# Does the Room Matter? Active Learning in Traditional and Enhanced Lecture Spaces

#### Jon R. Stoltzfus<sup>†\*</sup> and Julie Libarkin<sup>‡</sup>

<sup>†</sup>Biochemistry and Molecular Biology and <sup>†</sup>Geocognition Research Lab, Michigan State University, East Lansing, MI 48824-1319

#### ABSTRACT

SCALE-UP-type classrooms, originating with the Student-Centered Active Learning Environment with Upside-down Pedagogies project, are designed to facilitate active learning by maximizing opportunities for interactions between students and embedding technology in the classroom. Positive impacts when active learning replaces lecture are well documented, both in traditional lecture halls and SCALE-UP-type classrooms. However, few studies have carefully analyzed student outcomes when comparable active learningbased instruction takes place in a traditional lecture hall and a SCALE-UP-type classroom. Using a quasi-experimental design, we compared student perceptions and performance between sections of a nonmajors biology course, one taught in a traditional lecture hall and one taught in a SCALE-UP-type classroom. Instruction in both sections followed a flipped model that relied heavily on cooperative learning and was as identical as possible given the infrastructure differences between classrooms. Results showed that students in both sections thought that SCALE-UP infrastructure would enhance performance. However, measures of actual student performance showed no difference between the two sections. We conclude that, while SCALE-UP-type classrooms may facilitate implementation of active learning, it is the active learning and not the SCALE-UP infrastructure that enhances student performance. As a consequence, we suggest that institutions can modify existing classrooms to enhance student engagement without incorporating expensive technology.

## **INTRODUCTION**

Incorporation of more active learning in instruction has become a major goal of efforts to improve science, technology, engineering, and mathematics (STEM) education at the undergraduate level. Active learning is based on constructivist theory-the idea that students must create their own knowledge in order for learning to persist (Dori and Belcher, 2005). One core feature of active learning in the classroom is a decrease in lecturing during which students passively listen and an increase in outcome-related activities in which students actively develop their own understanding (Andrews et al., 2011). The use of writing-to-learn (Reynolds et al., 2012), drawing-to-learn (Quillin and Thomas, 2015), and talking-to-learn (Tanner, 2009) can all facilitate students' construction of their own knowledge and can be used to promote active learning. When implemented in the classroom to replace lecture, active learning typically involves both student-student and student-instructor interactions (Andrews et al., 2011). Because of these student-student interactions, active learning is frequently linked to peer instruction and cooperative learning. Peer instruction frequently puts students together in informal ways that promote discussion of questions during class (Crouch and Mazur, 2001). Cooperative learning typically puts students together in more formal situations in which they must work together to promote each other's success; cooperative learning also usually involves some form of peer instruction. Successful implementation of cooperative learning requires careful instructional design to ensure positive interdependence among group members, including promoting one another's success, being held accountable both as individuals and as a team, using appropriate interpersonal skills, and spending time evaluating group function (Johnson *et al.*, 1991).

Michèle Shuster, Monitoring Editor Submitted March 15, 2016; Revised August 16, 2016; Accepted August 17, 2016

CBE Life Sci Educ December 1, 2016 15:ar68 DOI:10.1187/cbe.16-03-0126

\*Address correspondence to: Jon R. Stoltzfus (stoltzfu@msu.edu).

© 2016 J. R. Stoltzfus and J. Libarkin. CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/ by-nc-sa/3.0).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

The impact of well-designed active learning is so well documented that a recent meta-analysis of studies comparing carefully designed active learning with traditional lecture concluded that the positive impacts of active learning over lecture are so well established that future research should focus on the relative efficacy of different active-learning approaches rather than comparing active learning with lecture (Freeman et al., 2014). Many aspects of instructional design will influence the impact of active learning on student outcomes. Instructional approaches such as peer instruction, cooperative learning, and flipped instruction are common methods used to facilitate active learning (Johnson et al., 1991; Crouch and Mazur, 2001; Smith et al., 2009; Strayer, 2012). All three techniques depend on students working together within the classroom, and these techniques are often used in tandem; in addition, the application of flipped instruction, wherein students learn concepts and vocabulary before attending class, is typically thought to free up class time for engagement in active learning. Flipped instruction has been shown to improve student performance, attendance, satisfaction, metacognition, and cooperative learning strategies (Stockwell et al., 2015; van Vliet et al., 2015; Hibbard et al., 2016).

Likewise, classroom space and infrastructure design impact implementation of active learning in a classroom. A number of universities have designed classrooms intended to facilitate the use of active-learning techniques. In these spaces, students are seated around tables rather than in traditional rows and are provided with technology, such as computers, large monitors, and/or large whiteboards, to allow access to digital instructional material and to facilitate collaboration within and between groups (Dori and Belcher, 2005; Beichner et al., 2007; Whiteside et al., 2010). For the remainder of this paper, we adopt the terminology of "SCALE-UP-type" to describe these flexible instructional spaces, patterned after perhaps the most familiar project (Student-Centered Active Learning Environment with Upside-down Pedagogies), to design and implement these spaces within university settings (Beichner et al., 2007). Several studies suggest that SCALE-UP-type classrooms can improve a wide range of desirable student outcomes (Dori and Belcher, 2005; Dori et al., 2007; Beichner et al., 2007; Gaffney et al., 2008; Brooks, 2011; Cotner et al., 2013).

Not all studies have shown positive outcomes from active learning, flipped instruction, or SCALE-UP-type spaces. A large study of introductory biology courses at universities across the United States found no correlation between reported active-learning levels in the classroom and student learning gains related to key evolutionary concepts (Andrews et al., 2011). The authors of this study suggest that the ability to create well-designed instruction produces positive student outcomes; such positive outcomes are unlikely in the presence of any instruction, including active learning, that lacks high-quality instructional design. Similarly, a study of the impacts of flipped instruction found no differences in student outcomes between students from a section taught using flipped instruction and students from a section incorporating carefully designed active learning and postclass homework. (Jensen et al., 2015). The authors of this study suggest that carefully designed instruction using active learning can be effective whether implemented in or out of the classroom. Finally, two studies of SCALE-UP-type classrooms suggest mixed results (Brooks and Solheim, 2014; Straumsheim, 2014). Taken together, these studies suggest that carefully designed instruction is a necessary prerequisite to learning, regardless of the amount of active learning implemented or the room type used.

The mixed results from prior studies suggest a need to further dissect how interactions between active learning, flipped instruction, and instructional spaces facilitate increased student learning gains and performance on course assessments. The goal of the current research is to determine how students' perceptions and performance are impacted by equivalent active learning–based instruction in SCALE-UP–type and traditional lecture-type classroom spaces. This research has many similarities to previous studies (Brooks, 2011; Cotner *et al.*, 2013) but includes a measure of pre- and postcourse content knowledge using a validated assessment instrument and incorporated instruction specifically designed around a flipped classroom model.

### MATERIALS AND METHODS

## **Ethics Statement**

The institutional review board of our university's Human Research Protection Program granted permission for this research, and all participants were given the option to opt out of the research. One student who chose to opt out of the study was not included in the study description or analysis.

## Subjects

This study was carried out at a large, public, doctorate-granting university in the midwestern United States. Students in the study (n = 110) were predominantly non–science majors enrolled in an integrative studies biology course designed to fulfill the university's general education requirements.

Students self-selected sections of the course without any a priori knowledge of the research study. Slightly more students were enrolled in the section using a traditional classroom space (Traditional) than in the section using a SCALE-UP-type space (SCALE-UP). Enrollment in both courses was capped at 72 students. Enrollment on the first day of class was 60 students in the SCALE-UP-type classroom and 68 students in the Traditional classroom. Seven to 11 students dropped the course in each section, leaving final enrollments of 49 students in the SCALE-UP-type classroom and 61 students in the Traditional classroom. Students who completed the content knowledge survey both at the start of the course and the end of the course were included in analysis of performance (SCALE-UP: n = 34; Traditional: n = 37). Students who completed the classroom and technology and infrastructure survey at the end of the course were included in the analysis of student perceptions (SCALE-UP: n = 33; Traditional: n = 42). Average ACT scores, grades, and other demographic variables for each section and the subsets of students included in these analyses are shown in Table 1.

## Study Design

This study used a comparative quasi-experimental design to test the impact of instructional space design and technology on student performance in two different sections of the same course. The two sections of a nonmajors biology course were taught in parallel by the same instructor. The SCALE-UP section met from 12:40–2:00 pm on Tuesdays and Thursdays in a room

TABLE 1. Average A( in each section who	CT scores, aver completed bot	age incomi h the pre- a	ing GPA, i and postc	average c	ourse gra ntent knc	ide, and c wiledge :	other der survey, a	nographic nd the stu	data for dents in	all stude each seo	ents in th ction who	e Traditic comple	onal sections ted the te	on and Si schnolog	CALE-UP	section a	ind the si re survey	udents
	ACT	and grades	5	Gen	ıder				Ethni	city					Year	in colle	ge	
	ACT score (percentage of sample with score) <sup>a</sup>	Average incoming GPA	Average course grade	Female	Male	White (non-Hispanic)	Asian (non-Hispanic)	Black or African American (non-Hispanic)	International	Not reported	Two or more races (non-Hispanic)	Hispanic ethnicity	Hawaiian/ Pacific islander (non-Hispanic)	Freshman	Sophomore	Junior	Senior	Lifelong education
Traditional section																		
All students	25.25 (87%)	2.92	3.05	%69	31%	64%	5%	11%	7%	2%	3%	7%	2%	62%	28%	5%	5%	%0
Pre-post survey	25.25 (86%)	0 2.94	3.28	68%	32%	70%	5%	8%	8%	%0	%0	5%	3%	68%	22%	5%	5%	%0
Tech survey	25.37 (83%)	3.07	3.33	81%	19%	62%	7%	10%	7%	2%	5%	5%	2%	67%	24%	5%	5%	%0
SCALE-UP section																		
All students	25.08 (80%)	) 2.75	3.36	51%	49%	71%	6%	2%	16%	4%	%0	0%0	0%0	57%	29%	8%	4%	2%
Pre-post survey	25.33 (79%)	) 2.78	3.51	53%	47%	76%	6%	3%	12%	3%	%0	%0	0%0	53%	29%	%6	6%	3%
Tech survey	24.92 (76%)	2.73	3.53	52%	48%	73%	6%	3%	15%	3%	%0	%0	0%0	55%	27%	9%6	6%	3%
TOT correct more and	bitable for all chid	onte on collog	in the cor	0311														

CBE—Life Sciences Education • 15:ar68. Winter 2016

containing circular tables and movable chairs that allowed students to conveniently work in groups of three or four students (Figure 1). Each table had power outlets and flat-screen monitors on which the instructor could display digital content or to which the group could connect laptop computers or tablets and view the digital content of their choice. The room had a seating capacity of 72 students. The Traditional section met from 2:40-4:00 pm on Tuesdays and Thursdays in a traditional lecture hall with fixed desks in a tiered arrangement with no power outlets or flat-screen monitors available to students. The room had a seating capacity of 150 students (Figure 1). Finally, large screens allowing the instructor to display digital media to all students were available in both classrooms (Figure 1). The SCALE-UP-type classroom had four large screens located at the corners of the room, and the Traditional classroom had one large screen at the front of the lecture hall.

Both classrooms shared some common technology. Wireless Internet was available in both classrooms, and students were required to bring a mobile device with which they could connect to the Internet. These mobile devices included laptop computers, tablets, and cell phones. Each student used his or her device to respond to in-class questions related to content acquisition from the preclass reading and homework or as scaffolding for the group assignments. Each group of students was also provided with a tablet and stylus that allowed them to make freehand drawings as part of modeling activities. Students without wireless devices could also use these tablets when Internet connectivity was required.

Both sections were taught by the same instructor using an identical flipped approach. Lesson plans, preclass assignments, homework, formative assessments, and summative assessments were identical for both sections. The goal was for student concept acquisition to take place before class using lecture videos, readings, and homework. During class, students worked in formal groups to apply concepts to solve problems or build new understanding.

Formal groups were assigned by the instructor during the second week of class based on responses to a survey students completed during the first week of class. The target group size was three students, as this allowed students in the Traditional classroom to easily sit together and share resources. Group size varied from two to five students as students dropped the course and some groups were combined. The amount of time students planned to spend working on the class was used to put students with similar work ethics together. Self-reported grade point average (GPA) allowed groups to be mixed across potential ability level.

A typical class began with students using mobile devices, either personal devices or the tablets provided by the instructor, to answer several questions related to the readings and homework. Question types, including multiple choice, multiple select, circling regions on a diagram, numeric, and short answer, were answered using Pearson's Learning Catalytics platform. During this time, students were encouraged to work collaboratively and to ask the instructor questions about anything that was unclear regarding the preclass material. Following each question, the instructor would lead a whole-class discussion about the question. This set the stage for the group work, during which students either developed a model, analyzed an article from the popular press, or developed a scientific argument.



FIGURE 1. Students in the SCALE-UP classroom section sat in movable chairs at round tables with monitors on the tables (A), while students in the Traditional section sat in fixed desks in tired rows (B). SCALE-UP room dimensions are  $41 \times 56$  feet, with an  $8 \times 14$  foot alcove for the instructor podium. Traditional room dimensions are  $34 \times 57$  feet.

During this time, the instructor and an undergraduate learning assistant answered questions and interacted with groups, asking students about their progress and checking student understanding of key concepts. Following the group work, one or two groups would present their models, analyses, or arguments to the entire class for a critique. The class was then given time to critique the presentation, and several groups were selected to share their critiques with the whole class. All groups were then allowed to revise their models, analyses, or arguments based on the class critique and submit their final products to the learning-management system for evaluation and feedback. These daily group assignments were evaluated for completeness and quality using a predefined rubric. The majority of the points were earned for following instructions, with the remaining points based on the quality of the work.

For example, students were required to complete online readings and homework about gene expression on their own. In class, groups of students 1) answered Learning Catalytics questions about preclass material; 2) developed models explaining why an undifferentiated stem cell can become either an eye cell or a heart cell; 3) evaluated models developed by other groups; and 4) revised their own models. Before the next class, individual students then read an article from the popular press related to stem cell differentiation and use of stem cells in regenerative medicine. During class, groups of students 1) answered Learning Catalytics questions about the article; 2) used their models to analyze the article and make scientific arguments supporting or refuting claims made in the article; 3) evaluated arguments made by other groups; 4) revised their own arguments; and 5) submitted their groups' arguments to the learning-management system for evaluation and feedback.

Instruction was designed to maximize opportunities to use technology for sharing ideas, providing feedback, and fostering group interaction. Students in both sections collaborated in groups to answer questions and then used mobile devices to send in answers and receive feedback. Each group in both sections was provided with a Microsoft Surface Pro 2 tablet and stylus that allowed groups to create and submit freehand drawings as part of modeling assignments. In the SCALE-UP– type classroom, students could connect this tablet to the group monitor so that all group members could easily see the screen; in the Traditional classroom students sat close together to view the tablet screen. In the SCALE-UP–type classroom, group presentations could be displayed on each group's monitor and on the large classroom screens, while in the Traditional classroom, presentations were displayed on the large projection screen at the front of the room. Overall, implementation of instructional design was as similar as possible in the two sections, considering the differences in infrastructure and technology in the two classrooms.

Because classroom assessments were used to compare sections and the instructor knew of the study, the potential for bias during the grading process exists. To reduce this possibility, exams were assigned a random code and deidentified, and exams from the two sections were randomly mixed before being graded. Assessments other than exams could not be deidentified, but bias was reduced by alternating between the different sections during the grading process.

## **Measures of Student Performance**

Multiple measures of student outcomes were developed, including validation steps as needed, with analyses carried out both within and between sections. All statistical analysis was carried out using SPSS, version 21.

*Classroom Effort*. Measures of individual effort and group effort were collected to determine similarity of effort across the two sections and the impact of effort on postcourse content knowledge. Individual assignments included online homework, a short paper, and exam grades. In addition, student responses sent in through Learning Catalytics with their personal mobile devices were used to determine individual participation. Individual effort was calculated by multiplying the number of days a student interacted using Learning Catalytics by the student's total scores on individual assignments. Group assignments included daily group assignments and a final group presentation.

A portion of the daily group assignment scores was earned simply by turning in the group assignment. Group effort was calculated by combining group scores on all group assignments. Thus, both individual effort and group effort included a mixture of participation and performance metrics and provide insight beyond purely performance-based measures into how much effort individuals and groups put into the course. Group effort and individual effort for individuals from each section were compared using independent-samples *t* tests.

Pre-Post Content Knowledge. This course focused primarily on biotechnology. Eighteen multiple-choice items covering relevant topics were chosen from several published biology concept inventories (Klymkowsky et al., 2003; Bowling et al., 2008; Smith et al., 2008; Shi et al., 2010). These items were modified to increase validity through alignment with item construction standards (Haladyna and Downing, 1989; Haladyna et al., 2002; Frey et al., 2005) and expected student knowledge based on course content. A 19th quality control item asking students to select a specific answer was also included to identify students who were not carefully reading and answering the questions. This 19-item survey was administered online via the course management system before the second day of class and before the last week of class in order to collect data on students' pre- and postcourse knowledge of course content. Students were instructed to "answer each question based on what you know without using any additional resources (Google, a textbook, your friends, etc.)" to "make an honest effort to carefully read and answer all questions" and were told that they would earn the extra credit "regardless how many correct answers you select as long as you make an honest effort." Students earned extra credit worth 0.33% of their overall course scores for completion of each survey regardless of their scores on the assessment.

To ensure that a single construct was being measured by this set of items, we used factor analysis to establish unidimensionality. Unidimensionality is necessary for establishing that the set of items together measure a single, meaningful construct. Without unidimensionality, the score on the content knowledge survey would have little meaning. We investigated unidimensionality of the 18 conceptual items through exploratory factor analysis of posttest data, with iterative removal of items as the number of constructs measured by the test was evaluated. Despite emerging from established instruments, the set of items did not align with the instruments from which they were gathered and overall exhibited strong nonunidimensional behavior. Ultimately, a set of eight items was identified that adequately measured a single construct (Table 2). Given that unidimensionality was not established in any of the four studies from which items were sourced, the removal of 10 items from the set was not surprising. For this set of eight items, the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.63, above the 0.6 value recommended for factor analysis, and Bartlett's test of sphericity was significant ( $\chi^2(28) = 61.3, p < 0.001$ ). The scree plot for this set of items suggested one dominant factor, and only one eigenvalue was meaningfully above one. Finally, a Cronbach's alpha of 0.61 was calculated for this scale; an alpha above 0.6 is considered adequate for small samples (Hair et al., 2006). Taken together, this set of eight items is considered to measure a unidimensional construct related to molecular biol-

TABLE 2. Factor loadings for eight items that factored together to produce a single scale and that were used as the bases for pre- and postcourse content knowledge scores

Question	Loading	Communality
2	0.643	0.413
4	0.491	0.241
5	0.478	0.228
6	0.401	0.161
8	0.461	0.212
12	0.502	0.252
14	0.584	0.341
18	0.551	0.304

ogy relevant to biotechnology. The total number correct out of these eight items was used to generate pre- and postcourse content knowledge scores for each student in both course sections. The complete content knowledge survey is available in the Supplemental Material.

Precourse content knowledge scores and postcourse content knowledge scores from each section were compared using mixed-design analysis of variance (ANOVA) for a betweenwithin subjects analysis. The influence of precourse knowledge scores, gender, class level, group effort, individual effort, and section on postcourse knowledge scores was evaluated through linear regression. These covariates were used because of demographic differences across individuals and sections.

#### **Measures of Student Perceptions**

Students' perceptions of how classroom technology and infrastructure influenced their experiences in the course were obtained via a survey administered at the end of the semester just before submission of grades. The survey was administered online through the university's learning-management system. The survey consisted of 13 multiple-choice questions and four open-ended questions. The questions focused on the usefulness of specific aspects of the classroom design for students, including how students thought their performance in the course would have changed if that aspect of the course had changed. Students earned 0.33% extra credit on their overall course scores for completing the survey.

Five multiple-choice questions were identical between the sections and asked students about some aspect of the classroom, technology, or instruction that was identical between the sections. Seven questions asked students about unique aspects of their specific classroom setting, such as the value of group monitors in the SCALE-UP-type classroom. In two questions, students were asked how they thought their performance in the course might have changed if their classroom had contained features from the other classroom type. One multiple-choice question was a quality-control question designed to check for students who were clicking through the survey without reading the questions. Finally, four open-ended questions allowed students to explain aspects of the classroom and technology that were helpful or that could be improved. The responses to openended questions were analyzed by counting the number of instances in which students mentioned a particular aspect of the course. The complete surveys are available in the Supplemental Material.

TABLE 3.	Variable	means	across	sections
----------	----------	-------	--------	----------

	Section	n	Mean	SD	SEM
Precourse content	Traditional	37	2.757	1.422	0.234
knowledge	SCALE-UP	34	2.147	1.676	0.302
Postcourse content	Traditional	37	4.595	1.832	0.301
knowledge	SCALE-UP	34	3.706	2.048	0.344
Group effort	Traditional	61	39.508	7.066	0.905
	SCALE-UP	49	41.949	2.900	0.414
Gender	Traditional	61	1.690	0.467	0.060
	SCALE-UP	49	1.510	0.505	0.072
Class level	Traditional	61	1.510	0.766	0.098
	SCALE-UP	49	1.670	0.944	0.135
Individual effort	Traditional	61	802.062	236.549	30.287
	SCALE-UP	49	841.360	223.219	31.890

#### RESULTS

#### Measures of Student Performance

*Classroom Effort*. Independent-samples *t* test comparison for individual effort and group effort variables were used to compare student effort across the Traditional section and SCALE-UP section. Results indicate no difference for either individual or group effort between students in the SCALE-UP–type classroom and students in the Traditional classroom (Table 3). These results suggest students' efforts across sections were comparable and students in both sections performed equally well on classroom assessments, including exams, homework, in-class activities, and group projects.

**Pre–Post Content Knowledge**. Only students who completed both the pre- and postinstruction content knowledge survey were included in the analysis. Response rates between classrooms yielded similar total numbers of students, with n = 34 (69%) of SCALE-UP students and n = 37 (61%) of Traditional students completing both tests. No obvious differences in average ACT scores, grades, or other demographic variables were observed between the subset of students who completed both the pre- and posttest and the sections as a whole (Table 1).

An independent-samples *t* test comparison of precourse content knowledge scores indicates no difference in precourse content knowledge between students in the SCALE-UP-type classroom and students in the Traditional classroom (Table 3). These data suggest that students in both sections began the course with similar content knowledge.

A mixed-design ANOVA for between–within subjects analysis was conducted to evaluate the impact of the classroom type (SCALE-UP, Traditional) on student content knowledge across two time periods (preinstruction and postinstruction); the impact of other variables is addressed below in the section on *Linear Regression Analysis*. No significant interaction between section and time was observed, Wilks lambda = 0.996, *F*(1, 69) = 0.302, *p* = 0.59, partial eta-squared = 0.004. A significant main effect for time existed, Wilks lambda = 0.61, *F*(1, 69) = 44.72, *p* < 0.001, partial eta-squared = 0.393. Although both groups exhibited increases in content knowledge after instruction, students in the Traditional group exhibited greater gains than those in the SCALE-UP group, with moderate effect size.

Although the full 18-item form was not unidimensional (as noted in *Materials and Methods*), mixed-design ANOVA was run

on total scores to address potential reader concerns that those data would support SCALE-UP spaces as more effective. Results were nearly identical to the valid eight-item form. Note that this 18-item form cannot be considered a valid measure in the absence of unidimensionality, hence the reporting of statistical analysis for the eight-item form only.

Linear Regression Analysis. Linear regression was used to investigate other variables that may impact post knowledge and explain the difference in Traditional and SCALE-UP sections. Linear regression indicates that interaction effects are not significant; main effects of gender, class level, and group effort were also insignificant predictors of post knowledge. A stepwise regression including only the significant variables of pre knowledge, individual effort, and section was then run (Table 4). Overall adjusted model fit was  $R^2 = 0.124$ , meaning that the model explains 12.5% of the variance in postknowledge scores. Prior knowledge and individual student effort are the only significant variables, together explaining 12.4% of the adjusted variance in postknowledge scores. All other variables, including section, do not explain any significant portion of the variance, suggesting that, when pre knowledge and individual effort are considered, section placement plays little role in explaining postknowledge scores.

## Student Perceptions: Technology and Infrastructure Survey

Students' perceptions of classroom technology and infrastructure were measured at the end of the semester in both sections. About two-thirds of the students in each section completed the survey with similar response rates across sections, with n = 33(67%) of SCALE-UP students and n = 42 (69%) of Traditional students completing the survey. With the exception of a larger percentage of females completing the technology and infrastructure survey in the Traditional section, no obvious differences in average ACT scores, grades, or other demographic variables were observed between subsets of students who completed the survey and the entire sections (Table 1).

Students reported a similar level of interaction with the instructor in both sections, in line with the instructor's intentions. Similarly, all respondents in both sections reported owning Wi-Fi-enabled mobile devices that could be used for Internet-based classroom activities. Despite access to personal devices, more than 70% of students in each section reported that the tablet was very useful during group activities and that team performance on classroom assignments would have suffered without this technology (Tables 5 and 6). When asked in the open-ended questions what technology in the course helped them learn, eight of 23 respondents from the SCALE-UP section and 10 of 21 respondents from the Traditional section mentioned the tablets in a favorable manner, including indications that they were useful for freehand drawing. For example, one student reported, "The tablets were very helpful when drawing scientific models because we didn't have to use a key pad or mouse which would have made it very difficult to draw detailed characteristics." Taken as a whole, responses suggest that students valued the tablets, because they were able to use them to create freehand drawings during the modeling activities, although some students found the tablets hard to use.

Students in the SCALE-UP section responded favorably to questions regarding the group monitor's utility and impact on

Does	the	Room	Matter?

	В	SE B	β	В	SE B	β	в	SE B	β	в	SE B	β
0	.366	0.139	$0.301^{*}$	0.404	0.138	0.333*	0.402	0.141	0.331	0.343	0.144	0.282
				0.004	0.002	$0.327^{*}$	0.003	0.002	0.318	0.003	0.002	0.309
				-0.057	0.080	-0.108	-0.055	0.082	-0.105	-0.029	0.082	-0.054
							0.012	0.248	0.006	0.088	0.248	0.041
emale)							0.152	0.461	0.038	0.037	0.459	0.009
nal, SCALE-UP)										0.268	0.157	0.207
	)	0.077			0.124			0 000				
					-			~~~~			0.125	
	,	277*			2 245			0.055			0.125 2.022	

TABLE 5. Percentage of respondents indicating each level of utility for different aspects of technology or infrastructure in their classrooms

	Q6: T util	ablet ity	Q4: monito	Group or utility <sup>a</sup>	Q9: Cla type	ssroom utility
Utility of resource	Traditional	SCALE-UP	Traditional	SCALE-UP	Traditional	SCALE-UP
Very useful	74%	70%	33%	73%	21%	73%
Somewhat useful	23%	27%	47%	24%	51%	27%
Not useful	2%	3%	19%	3%	26%	0%
Did not answer	0%	0%	2%	0%	2%	0%

<sup>a</sup>In the case of group monitors, students in the Traditional classroom were asked how useful a group monitor would have been had it been added to the classroom.

their performance, while responses from students in the Traditional section to question about the impact of adding monitors to the classroom were mixed (Tables 5 and 6). Several students from the SCALE-UP section noted issues with their group monitors not working in the open-ended portion of the survey.

Students in the SCALE-UP section also responded favorably to questions regarding the classroom layout, while responses from students in the Traditional section to questions about the impact of the classroom were less favorable (Tables 5 and 6). Answers on the open-ended section of the survey suggest that the classroom layouts influenced interactions between group members. In the SCALE-UP section, nine of the 22 responses regarding aspects of the classroom that helped learning mentioned interactions with other students, for example, "The big tables were very helpful rather than desks. It's easier for teams to work together that way." SCALE-UP students provided no suggestions related to improving student-student interactions. However, five of the 20 responses from the Traditional section regarding how the classroom could be improved mentioned the seating arrangements and interactions with other students. For example, one student wrote: "I think that it would be helpful if we were able to ... move our chairs around. At times it could be

TABLE 6. Percentage of respondents indicating each level of performance change if different aspects of technology or infrastructure in their classrooms were removed<sup>a</sup>

	Q7: Re tabl	emove ets	Q5: A remove mon	dd or e group itors	Q10: 0 the ty class	Change ype of room
Impact on performance if resource was changed	Traditional	SCALE-UP	Traditional	SCALE-UP	Traditional	SCALE-UP
Suffered	77%	65%	7%	54%	5%	78%
Not changed	23%	35%	65%	38%	53%	22%
Improved	0%	0%	28%	8%	40%	0%
Did not answer	0%	0%	0%	0%	0%	0%

<sup>a</sup>In the case of group monitors, students in the Traditional classroom were asked how their performance would have changed had group monitors been added to the classroom.

Pre knowled

/ariable

Individual eff 3roup effort Gention (Trad

**Class** level

djusted R<sup>2</sup> for change

< = 0.05

TABLE 4. Summary of hierarchical regression analysis of post knowledge

a bit disjointed trying to work with three people and not having all people be able to face each other/see the device." Another student said, "If we had a different classroom with tables and chairs, I feel like that would have helped communication when we needed to work on team projects and activities."

The presence of power outlets in the SCALE-UP-type classroom also resulted in differences in student experiences. Results from the survey indicate that 54% of respondents in the SCALE-UP-type classroom charged their devices during class once or more a week and that 12% of respondents from the Traditional classroom had issues with their mobile devices and tablets losing power during class at least once during the semester. When asked in the open-ended portion of the survey how the classroom could be improved, three of the 20 respondents from the Traditional classroom noted the need for power outlets in the open-ended portion of the survey.

# DISCUSSION

The major result of our study is that the SCALE-UP-type classroom did not enhance student performance relative to the Traditional-type classroom. We found no significant difference in individual or group effort in the two sections and no significant difference on classroom assessment performance between the two sections. Although mixed-design ANOVA suggests a difference, in favor of the Traditional section, on postknowledge scores, linear regression results indicate that individual students' prior knowledge and individual efforts explain section differences. We acknowledge that our study suffers from small sample size, which may have inhibited our ability to detect differences across treatments and pseudoreplication (Hurlbert, 1984), as we applied statistical analysis to data gathered in a study that lacks replication across the hypothesis space being tested. However, these results are significant, as they contradict the results of two similar previously published studies (Brooks, 2011; Cotner et al., 2013). All three studies are relatively small and suffer from pseudoreplication, common faults of this type of study due to constraints such as the limited number of sections typically taught by the same instructor, changes in teaching assignments over time, and the availability of specific instructional spaces. Completing more studies that include independent replicates, ideally across institutions, and paying careful attention to details, as per the discussion that follows, would enable examination of why the outcomes of these three small studies differ.

The two previous studies (Brooks, 2011; Cotner *et al.*, 2013) used research questions and experimental design similar to this study and found that student performance was enhanced in the SCALE-UP–type classrooms. These previous studies compared performance of students in active-learning spaces and traditional classrooms when the courses were taught using the same instructor, teaching methods, and assessments (Brooks, 2011; Cotner *et al.*, 2013). All three studies were carried out in introductory biology courses for non–science majors; and all three studies used teaching approaches that were learner centered. Differences between the three studies cannot be attributed to class size, as Brooks (2011) reported sections of similar sizes to those discussed here.

Some important differences in data collection and analysis between our study and both prior studies might explain the different results. In the current study, we used a validated content knowledge measure rather than course grades as our outcome variable. While grades provide some insight into student learning, grades may exhibit more noise than a psychometrically defined unidimensional scale.

Subtle differences in instructional approach may also explain the differences across studies. While all three studies used "active-learning" approaches, the description of the active-learning strategies used in each study varied. We describe our active-learning strategy as flipped instruction; the active-learning strategy in Brooks (2011) was described as a hybrid lecture/problem-solving approach; and Cotner et al. (2013) describe use of active-learning techniques without further explanation. In our instructional approach, groups were provided with instructions for developing a model or analyzing an article and often spent one-half to two-thirds of each class period working on the activity before a group was chosen to report to the entire class. The instructor and students interacted throughout each class as students asked questions and shared progress. Students in the current study reported similar levels of interaction across the SCALE-UP and Traditional sections. In contrast, Brooks (2011) and Cotner et al. (2013) found that instructors interacted more with students in the SCALE-UPtype classrooms than in traditional rooms. This equality of student-instructor interactions across sections may have contributed to the similar learning across our two sections. If this is the case, there are important implications for large-enrollment sections, as one instructor can interact effectively only with a limited number of groups during any one class period. One possible solution is the use of instructional teams of well-trained graduate and undergraduate students who interact with students during class (Smith et al., 2005). However, more research is needed to determine the impact of student-instructor interactions on student performance and the effectiveness of substituting graduate or undergraduate students for instructors during these interactions (Kendall and Schussler, 2012; Knight et al., 2015).

Several factors likely contributed to common levels of interactions across sections. First, the tablet provided a focal point for the groups during many activities across both sections. In the Traditional section, the rows were curved. This allowed students to sit in three consecutive desks as well as adjacent rows and still be able to simultaneously view a common tablet or device. The Traditional classroom design and the use of a common device may have enhanced student–student interactions in the Traditional section and contributed to the lack of differences in student outcomes.

Students in this study exhibited positive views of the technology and infrastructure in SCALE-UP-type classrooms and felt the room layout enhanced their performance even when our postcourse analysis indicated that it did not. Previous studies have also shown that students have positive views of SCALE-UP-type rooms (Dori and Belcher, 2005; Beichner *et al.*, 2007; Whiteside *et al.*, 2010). In the current study, SCALE-UP students reported that technology in the room and the movable chairs and round tables helped them learn and that their performance in the class would have suffered if these items had not been present. Similarly, Traditional classroom students indicated a preference for a more flexible seating arrangement. Taken together, these data support the idea that students prefer a SCALE-UP-type classroom layout for active learning and suggest that this layout fosters interactions between group members.

Finally, our study agrees with results from other recent studies showing that student ownership of mobile devices is approaching 100% (Cassidy et al., 2014) and that students are using their own mobile devices in classrooms to support and enhance their classroom learning (Biddix et al., 2015). Early reports on learning gains in SCALE-UP-type classrooms took place in an era when including technology in the classroom was important, because the technology allowed students to retrieve and interact with digital content that they would not otherwise have been able to access (Dori and Belcher, 2005; Beichner et al., 2007). In 2001, around the time the initial versions of SCALE-UP-type classrooms were developed, a minority of students owned laptops. For example, only about one-third of students at UVA owned a laptop (University of Virginia Instructional Technology Services, 2009). In 2001, students also did not own smartphones, as such devices did not exist; Blackberry released the first cell phone with email capability in 2001, and the first iPhone was not released until 2007. This meant that, in order for students to interact with digital media, technology needed to be provided by the instructor. Today, most students own a laptop and/or a smartphone. In 2014, student ownership of laptop computers was greater than 95%, and ownership of cell phones was greater than 98% (Cassidy et al., 2014). In our study, 100% of students reported owning a laptop. Because Wi-Fi is also present in classrooms on most college campuses, students can easily access and share digital information using their own devices. The current level of individual access to technology suggests that technology provided in most SCALE-UP-type classrooms, such as monitors, may no longer be necessary.

It is clear that active learning can improve student outcomes (Michael, 2006; Freeman et al., 2014). However, our understanding of the specific aspects of classroom technology, instructional approaches, and contextual factors that can lead to improvements in student outcomes is limited. For example, a major feature of flipped instruction that is thought to increase student learning is moving the less difficult task of concept acquisition out of the classroom and using valuable class time to focus on the more difficult task of concept application and problem solving. Counter to this idea, a recent study found no differences in student performance when class time was spent focusing on either content acquisition or content application and problem-solving (Jensen et al., 2015). The study found that flipped instruction did not improve student performance over that achieved in a previous section of the course that incorporated active-learning strategies. The authors of the study therefore suggested that well-designed active learning is the most important feature of effective instruction, not where (in class or at home) the active learning occurs.

Prior work coupled with the current study suggests a need for future work. Specifically, studies should investigate which specific aspects of instructional approaches, instructional technology, and learning spaces increase student learning. In addition, the specific methods used to measure student performance and learning gains likely impact findings, and future studies should incorporate a broad spectrum of research-quality metrics to help delineate how instructional approaches, technology, and instructional spaces impact student learning.

# CONCLUSIONS

On the basis of results of our study in conjunction with results from prior work (Andrews *et al.*, 2011; Brooks and Solheim, 2014; Straumsheim, 2014; Jensen *et al.*, 2015), we conclude that adding technology to a classroom, remodeling classrooms to facilitate interactions, flipping instruction, or even adding active learning to a course is not a panacea that produces better outcomes for students. Factors unique to each instructional situation likely influence outcomes, and care must be taken when assuming strategies shown to work in one situation will transfer to learning gains in similar situations.

Building instructional spaces requires significant expenditures, especially when technology is included in the classroom. In 2007, the University of Minnesota renovated two classrooms following a SCALE-UP-type model. Design, technology purchase and installation, and furniture for a room with 45 seats cost \$147,000, and a room with 117 seats cost \$269,000 (Whiteside et al., 2009). These costs did not include other expenses associated with renovating the rooms. At our institution, renovating the room in which the SCALE-UP section was taught cost \$192,000, and renovating a smaller SCALE-UPtype classroom with 36 seats cost \$128,000. The cost is substantially reduced to \$30,000, or one-fourth to one-sixth of the cost, when technology is not included in the renovation (S. Grabski, personal communication). Clearly, room renovations to facilitate group work are much cheaper in the absence of embedded technology. This greatly reduced cost and student preference for SCALE-UP spaces, coupled with the equivalent learning observed across the two sections in our study, suggests that altering classrooms to facilitate student-student and student-instructor interactions may be worth the cost. This is especially true in spaces designed for large numbers of students, where instructors will be unlikely to replicate the interactions made possible by the small number of students enrolled in the Traditional section in the current study.

As described by other authors and experienced firsthand in this study, there is a significant cost for remodeling classrooms to facilitate active learning and adding technology to classrooms (Cotner et al., 2013) and to developing flipped instruction (Jensen et al., 2015). Until we better understand the mechanisms by which these changes produce improved student outcomes, we should be cautious with our investments of scarce resources. Based on evidence that highly skilled instructors with intimate understanding of education research and active-learning pedagogy can have significant favorable impact on student learning, scarce resources may be best spent on 1) training faculty, 2) providing flexible spaces rather than embedding expensive technology into rooms, and 3) maintaining smaller sections or providing well-trained graduate teaching assistants or undergraduate learning assistants to maintain frequent and high-quality interactions between students and the instructional staff during class time. Finally, we should be cautious and not measure progress in education simply by the amount of active learning reported in classrooms or by the number of remodeled instructional spaces made available for faculty. Scientific teaching requires constant evaluation of student outcomes to determine what works and what does not work (Handelsman et al., 2004; American Association for the Advancement of Science, 2011). We encourage the use of validated research assessments in tandem with grades to investigate advances in undergraduate

learning. The use of multiple measures of learning is likely our best avenue for effective assessment and research into the impacts of instruction on students.

## **ACKNOWLEDGMENTS**

Funding from an internal LPF-CMP2 Innovation Grant awarded through the CREATE for STEM Institute provided partial support for this project. We also thank faculty in the Center for Integrative Studies in General Science and members of the Geocognition Research Lab for their assistance with this study and review of this article.

#### REFERENCES

- American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.
- Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST (2011). Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ 10, 394–405.
- Beichner RJ, Sual JM, Abbot DS, Morse JJ, Deardorff DL, Allain RJ, Bonham SW, Dancy MH, Risley JS (2007). The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) Project. In: Reviews in PER, vol. 1: Research-Based Reform of University Physics, ed. E Redish and P Cooney, College Park, MD: American Association of Physics Teachers. www.percentral.com/PER/per\_reviews/media/volume1/ SCALE-UP-2007.pdf (accessed 7 September 2014).
- Biddix JP, Chung CJ, Park HW (2015). The hybrid shift: evidencing a student-driven restructuring of the college classroom. Comput Educ 80, 162–175.
- Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics 178, 15–22.
- Brooks DC (2011). Space matters: the impact of formal learning environments on student learning. Br J Educ Technol 42, 719–726.
- Brooks DC, Solheim CA (2014). Pedagogy matters, too: the impact of adapting teaching approaches to formal learning environments on student learning. New Dir Teach Learn 137, 53–61.
- Cassidy ED, Colmenares A, Jones G, Manolovitz T, Shen L, Vieira S (2014). Higher education and emerging technologies: shifting trends in student usage. J Acad Libr 40, 124–133.
- Cotner S, Loper J, Walker JD, Brooks DC (2013). "It's not you, it's the room" are the high-tech, active learning classrooms worth it? J Coll Sci Teach 42, 82–88.
- Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. Am J Phys 69, 970–977.
- Dori YJ, Belcher J (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? J Learn Sci 14, 243–279.
- Dori YJ, Hult E, Breslow L, Belcher J (2007). How much have they retained? Making unseen concepts seen in a freshman electromagnetism course at MIT. J Sci Educ Technol 16, 299–323.
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci USA 111, 8410–8415.
- Frey BB, Petersen S, Edwards LM, Pedrotti JT, Peyton V (2005). Item-writing rules: collective wisdom. Teach Teach Educ 21, 357–364.
- Gaffney JD, Richards E, Kustusch MB, Ding L, Beichner RJ (2008). Scaling up education reform. J Coll Sci Teach 37, 48–53.
- Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006). Multivariate Data Analysis, 6th ed., Upper Saddle River, NJ: Pearson Education.
- Haladyna TM, Downing SM (1989). A taxonomy of multiple-choice item-writing rules. Appl Meas Educ 2, 37–50.
- Haladyna TM, Downing SM, Rodriguez MC (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 15, 309–333.

- Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, et al. (2004). Scientific teaching. Science 304, 521–522.
- Hibbard L, Sung S, Wells B (2016). Examining the effectiveness of a semi-selfpaced flipped learning format in a college general chemistry sequence. J Chem Educ 93, 24–30.
- Hurlbert SH (1984). Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54, 187–211.
- Jensen JL, Kummer TA, Godoy PDdM (2015). Improvements from a flipped classroom may simply be the fruits of active learning. CBE Life Sci Educ 14, ar5.
- Johnson DW, Johnson RT, Smith KA (1991). Cooperative Learning: Increasing College Faculty Instructional Productivity (ASHE-ERIC Higher Education Report No. 4), Washington, DC: George Washington University.
- Kendall KD, Schussler EE (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. CBE Life Sci Educ 11, 187–199.
- Klymkowsky MW, Garvin-Doxas K, Zeilik M (2003). Bioliteracy and teaching efficacy: what biologists can learn from physicists. Cell Biol Educ 2, 155–161.
- Knight JK, Wise SB, Rentsch J, Furtak EM (2015). Cues matter: learning assistants influence introductory biology student interactions during clicker-question discussions. CBE Life Sci Educ 14, ar41.
- Michael J (2006). Where's the evidence that active learning works? Adv Physiol Educ 30, 159–167.
- Quillin K, Thomas S (2015). Drawing-to-learn: a framework for using drawings to promote model-based reasoning in biology. CBE Life Sci Educ 14, es2.
- Reynolds JA, Thaiss C, Katkin W, Thompson RJ (2012). Writing-to-learn in undergraduate science education: a community-based, conceptually driven approach. CBE Life Sci Educ 11, 17–25.
- Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK (2010). A diagnostic assessment for introductory molecular and cell biology. CBE Life Sci Educ 9, 453–461.
- Smith AC, Stewart R, Shields P, Hayes-Klosteridis J, Robinson P, Yuan R (2005). Introductory biology courses: a framework to support active learning in large enrollment introductory science courses. Cell Biol Educ 4, 143–156.
- Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT (2009). Why peer discussion improves student performance on in-class concept questions. Science 323, 122–124.
- Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ 7, 422–430.
- Stockwell BR, Stockwell MS, Cennamo M, Jiang E (2015). Blended learning improves science education. Cell 162, 933–936.
- Straumsheim C (2014). Room to experiment. Inside Higher Ed. www.insidehighered.com/news/2014/12/12/interactive-learning-spaces-center -ball-state-us-faculty-development-program (accessed 7 September 2015).
- Strayer J (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. Learn Environ Res 15, 171–193.
- Tanner KD (2009). Talking to learn: why biology students should be talking in classrooms and how to make it happen. CBE Life Sci Educ 8, 89–94.
- University of Virginia Instructional Technology Services (2009). UVa First-Year Student Computer Inventory: Year-to-Year Comparison, 1997–2009. http://its.virginia.edu/students/inventory/compare (accessed 7 September 2015).
- van Vliet EA, Winnips JC, Brouwer N (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. CBE Life Sci Educ 14, ar26.
- Whiteside A, Brooks DC, Walker JD (2010). Making the case for space: three years of empirical research on learning environments. EDUCAUSE Q. www.educause.edu/ero/article/making-case-space-three-years -empirical-research-learning-environments (accessed 22 August 2015).
- Whiteside A, Jorn LA, Duin AH, Fitzgerald S (2009). Using the PAIR-up model to evaluate active learning spaces. EDUCAUSE Q. http://er.educause .edu/articles/2009/3/using-the-pairup-model-to-evaluate-active -learning-spaces (accessed 22 August 2015).