidence supporting team-based learning[16] noted, *"Establishing the efficacy of TBL (Team-Based Learning) is limited in all of the studies because of the lack of true experimental studies."* The author further states, *"Most of the studies about TBL are descriptive, rather than experimental. When comparisons are used, random assignment is rare. Planning true experiments is difficult in education. However, prospective studies in which students are assigned to TBL and non-TBL sections are needed to demonstrate efficacy of TBL."* A more recent review of the 67 published peer-reviewed studies evaluating the effectiveness of peer-led team learning[38] did not identify any studies that used a randomized controlled trial to objectively assess the impact of team-based approaches on student learning.

Thus, even for the specific approach termed "team-based learning", which has been extensively studied and implemented, randomized controlled trials to evaluate efficacy are lacking. While studies reported to date provide useful information about team learning in a variety of subjects, they cannot definitively determine the impact of team-based learning compared to individual learning among the same pool of students, with the same educational material, exam, and instructor, isolating the variable of interest—group versus individual learning; thus, randomized controlled trials would add to the literature on the effects of learning in groups. We sought to test experimentally the impact of learning in small groups versus individual learning in an undergraduate biochemistry course using such a randomized controlled trial. We did not examine the structured learning format known as "team-based learning", but instead examined the more general question of how well students learn when they answer questions in small groups of four students, versus answering these questions as individuals.

A RANDOMIZED CONTROLLED TRIAL TO EVALUATE LEARNING IN GROUPS

The 166 students enrolled in Biochemistry I: Structure and Metabolism, an undergraduate biochemistry course at Columbia University, were invited to participate in a randomized controlled trial in the fall of 2015 (Figure 1). A total of 80 students enrolled in the study. Students were randomized 1:1 to one of two arms—learning in groups (n = 40) or individual learning (n = 40). The students were stratified for randomization by prior exam performance (lower quarter vs upper three-quarters), to ensure equal representation of these differently performing students in each study arm. In a previous study,[5] we stratified biochemistry students at Columbia in a blended learning randomized controlled trial based on gender and prior exam performance. We found that gender had no effect on performance in the class or in the study, whereas prior exam performance was correlated with study performance. Therefore, in the current study presented in this manuscript, we did not stratify based on gender, only on prior exam performance.

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

A

	Team learning	Individual learning	P value	Effect size
Number of students	40	40		
Median score, predict questions	82 +/- 3%	74+/-3%	0.036	0.48
Median score, recall questions	83+/- 2%	87+/- 2%	ns	0.24
Satisfaction	4.2 +/- 0.2	4.0 +/- 0.1	ns	0.17

B

	Students who prefer teams	Students who prefer indiv.	P value	Effect size
Median score	82 +/- 2%	82 +/- 3%	ns	0.01

C

	Students with high grade	Students with low grade	P value	Effect size
Median score (predict)	81 +/- 2%	68 +/- 5%	0.004	0.57

Figure 1. Results of randomized controlled trial of team-based learning versus individual learning. (A) 80 students were recruited into the study, and randomized to learning in a small group or individually. The mean scores (± SEM) on *predict* and *recall* questions are shown for the two groups, along with the significance (Student's *t* test) and effect size (calculated as difference in means between groups divided by pooled standard deviation, using Cohen's *d*). In addition, students were asked how satisfied they were with the learning environment, and there was no significant difference between the study groups. (B) Students who indicated that they prefer to learn in teams had no significant difference in exam scores compared to students who indicated that they prefer to learn as individuals (*P* value calculated using Student's *t* test and effect size calculated using Hedges' *g*, due to the different group sizes). (C) The median score on predict questions of students with a high prior course grade was significantly higher than for students with a low prior course grade, indicating that performance on these questions was correlated with prior performance in the course (*P* value calculated using Student's *t* test, and effect size calculated using Cohen's *d*).

In addition, we recognized that there may be unknown covariates of study performance that were not equally distributed between the study groups, despite the randomization. To test explicitly for two such possibilities, we compared the distribution of learning preferences between the two groups. We found that the team-learning group had 8 students who stated that they prefer learning on their own versus 32 students who stated that they prefer learning in a group; in the individual learning group, the corresponding numbers were 10 and 30 (p = 0.59). Thus, stated learning preference, although not explicitly considered in the randomization process, was not significantly different between the groups.

Similarly, we analyzed the distribution of male and female students in the two study groups, even though gender was not explicitly considered in the randomization process. We found that the team learning group had 11 men and 29 women, whereas the individual learning group had 13 men and 27 women (p = 0.63). Thus, gender differences are not likely to explain performance differences between the two groups.

Despite these analyses, caution is needed in assessing the impact of randomized controlled trials, as useful as they can be. We cannot control for all possible covariates with study performance, and it is possible, although statistically unlikely, that some unknown covariate varied substantially between the study groups.

One week prior to class, all enrolled students were sent a link to a video providing an introduction to the material to be discussed in class, based on our prior finding that video preparation increases student satisfaction and attendance, compared to textbook preparation.[5] In class, students were then directed to one particular side of the classroom, as per their randomization assignment; students on one side of the classroom were instructed to solve problems and answer questions as individuals, whereas students on the other side of classroom were instructed to work in assigned groups of four students. During class, the instructor presented a lecture, stopping periodically to ask students to solve a problem or answer a question, and to submit their answers using an audience response system (Learning Catalytics, Pearson). During the audience response portion of the class (i.e., while students were answering questions), the instructor scanned the room to ensure that the students in the group learning environment worked in groups and that the students in the individual learning group worked as individuals. In addition, four teaching assistants and the study coordinator walked around the room, monitoring the students, to ensure that each student worked in the manner in which they had been instructed to do so. There were no reports of students working in a manner other than as they had been instructed to.

At the end of the class, all students took a 15-question, 20 min exam as individuals to test their understanding of the class material. There were two types of questions purposefully included on the exam—8 recall questions that simply asked the students to recall information presented to them in class, and 7 predict questions that asked students to apply the knowledge they gained in class to a new context that had not been explicitly described to them during class. The recall questions were defined as recall questions if the answer to the question was provided during the class prior to asking the question—these questions simply tested the ability of students to remember the information that was provided during class. The predict questions were defined on the basis that the answer to the question was not explicitly provided during class—for these questions, students had to deduce the answers by applying the knowledge they obtained during class.

STUDENTS WHO LEARNED IN TEAMS PERFORMED BETTER AS INDIVIDUALS ON *PREDICT* QUESTIONS

We examined the performance of students in the two study arms on the two types of exam questions to determine the effects of individual vs team learning on student performance. We found that students who worked in groups during class had a higher subsequent score as individuals on *predict* questions than students who worked individually during class (82 ± 3% vs 74 ± 3%, p = 0.036, Figure 1). In contrast, there was no significant difference in the performance of students working individually versus in teams on simple *recall* questions (83 ± 2% vs 87 ± 2%, p = 0.29, Figure 1), demonstrating that working in teams did not affect recall of class material. This suggests that the benefit of working in groups is not related to recalling aspects of the class material that individual students may have missed.

STUDENTS LEARNING PREFERENCES DO NOT ALTER THE BENEFIT OF LEARNING IN TEAMS

We recognized that students have varied preferences regarding their learning environment, and that some students may prefer to learn as individuals, whereas others may prefer to learn in groups. To determine the preferences of each student in the study, we asked all of the students in the study whether they preferred to learn as individuals or in teams. We found that 62/80 students indicated that they prefer to learn in teams, whereas 18 students indicated that they prefer to learn as individuals. We then analyzed whether student learning preference (team vs

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

individual) correlated with performance on the study exam. We found that students with these different preferences had identical overall performance (82%, Figure 1) on the study exam. Consistent with this observation, these different learning preferences (team vs individual) did not impact the relationship seen between learning condition and performance on the *recall* or *predict* type of questions. That is, the 32 students who indicated that they prefer to learn in teams and who were in the team learning condition indeed performed better than the 30 students who indicated that they prefer to learn in teams, but who were randomized to the individual learning conditions, on *predict* questions (82 ± 3% vs 74 ± 3%, $p = 0.03$). Additionally, there was no significant difference in the performance of these two groups on *recall* questions (84 ± 2% vs 86 ± 3%, ns). Interestingly, even the 18 students who indicated that they prefer to learn as individuals had a trend toward a higher score on the *predict* questions (84 ± 7% vs 75 ± 6%, $p = 0.17$), but not on the recall questions (81 ± 7% vs 88 ± 2%, ns) only for those randomized to the group learning condition.

We also evaluated the satisfaction of students with their learning environment, asking them to rate their satisfaction on a scale of 1 to 5, with 5 being fully satisfied and 1 being unsatisfied. There was no difference in the satisfaction of students in the team learning and individual learning environments (4.2 ± 0.2 vs 4.0 ± 0.1, ns). Moreover, we found no significant difference in the satisfaction of students who reported that they prefer to learn in teams or as individuals (4.1 ± 0.1 vs 4.2 ± 0.2, ns). When we analyzed the satisfaction of students based on learning preference (team vs individual), we also found no significant difference in satisfaction in the team and individual learning environment; that is, students who indicated that they prefer learning in teams reported themselves to be equally satisfied in the team learning environment and the individual learning environment (4.1 ± 0.2 vs 4.1 ± 0.1). Similarly, students who indicated that they prefer to learn as individuals actually had a trend toward higher satisfaction in the team-learning environment, although it did not reach statistical significance due to the small number of students in this category (4.4 ± 0.3 vs 4.0 ± 0.2, $p = 0.28$). Thus, students' reported learning preference did not predict either their actual performance or satisfaction with their learning environment. This is something to be aware of when determining whether to use a team-learning environment for students—their stated preferences may not predict either their satisfaction or performance.

We then analyzed the performance of students (15/80) who rated their satisfaction as being only moderate or low (≤3); we wondered if this group of students would show less benefit of working in teams. On the contrary, we found indications of the same benefit of group-based work in this group of students. Among this group of students, those who worked in teams had a trend toward better performance than those who worked as individuals (81 ± 6% vs 69 ± 10%, $p = 0.31$) on *predict* questions, but not on *recall* questions (85 ± 5% vs 88 ± 5%). Therefore, even less satisfied students may perform better in teams on *predict* questions.

To test whether the study measured the same performance characteristics found in the rest of the course, we compared the study exam performance for students with a higher prior course exam performance (top 76% of students) versus a lower one (bottom 24%). Indeed, we found that prior course performance (high vs low) correlated with the performance on the *predict* questions in the study exam: the median score for students in the high prior exam group (81%) was significantly ($p = 0.004$) greater than the median score for students in the low prior exam group (68%) on these questions, suggesting that the students exerted their typical effort and exhibited similar performance characteristics, despite the fact that participation was voluntary (Figure 1). Interestingly, performance on the *recall* questions in the study was not correlated with prior exam performance, as both students with high and low prior exam scores had a median score of 85% on these recall questions in the study. This could indicate that predict questions on exams throughout the course are primarily responsible for the gap in test scores between the high and low performers. Therefore, group-based learning, which has been found here to increase scores of predict questions specifically, could play an even more significant role in the classroom.

We should note that while we quantitatively assessed the satisfaction of the students, we did not do a qualitative assessment through interviews to explore the students' experiences working in teams. We believe that in this study, the quantitative assessment stands alone, but a qualitative assessment would be a valuable avenue to explore in the future, similar to other efforts that have been performed.[39,40]

We also examined the validity of the study exam instrument. First, as noted above, we compared the performance of students on the study exam to their prior performance in the course and found a statistically significant correlation for predict questions (Figure 1): students who had a high prior course grade (top 3/4 of class) had a median score on the study exam of 81 ± 2%, while students who had a low prior course grade (bottom 1/4 of class) had a median score on the study exam of 68 ± 5%. The predict questions were correlated with prior course performance and suggest that they measured the attributes needed for success in this biochemistry course. The recall questions were not correlated with prior exam performance ($p = 0.67$). This is likely because this course requires students to make predictions and apply their knowledge on exams, rather than simply recalling information provided during class; thus, we would not expect recall performance to correlate with course performance for this course, although the results could be different for other courses in which grades are based on memorization and recall.

To analyze further the reliability of the study exam in terms of predict vs recall questions, we analyzed the performance of the two study groups on each question, as shown in Figure 2. We found that the students randomized to group learning performed better on every predict question, but on only one recall question. This suggests that the questions were properly categorized and consistently reported on the effect of the group vs individual learning environment in the study.

We also examined whether the two study groups had a similar capacity to learn the study topic. Prior exam performance in the course is the best indicator we have of the knowledge and performance capability of the students in two groups. For this reason, as noted above, we stratified the students in the randomization process based on prior exam performance in the course, which explicitly controlled for their varying levels of knowledge and performance related to undergraduate biochemistry.

However, we recognized that students could have different background knowledge related to the specific biochemistry topic covered during the study (amino acid metabolism), but not in the rest of the course. To test for this, we analyzed the performance of students prior to the study on the same

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

Question	Type	Group	Indiv.	P value	Discrim. Index	Point Biserial Correl.
1	Recall	0.85	0.95	0.96	0.14	0.09
4	Recall	0.93	0.95	0.72	0.23	0.27
5	Recall	0.90	0.98	0.94	0.14	0.14
7	Recall	0.93	0.98	0.88	0.09	0.22
8	Recall	0.60	0.65	0.87	0.09	-0.05
11	Recall	0.73	0.73	0.50	0.59	0.36
14	Recall	0.93	0.93	0.50	0.23	0.34
15	Recall	**0.83**	0.80	0.34	0.46	0.35
2	Predict	**0.75**	0.68	0.14	0.59	0.37
3	Predict	**1.00**	0.98	0.00	0.00	-0.07
6	Predict	**0.80**	0.75	0.22	0.46	0.33
9	Predict	**0.98**	0.93	0.02	0.14	0.11
10	Predict	**0.80**	0.58	2e-4	0.77	0.43
12	Predict	**0.45**	0.33	0.06	0.50	0.16
13	Predict	**0.98**	0.94	2e-4	0.09	0.22

Figure 2. Performance of study groups in each exam question. The recall questions are shaded white and the predict questions are shaded gray, while the questions on which the students in the group learning environment performed better are highlighted in bold. While the study was not powered to detect differences on every question, the results show the consistent nature of the study questions and student performance. The fraction of students in each group who provided the correct answer is listed for each question. The discrimination index is shown for each question and was calculated by subtracting the number of correct responses provided by the lowest scoring 22 students (~27% of 80 students) from the number of correct responses provided by the highest scoring 22 students on the exam, divided by 22. The point-biserial correlation is shown for each question. *P* values were calculated using the *Z* test, comparing the fraction of students who provided the correct answer in group learning vs individual learning.

biochemistry topic, amino acid metabolism, in an earlier part in the course. In this earlier part of the course, we had administered three tests to the students related to amino acid metabolism, which serve as a prestudy performance measure. First, we assigned students a video to watch on amino acid metabolism, and then had them answer questions at home individually based on the video content, to test the extent to which they mastered the video content prior to attending class. We found that the students who would later be randomized to the group learning condition in our study had an identical performance to the students who would later be randomized to the individual learning condition (both groups had a mean 95% on this quiz, $p = 1.0$); this suggests that these two groups, after randomization, had similar knowledge and capacity to learn amino acid metabolism from this video.

Second, we analyzed the ability of these students to work in teams. During the prior part of the course on amino acid metabolism, we had the students work in teams and answer questions about amino acid metabolism as a team. We found that in this environment, the students who would later be randomized to the team learning condition in the study had a mean score of 91%, whereas the students who would later be randomized to the individual learning condition had a mean score of 96%, but this difference was not statistically significant ($p = 0.10$). This suggests that the two groups from the study had a similar capacity to work in teams (if anything, those who would be in the individual learning group actually performed slightly higher in teams, although this was not statistically significant).

Third, in this prior portion of the course, we asked the students questions as individuals during class about amino acid metabolism to assess how well they were mastering the content during class. We found that the group that would later be randomized to the team learning condition had a mean score of 97.6% in class and the group that would later be randomized to the individual learning condition had a mean score of 97.3%; these were not significantly different ($p = 0.79$).

We also examined the reliability and validity of the exam instrument used in the study. Reliability refers to the extent to which an assessment will produce a similar result upon repeated use. Unfortunately, we have no means of establishing the reliability of the study instrument upon repeated use by the same students, because it was not possible to have the same students take the study exam a second time. We note that this could be an interesting aspect of future study designs. Validity refers to the extent to which as assessment measures the intended characteristics of those who take the assessment. Kibble[11] has noted five domains for establishing validity, involving exam content, response process, internal structure, relationship to other variables, and consequences of testing.

The study exam content related to amino acid metabolism, the subject of the lecture and preparatory video provided to the students. One advantage of performing this randomized controlled trial on a single topic is that it is less likely that there would be mismatch between distribution of exam content and the content of the class material, since the topics are closely aligned. Indeed, due to the focused nature of the content, we were able to ensure that every "recall" question asked information that was explicitly described in the lecture, and that every "predict" question related to the content of the lecture, but was not explicitly described during the lecture.

To avoid the possibility of human error or subjectivity in grading the student exams, we used multiple-choice questions that were computer graded through the Columbia Learning Management System. This increased the likelihood that the scores were a valid reflection of the answers the students intended to provide.

To assess whether the different questions individually measured the same properties and were therefore consistent, we examined the performance of the two study groups on each question (Figure 2). We calculated the difficulty of each question for each group (fraction of correct responses, Figure 2). The students who worked in groups performed better on predict questions, supporting the consistency of the questions. We calculated Cronbach's α, as a measure of internal consistency of the study exam questions, and found it to be 0.60. While this is on the low end for measuring a single construct, the study exam measured understanding of several different aspects of amino acid metabolism and so would not necessarily have a high correlation among responses to all test questions; nonetheless, future replication studies would be of value, focusing on measuring a single construct with high α.

We also calculated the discrimination index for each question (the difference between the top 22 (27% of 80) scoring students and bottom 22 (27% of 80) scoring students to each question, divided by 22) to determine which questions differentiated the top scoring students from the lowest performing students on this exam. We found a range of discrimination indices for both recall and predict questions (Figure 2). We also calculated the point-biserial correlation for each question, to determine the extent to which performance on each question correlated with each student's overall exam

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

performance. Again, we found a range of correlations for both the recall and predict questions (Figure 2). Together, these data suggest that the study exam had a reasonable degree of internal validity, although improvements are certainly possible in the future.

In terms of correlating with other variables, as noted above, there was a significant correlation between prior course performance and performance on the study exam, indicating that the study exam was measuring similar characteristics of the students as the prior course exams. Since our goal was to evaluate the effect of learning in a group versus individually, but not the ability of the exam to evaluate future trajectories of the students, we did not evaluate other consequences of the study or the exam. However, such analyses could be designed into future studies to examine the long-term impact of learning in groups.

Thus, this analysis suggests that the students in the two groups of this study had similar knowledge, capacity to learn the topic of the study, and ability to work in teams on this topic, and that the study instrument had a degree of internal consistency and validity. As a result, we can have reasonable confidence in the conclusion that the students assigned to the group learning condition performed better on *predict* questions.

DISCUSSION

We found that learning in groups significantly improved the subsequent performance of students as individuals, irrespective of the satisfaction or stated learning preference of the students. There are a variety of reasons to suspect that students might learn more or less effectively in a group environment compared to how they learn as individuals. We approached this question empirically, by executing a randomized controlled trial to compare how well students learned a new topic in biochemistry in small groups versus as individuals. All other aspects of their learning environment were identical, so that we could isolate the impact of this variable on learning. A limitation of this study is that it was only performed on a particular set of students on a single topic in biochemistry, and that future studies will need to examine whether the findings are transferable to other topics. However, we also note that there is nothing intrinsic to the topic of biochemistry or amino acid metabolism in particular that would lend itself to group-based learning.

We suggest that it would be helpful to look in the future at other student characteristics and to know whether team learning is particularly effective for underrepresented groups in science. This was not feasible for this study for several reasons. First, we did not have ready access to other characteristics of the students (e.g., GPA, race, major, or first generation status). Second, with the relatively modest number of students available to study in the course, it is likely not feasible to subdivide the students by multiple characteristics, as the power would be reduced beyond the point that significant differences could be identified—to do that, we would need a significantly larger study, which would either need to be performed in a very large introductory science course with a high participation rate in the study, or by combining students from multiple different courses. We believe that this would be valuable for a future larger study.

We found that students' reported learning preference did not predict either their actual performance or satisfaction with their learning environment. We emphasize that this is something to consider when assessing the role of a group-learning environment for students—as their preferences may not predict either their satisfaction or performance. In general, it is important to remember that satisfaction and preferences are valuable to assess, but do not equate with learning. In this study, we compared a lower level cognitive skill (recall questions) with a higher order cognitive skill (predict questions). It would be valuable to expand upon this comparison for higher and lower cognitive activities more generally in the future.

Overall, we found that students who worked in small groups during class ultimately performed better as individuals, compared to students who worked individually to answer queries in class. Of note, even large introductory science courses can make use of group learning environments—an undergraduate analytical chemistry course instructor reported that group-based problem solving was possible in a large course.[41] Moreover, a number of studies have shown increased student satisfaction with group learning compared to traditional lecture-based learning.[42,43]

Interestingly, the benefit of group learning on exam performance in our study only applied to questions that asked students to make new predictions based on their newly acquired knowledge—students working in small groups did not perform better at simple recall of material presented in class. This suggests that the benefit of working in small groups is not due to enabling students to remind each other of aspects of the class that they might individually have missed hearing, due to a lack of attention or focus. Instead, the benefit of learning in small groups appears to enable students to understand the class material at a deeper level, so they can apply it to new contexts. We suspect that this effect of learning in small groups may be due to (i) persistence to work on problems in a team caused by the social stigma associated with giving up prematurely and the social benefit of seeing a problem through to completion with peers, (ii) the ability of students to provide background context to each other that each individual student may be missing to fully grasp the presented material in class, or (iii) providing students an opportunity to discuss the topics in an interactive forum that offers a further outlet for students who learn best by speaking and reciting information, rather than just listening to a presentation. Alternatively, the benefit of learning in small groups may be due to a combination of these factors, as well as additional factors that we have not considered. Future studies may shed increased light on the mechanisms governing the benefit of learning in small groups, as well as how general this effect may be in other student populations. In summary, these results suggest that instructors should consider adopting and evaluating the impact of group work on student learning in diverse contexts.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bstockwell@columbia.edu.

ORCID

Brent R. Stockwell: 0000-0002-3532-3868

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support for this study was provided by Columbia University's Office of the Provost.

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX

REFERENCES

(1) Handelsman, J. *Scientific Teaching*; W.H. Freeman and Co.: New York, 2007.

(2) Reich, J. Education research. Rebooting MOOC research. *Science* **2015**, *347* (6217), 34−35.

(3) Glazer, F. S. *Blended Learning: Across the Disciplines, Across the Academy*; Stylus: Sterling, VA, 2012.

(4) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (23), 8410−8415.

(5) Stockwell, B. R.; Stockwell, M. S.; Cennamo, M.; Jiang, E. Blended Learning Improves Science Education. *Cell* **2015**, *162* (5), 933−936.

(6) Couch, B. A.; Brown, T. L.; Schelpat, T. J.; Graham, M. J.; Knight, J. K. Scientific Teaching: Defining a Taxonomy of Observable Practices. *Cbe-Life Sci. Educ.* **2015**, *14* (Spring), 1−12.

(7) Haak, D. C.; HilleRisLambers, J.; Pitre, E.; Freeman, S. Increased structure and active learning reduce the achievement gap in introductory biology. *Science* **2011**, *332* (6034), 1213−1216.

(8) Hake, R. R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* **1998**, *66*, 64−74.

(9) Knight, J. K.; Wood, W. B. Teaching more by lecturing less. *Cell Biology Education* **2005**, *4*, 298−310.

(10) National Research Council. *How People Learn Brain, Mind, Experience, and School*, expanded ed.; National Academy Press: Washington, D.C., 2000.

(11) Kibble, J. D.; Bellew, C.; Asmar, A.; Barkley, L. Team-based learning in large enrollment classes. *Adv. Physiol Educ* **2016**, *40* (4), 435−442.

(12) Haidet, P.; Levine, R. E.; Parmelee, D. X.; Crow, S.; Kennedy, F.; Kelly, P. A.; Perkowski, L.; Michaelsen, L.; Richards, B. F. Perspective: Guidelines for reporting team-based learning activities in the medical and health sciences education literature. *Acad. Med.* **2012**, *87* (3), 292−299.

(13) Hrynchak, P.; Batty, H. The educational theory basis of team-based learning. *Med. Teach* **2012**, *34* (10), 796−801.

(14) Andersen, E. A.; Strumpel, C.; Fensom, I.; Andrews, W. Implementing team based learning in large classes: nurse educators' experiences. *Int. J. Nurs. Educ. Scholarsh.* **2011**, *8*.10.2202/1548-923X.2197

(15) Farland, M. Z.; Sicat, B. L.; Franks, A. S.; Pater, K. S.; Medina, M. S.; Persky, A. M. Best practices for implementing team-based learning in pharmacy education. *Am. J. Pharm. Educ.* **2013**, *77* (8), 177.

(16) Sisk, R. J. Team-based learning: systematic research review. *J. Nurs Educ* **2011**, *50* (12), 665−669.

(17) Hills, H. *Team-Based Learning*; Gower: Aldershot, Hampshire, England, 2001.

(18) Sibley, J. *Getting Started with Team-Based Learning*, 1st ed.; Stylus Publishing: Sterling, VA, 2014.

(19) Haidet, P.; Kubitz, K.; McCormack, W. T. Analysis of the Team-Based Learning Literature: TBL Comes of Age. *J. Excell. Coll. Teach.* **2014**, *25* (3−4), 303−333.

(20) Sweet, M.; Michaelsen, L. K. *Team-Based Learning in the Social Sciences and Humanities: Group Work That Works To Generate Critical Thinking and Engagement*, 1st ed.; Stylus Pub.: Sterling, VA, 2012.

(21) Michaelsen, L. K. *Team-Based Learning for Health Professions Education: a Guide to Using Small Groups for Improving Learning*, 1st ed.; Stylus: Sterling, VA, 2008.

(22) Michaelsen, L. K.; Knight, A. B.; Fink, L. D. *Team-Based Learning: a Transformative Use of Small Groups*; Praeger: Westport, CT, 2002.

(23) McInerney, M. J.; Fink, L. D. Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course. *Microbiol. Educ.* **2003**, *4*, 3−12.

(24) Haberyan, A. Team-based learning in an Industrial/Organizational Psychology course. *North Am. J. Psychol.* **2007**, *9* (1), 143−152.

(25) Kuhne-Eversmann, L.; Eversmann, T.; Fischer, M. R. Team- and case-based learning to activate participants and enhance knowledge: an evaluation of seminars in Germany. *J. Contin Educ Health Prof* **2008**, *28* (3), 165−171.

(26) Wiener, H.; Plass, H.; Marz, R. Team-based learning in intensive course format for first-year medical students. *Croat. Med. J.* **2009**, *50* (1), 69−76.

(27) Koles, P.; Nelson, S.; Stolfi, A.; Parmelee, D.; Destephen, D. Active learning in a Year 2 pathology curriculum. *Med. Educ* **2005**, *39* (10), 1045−1055.

(28) Lyon, D. C.; Lagowski, J. J. Effectiveness of Facilitating Small-Group Learning in Large Lecture Classes. *J. Chem. Educ.* **2008**, *85* (11), 1571−1576.

(29) Thomas, P. A.; Bowen, C. W. A controlled trial of team-based learning in an ambulatory medicine clerkship for medical students. *Teach Learn Med.* **2011**, *23* (1), 31−36.

(30) Gosser, D. K.; Roth, V. The Workshop Chemistry project: Peer-led team learning. *J. Chem. Educ.* **1998**, *75* (2), 185−187.

(31) Lewis, S. E. Retention and Reform: An Evaluation of Peer-Led Team Learning. *J. Chem. Educ.* **2011**, *88* (6), 703−707.

(32) Eberlein, T.; Kampmeier, J.; Minderhout, V.; Moog, R. S.; Platt, T.; Varma-Nelson, P.; White, H. B. Pedagogies of engagement in science: A comparison of PBL, POGIL, and PLTL. *Biochem. Mol. Biol. Educ.* **2008**, *36* (4), 262−273.

(33) Lewis, S. E.; Lewis, J. E. Departing from Lectures: An Evaluation of a Peer-Led Guided Inquiry Alternative. *J. Chem. Educ.* **2005**, *82* (1), 135−139.

(34) McCreary, C. L.; Golde, M. F.; Koeske, R. Peer Instruction in the General Chemistry Laboratory: Assessment of Student Learning. *J. Chem. Educ.* **2006**, *83* (5), 804−810.

(35) Wamser, C. C. Peer-led team learning in organic chemistry: Effects on student performance, success, and persistence in the course. *J. Chem. Educ.* **2006**, *83* (10), 1562−1566.

(36) Lyle, K. S.; Robinson, W. R. A Statistical Evaluation: Peer-Led Team Learning in an Organic Chemistry Course. *J. Chem. Educ.* **2003**, *80* (2), 132−134.

(37) Tien, L. T.; Roth, V.; Kampmeier, J. A. Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *J. Res. Sci. Teach.* **2002**, *39* (7), 606−632.

(38) Wilson, S. B.; Varma-Nelson, P. Small Groups, Significant Impact: A Review of Peer-Led Team Learning Research with Implications for STEM Education Researchers and Faculty. *J. Chem. Educ.* **2016**, *93* (10), 1686−1702.

(39) Towns, M. H.; Kreke, K.; Fields, A. An action research project: Student perspectives on small-group learning in chemistry. *J. Chem. Educ.* **2000**, *77* (1), 111−115.

(40) Mahalingam, M.; Schaefer, F.; Morlino, E. Promoting Student Learning through Group Problem Solving in General Chemistry Recitations. *J. Chem. Educ.* **2008**, *85* (11), 1577−1581.

(41) Wenzel, T. J. Cooperative group learning in undergraduate analytical chemistry. *Anal. Chem.* **1998**, *70* (23), 790a−795a.

(42) Evans, H. G.; Heyl, D. L.; Liggit, P. Team-Based Learning, Faculty Research, and Grant Writing Bring Significant Learning Experiences to an Undergraduate Biochemistry Laboratory Course. *J. Chem. Educ.* **2016**, *93* (6), 1027−1033.

(43) Jansson, S.; Soderstrom, H.; Andersson, P. L.; Nording, M. L. Implementation of Problem-Based Learning in Environmental Chemistry. *J. Chem. Educ.* **2015**, *92* (12), 2080−2086.

DOI: 10.1021/acscentsci.7b00133
ACS Cent. Sci. XXXX, XXX, XXX−XXX