

# A SCALE-UP Mock-Up: Comparison of Student Learning Gains in High- and Low-Tech Active-Learning Environments

Paula A. G. Soneral\* and Sara A. Wyse

Department of Biological Sciences, Bethel University, St. Paul, MN 55112

## ABSTRACT

Student-centered learning environments with upside-down pedagogies (SCALE-UP) are widely implemented at institutions across the country, and learning gains from these classrooms have been well documented. This study investigates the specific design feature(s) of the SCALE-UP classroom most conducive to teaching and learning. Using pilot survey data from instructors and students to prioritize the most salient SCALE-UP classroom features, we created a low-tech “Mock-up” version of this classroom and tested the impact of these features on student learning, attitudes, and satisfaction using a quasi-experimental setup. The same instructor taught two sections of an introductory biology course in the SCALE-UP and Mock-up rooms. Although students in both sections were equivalent in terms of gender, grade point average, incoming ACT, and drop/fail/withdraw rate, the Mock-up classroom enrolled significantly more freshmen. Controlling for class standing, multiple regression modeling revealed no significant differences in exam, in-class, preclass, and Introduction to Molecular and Cellular Biology Concept Inventory scores between the SCALE-UP and Mock-up classrooms. Thematic analysis of student comments highlighted that collaboration and whiteboards enhanced the learning experience, but technology was not important. Student satisfaction and attitudes were comparable. These results suggest that the benefits of a SCALE-UP experience can be achieved at lower cost without technology features.

## INTRODUCTION

College classrooms using learner-centered, active-learning strategies are unequivocally more effective for student learning compared with traditional pedagogies (Freeman *et al.*, 2014). National calls for reform in biology education have emphasized the development of such evidence-based pedagogies to help students develop the concepts and competencies held by practicing scientists (National Research Council, 2003, 2012; American Association for the Advancement of Science, 2011). These pedagogies include, but are not limited to, the use of personal response devices (i.e., clickers; Mazur, 1997; Knight *et al.*, 2013), cooperative learning groups (Johnson *et al.*, 1998), science argumentation (Toulmin, 1958), problem-based learning (Hmelo-Silver, 2004; Savery, 2006), and more recently, the flipped or inverted classroom (Lage *et al.*, 2000; Mazur, 2009). Despite the well-documented learning gains from learner-centered pedagogies (e.g., Prince, 2004; Knight and Wood, 2005; Michael, 2006; Blanchard *et al.*, 2010; Freeman *et al.*, 2014) and widespread awareness of the need for reform, a significant challenge many institutions face in implementing such pedagogies (Bligh, 2000) is the physical layout of the classroom.

Physical space influences instructor teaching philosophy and practices, as well as the student experience (Brooks, 2011). Despite recent studies emphasizing the importance of aligning physical classroom spaces with active-learning pedagogies to achieve greater learning gains (Knaub *et al.*, 2016), learning environments remain poorly designed for active learning. For example, tiered stadium-style lecture halls, chairs and

Peggy Brickman, *Monitoring Editor*

Submitted July 20, 2016; Revised November 7,

2016; Accepted November 30, 2016

CBE Life Sci Educ March 1, 2017 16:ar12

DOI:10.1187/cbe.16-07-0228

\*Address correspondence to: Paula A. G. Soneral (p-soneral@bethel.edu).

© 2017 P. A. G. Soneral and S. A. Wyse. CBE—Life Sciences Education © 2017 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.

desks bolted to floors, and tightly packed seating density make it nearly impossible for faculty to meaningfully interact with students while using learner-centered pedagogy. Furthermore, such spaces create difficulties for students to face one another while engaging in in-class collaborative activities or discussions (Milne, 2006; Oblinger, 2006). Our university classrooms may thus be facilitating an emerging but unintentional paradox: as increasing numbers of students become accustomed to active learning and reap its benefits, their classrooms lock them into a mold not suited for their learner-centered expectations. Likewise, faculty desiring to use evidence-based teaching practices are restricted from using the full range of active-learning strategies in lecture-style physical spaces (Whiteside *et al.*, 2010).

### SCALE-UP Classrooms

Recognizing the need for spaces that foster active instruction, physicists at North Carolina State University and the TEAL Project at MIT designed student-centered learning environment with upside-down pedagogies (SCALE-UP) classrooms (Beichner *et al.*, 2000, 2007; Dori, 2007; <http://scaleup.ncsu.edu>). The overarching concept of the SCALE-UP model is to redesign learning spaces to promote the use of teaching and learning techniques that place students at the center of the learning process—working in groups, applying concepts, solving problems—bringing studio-style learning to larger numbers of students at one time (Knaub *et al.*, 2016). The SCALE-UP classroom model consists of tables, typically round, that enable students to sit with peers facing one another (i.e., pods). These classrooms also feature nearby “team spaces” that contain a large whiteboard or chalkboard writing space. In addition, pods are outfitted with technology that allows students to plug in their laptops and display work on a designated flat TV or computer monitor and with microphones. Technology at the teaching station allows an instructor to display work from an individual pod to the entire class across all screens or enables students to work simultaneously at individual pods. As its namesake implies, the SCALE-UP classroom model can be scaled to suit a variety of classroom sizes and institution types, wherever active learning is used as a pedagogical approach. Not only have these classrooms been extremely popular in large-enrollment physics courses that use learner-centered instructional practices, but they have also been implemented at more than 259 large and midsized institutions nationwide for teaching in a variety of disciplines (Gaffney *et al.*, 2008; Knaub *et al.*, 2016; <http://scaleup.ncsu.edu>).

A growing body of research on the impact of SCALE-UP learning spaces illustrates the manifold positive advantages of this space on both teaching and learning. In large-enrollment scenarios, the SCALE-UP classroom lessened the psychological distance between students and instructor(s), effectively making the classroom seem smaller to students. Instructors adopted teaching methodologies more aligned with a learning “coach” or moderator compared with lecturer, and students increased their engagement, particularly valuing the banquet-style seating, which promotes teamwork and collaboration (Whiteside *et al.*, 2010; Walker *et al.*, 2011; Cotner *et al.*, 2013; Lasry *et al.*, 2014).

In addition to these positive effects on attitudes and classroom behaviors for students and faculty, SCALE-UP classrooms have promoted gains in student learning, achievement, and

retention. Initial studies on learning impacts among physics students demonstrated increased cognitive gains on concept inventories and assessments and reduced failure rate compared with traditional approaches to learning (Hestenes *et al.*, 1992; Dori and Belcher, 2005). In quasi-experimental studies, students in SCALE-UP environments reproducibly earned course grades that exceeded ACT-based model predictions compared with peers enrolled in traditional classrooms (Cotner *et al.*, 2013). In a separate UK-based study controlling for a range of variables, technology-enhanced SCALE-UP classrooms contributed to increased academic performance (Brooks, 2011). Taken together, these findings emphasize that the SCALE-UP classroom as an educational intervention increased student engagement, attitudes, and performance. Over the long term, such collaborative spaces may be especially important as an intervention for underrepresented students and women susceptible to the leaky science, technology, engineering, and mathematics pipeline (Seymour and Hewitt, 1997).

While the SCALE-UP classroom offers positive returns on capital investment in the form of enhanced learning and experiences for students, the cost for construction and maintenance (Erol *et al.*, 2016) and effective pedagogical use of these spaces may be prohibitive for smaller, less-endowed institutions. With rapid changes in technology, some features of these classrooms become outdated very quickly. For example, with the advent of Google Docs, student collaboration is possible without the flat TV or computer monitors at each pod. Furthermore, once these innovative spaces are constructed, institutions must invest in faculty development programming to help instructors adjust their active-learning repertoire to make nuanced and full use of the learning space (Walker *et al.*, 2011; Knaub *et al.*, 2016). In light of serious budget constraints facing higher education, enrollment decline due to the rising cost of tuition, and increased online offerings, institutions must carefully discern the relative advantages of investing in SCALE-UP classrooms and pedagogy relative to costs. Even the creators of these classrooms identify the need to understand whether technology is useful in the design of a SCALE-UP classroom, since it is the most costly component of these learning spaces (Knaub *et al.*, 2016). This is especially true for small liberal arts institutions with operating budgets driven by enrollment-generated revenue streams.

### Study Aims

The purpose of this research study is to explore the efficacy of lower-cost alternatives to the high-tech, expensive, SCALE-UP design. Because a hallmark characteristic of the SCALE-UP model is its emphasis on peer–peer interactions and cooperative learning, we hypothesize that gains associated with these active-learning spaces can be achieved by simulating this salient attribute of the SCALE-UP environment (Cotner *et al.*, 2013). While previous studies have compared learning outcomes and overall initial reactions to the fully equipped SCALE-UP spaces, this study extends previously published studies on the effectiveness of SCALE-UP spaces in several ways. First, we identify specific features of the SCALE-UP classroom space most helpful to students and instructors. Second, using pilot responses to guide prioritizing classroom features, we compare student learning in a low-tech “Mock-up” of the SCALE-UP classroom using a quasi-experimental setup. An experienced instructor used

active-learning pedagogies in back-to-back sections of an introductory biology course, holding variables constant except for the high- and low- tech aspects of the SCALE-UP learning environment. These sections were taught in 1) a newly constructed SCALE-UP classroom with writable walls and collaborative pods; and 2) a low-tech adaptation of the SCALE-UP space, the Mock-up. Third, this design isolated the most expensive technology variable from the SCALE-UP room and tested its impact on student learning. While other researchers have suggested the technology features may not be essential (Knaub *et al.*, 2016), no studies have directly tested the impact of these features. In addition, the setup tested the impact of adapting a standard classroom for cooperative learning in order to mimic the SCALE-UP experience.

Three core questions guided this study: 1) Which component(s) of the SCALE-UP classroom—whiteboard space, seating arrangement, or technology—do students and faculty view as most helpful to their learning or teaching? 2) What is the impact of high-tech versus low-tech SCALE-UP classroom environments on student conceptual understanding and scientific reasoning skills? 3) Does a Mock-up of the most salient low-tech features of the SCALE-UP classroom yield comparable learning gains and overall student experience?

The study was designed and executed in two phases (Figure 1). In the first phase, we collected data from students and faculty to capture their initial reactions to the SCALE-UP classroom in its inaugural semester through focus groups, open-ended surveys, and forced-response surveys. In the second phase of the study, we used the SCALE-UP classroom features identified from phase 1 to design a low-tech, low-cost Mock-up of the classroom environment. Using this quasi-experimental approach, we compared student learning, attitudes, and overall

satisfaction in the high-tech SCALE-UP and low-tech Mock-up environments. These findings have the potential to help universities reconfigure existing classrooms into learning spaces that are centered around students but equally effective for achieving desired learning gains without the exorbitant price tag associated with building high-tech SCALE-UP spaces.

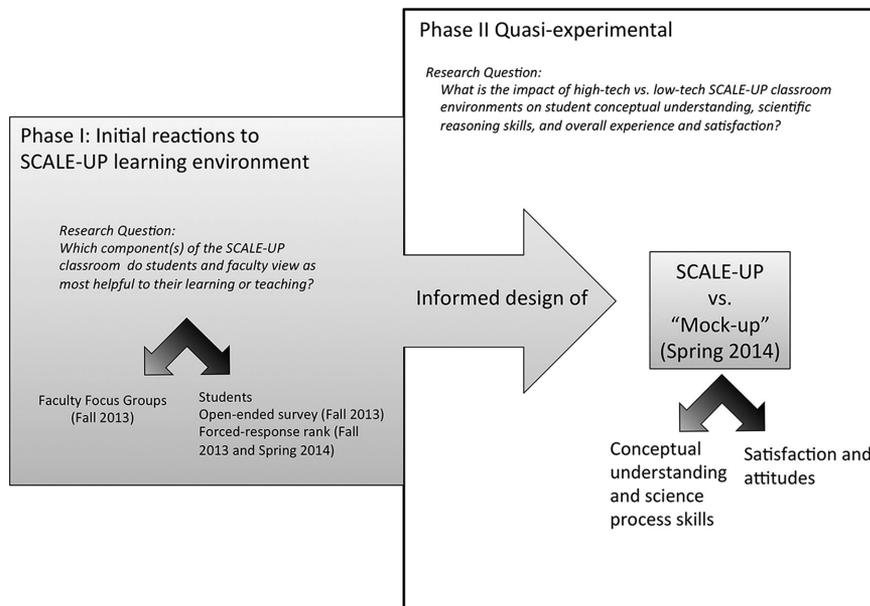
## METHODS

### Phase 1. Initial Reactions to the SCALE-UP Learning Environment

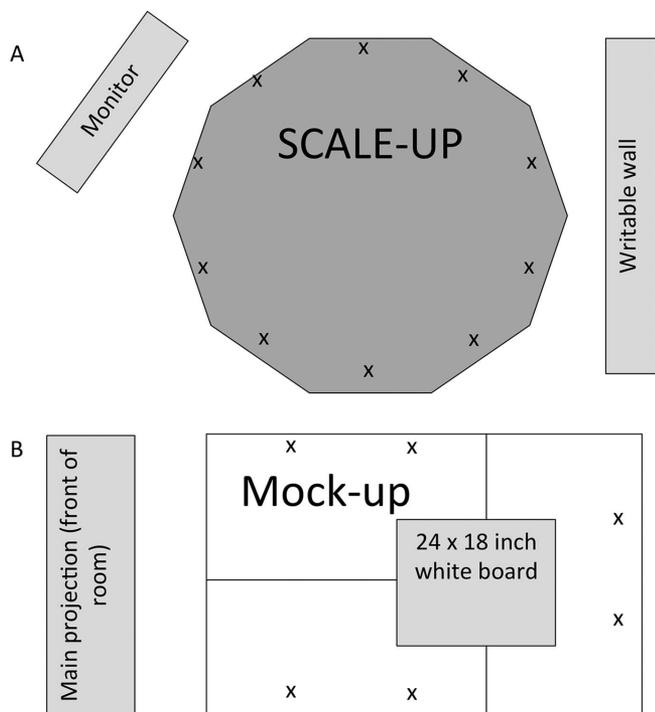
The high-tech SCALE-UP room was constructed during the Summer of 2013 and had capacity for 45 students. Five circular table pods each seated nine students, had dedicated wall-mounted monitors, one electronic plug-in for projection of a laptop computer, and volume control capabilities centralized to each pod. Additionally, the SCALE-UP room was equipped with floor-to-ceiling writable wall space and colored markers and a main projection screen at the front of the room (Figure 2A). A sixth pod was the instructor base station, which controlled the overall activation of monitors in the classroom—from individual displays at each pod to common displays from a single computer. Because the room was equipped with six monitors in total, students sitting at any seat in the space had multiple sight-line options for viewing projected content. The space allowed collaboration and visible display of student work both from a computer and from written work on walls.

To capture initial reactions to the classroom space during its inaugural semester (Fall 2013), we invited faculty and students teaching or enrolled in classes in the newly constructed SCALE-UP classroom to participate in our study. Eight faculty (of a possible 10) teaching in the SCALE-UP classroom for the first time participated in a focus group reflecting on their teaching experiences. These faculty came from a variety of disciplines—biology ( $n = 2$ ), social work, education ( $n = 2$ ), chemistry, communication, and business—and were asked, “Which component(s) of the SCALE-UP classroom helped your teaching?” Faculty comments were transcribed during the focus group and coded for themes (Glaser and Strauss, 1967).

At the conclusion of their first experience in the SCALE-UP space (Fall 2013 and Spring 2014), students enrolled in courses in the SCALE-UP room ( $n = 117$ ) were likewise asked, “Which component(s) of the SCALE-UP classroom helped your learning?” in an end-of-semester survey. These students consisted of 100- and 300-level students enrolled in two different 100-level biology courses and one biology-related general education course at the 300-level. Eighty-three responses (71% response rate) were coded by theme (Glaser and Strauss, 1967) into categories: writable space, seating arrangement (including pods), multiple sight lines for viewing monitors, or technology-related (e.g., plugging in to display work to the larger group). Frequencies for each



**FIGURE 1. Study design.** The first phase of the study was designed to capture initial reactions to the SCALE-UP learning environment using faculty focus groups and open-ended surveys to students. The responses informed the design of a quasi-experimental phase 2 study in which student learning, satisfaction, and attitudes were compared for the SCALE-UP and Mock-up environments.



**FIGURE 2.** Schematic representation of high-tech SCALE-UP and low-tech Mock-up classrooms. (A) Model of pod configuration in the SCALE-UP classroom. Nine students (X) are seated in a circular arrangement with a single dedicated monitor per pod. Additional monitors (not shown) are located throughout the classroom to create multiple sight lines. Writable wall space is located within 2–3 feet of each pod. (B) In the Mock-up classroom, tables were pushed together into pods of six students. All students view a single line-of-sight projection screen located at the front of the room. Each pod was equipped with 24 × 18 inch whiteboard, markers, and access to LiteShow software that enabled the sharing of work from an individual computer.

category were recorded. Categories that emerged from open-ended responses were subsequently used to generate a forced-response survey wherein students were asked to rank the relative utility of each classroom feature. This survey was administered to additional biology students enrolled in a SCALE-UP classroom in Spring 2014 ( $n = 37$ ).

**Phase 2. Quasi-experimental SCALE-UP Mock-Up: Comparison of Student Learning and Satisfaction**

Informed by the faculty focus group and student survey data from phase 1, we created a low-tech mock SCALE-UP classroom

using the most valuable low-tech features of the SCALE-UP classroom. We asked, What is the impact of the high-tech SCALE-UP versus the low-tech Mock-up room on student conceptual understanding, attitudes toward biology, and overall course satisfaction?

A standard laboratory space with movable rectangular tables was converted into an active-learning classroom, hereafter referred to as “Mock-up.” Individual tables were arranged into a T configuration, seating six students per pod (Figure 2B). Each pod was equipped with a 24 × 18 inch portable whiteboard that students could place on the tabletop and markers for visual display of collaborative work. The room had a large chalkboard along the front wall of the classroom and a projection screen that could be superimposed on this front wall. The instructor base station was located at the front of the room, near the main projection screen. Students used a single line of sight for viewing projected content (front of room; Figure 2B). The maximum capacity of the low-tech Mock-up room was 20 students, and the section enrollment for this study was 17. Although the Mock-up room did not use multiple monitors, low-cost software was used for large-group shared projection. LiteShow software ([www.infocus.com/peripherals/INLITESHOW4](http://www.infocus.com/peripherals/INLITESHOW4)) enabled computers to connect remotely to the main classroom projector via an IP address. The instructor controlled the display of student work to the class’s main projection system. A maximum of four individual computer screens were shared simultaneously to the main projection system. Students utilized Google Docs for collaborative purposes, rather than collaboratively working from one large mounted monitor and the associated student computer on display.

Although the majority of the SCALE-UP features were replicated, some noteworthy differences exist between the two spaces (Table 1). Pod size consisted of nine students in circular arrangement in the SCALE-UP room and six students in a T configuration in the Mock-up room (Figure 2). The circular arrangement of seating in the SCALE-UP classroom required multiple projection sight lines, whereas the Mock-up used a single line of sight to the front of the room. Students sharing ideas on walls were typically standing 2–3 feet away from their main pod table, whereas students using the portable whiteboards were seated during collaborative ideation and group work. Finally, the combination of Google Docs and LiteShow software projected student work to a main projection screen, rather than to dedicated pod monitors.

**Course Description**

The second phase of this research focuses on one of three introductory biology courses offered at a liberal arts college in the upper Midwest. The course of interest, Introduction to

**TABLE 1.** Experimental classroom layout

Classroom	Main projector	Sight lines(s)	Seating arrangement	Whiteboard space	Sharing capability
High-tech SCALE-UP	Yes	Multiple	Five pods of nine students; circular	Writable walls from floor to ceiling	Pod technology enables individual computers to plug in to mounted monitors to be shared with pod and/or entire class.
Low-tech Mock-up	Yes	Single	Four pods of six-students; T configuration	One large portable whiteboard at the center of each pod	LiteShow enables individual computers to share with the entire class via front projection screen only.

Molecular and Cellular Biology, requires one semester of chemistry as a co- or prerequisite and is typically taken during the second or third semester from the start of college by students majoring in biology, biochemistry/molecular biology, and a biology-related major, biokinetics.

Introduction to Molecular and Cellular Biology was selected for this study for two reasons: 1) the instructor had previous experience teaching in a SCALE-UP classroom, and 2) unusually high enrollments exceeded the capacity of the SCALE-UP classroom (maximum 45 students) and forced the creation of a third section offered in an alternative space. Thus, this study design uses a quasi-experimental approach, in which the same instructor taught an identical course, allowing the majority of confounding variables to be controlled as reasonably as possible. The sections were taught in 70-minute modules, back-to-back in the early morning on a Monday–Wednesday–Friday schedule, with 160-minute labs taught back-to-back on Thursdays. The first section was taught in a high-tech SCALE-UP classroom, and the second section in the converted biology laboratory (low-tech Mock-up). This course enrolled a total of 58 students in all sections in the spring of 2014 ( $n = 41$  in SCALE-UP and  $n = 17$  in Mock-up). We were unable to randomly assign students to the SCALE-UP and Mock-up sections. However, to minimize the potential confounds introduced by differences in instruction on student learning, the course used the same syllabus, instructor, course policies, assessments, learning resources, lesson plans, pedagogy (e.g., active, learner-centered, cooperative, daily use of writing on walls), and technology use (e.g., plugging in and using Google Docs one time per unit). In addition, we approximated the randomization of students in the section by evaluating equivalency in terms of grade point average (GPA), ACT scores, class standing, effort, and attitude. The university institutional review board approved all protocols used for collecting and analyzing student data, and informed consent was obtained for all volunteer participants in this study.

### Assessment of Student Learning

Student learning goals and objectives were broadly categorized into conceptual learning and science process skills. A summary of the learning goals for each category, the aligned summative assessments, and implementation are summarized in Table 2. Individual class and laboratory sessions were marked and

scaffolded by a number of specific learning objectives and formative assessments to help students achieve the larger course milestones reported here. The course used an upside-down or flipped structure, with readings assigned before class or lab and online activities targeting low-level Bloom's cognitive skills (Anderson *et al.*, 2001) completed before the class session on that topic. Class and lab sessions were devoted to group interactions on higher-order cognitive activities (Bloom's 3–6)

Summative unit assessments were identical in both sections and consisted of four to six multipart open-response questions that spanned all levels of Bloom's taxonomy, characterized using a weighted Bloom's index. Two raters independently assigned a Bloom's level (1–5, where 1 = knowledge, 2 = comprehension, 3 = analysis, 4 = application, 5 = synthesis and evaluation) to each question prompt for items on five unit assessments (i.e., 23 items total for five unit assessments). The agreement between the raters was substantial (Cohen's kappa = 0.76); therefore, these scores were averaged between the raters. To normalize the relative point differences associated with low and high cognitive levels (higher Bloom's levels were typically associated with more points), we computed a weighted Bloom's score for each assessment item (Momsen *et al.*, 2010):

$$\sum_{i=1}^n \beta(\omega/\gamma)$$

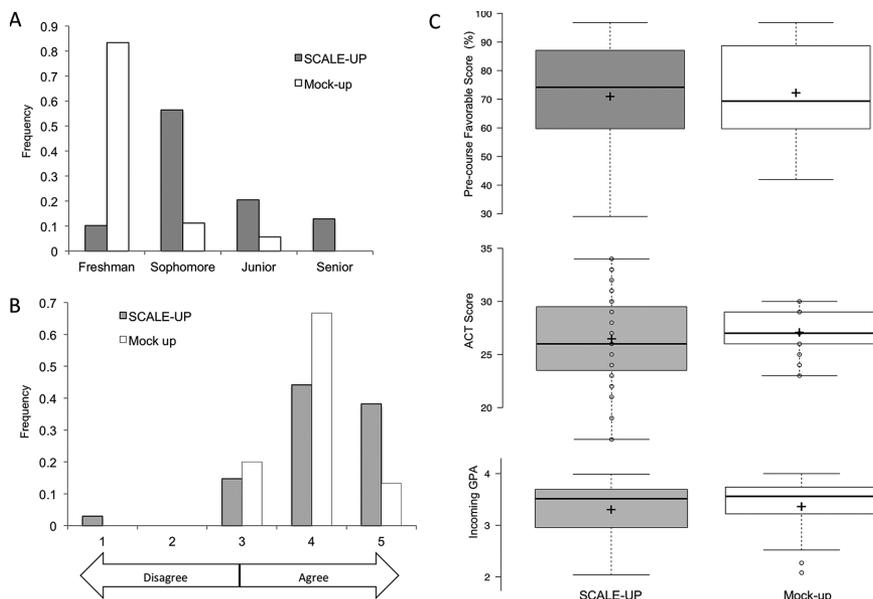
where the Bloom's index for each unit assessment represents the sum of the weighted average for each item ( $\beta$  = average Bloom score for an assessment item,  $\omega$  = the points associated with the item, and  $\gamma$  = the total number of points for the assessment).

### Measurements of Equivalency for Student Populations

Equivalency of student populations for SCALE-UP and Mock-up sections was evaluated in terms of college preparation and prior academic achievement using incoming ACT scores, college GPA, class standing, and major. Student incoming attitudes toward the discipline of biology were measured using the Colorado Learning Attitudes and Science Survey at the beginning of the semester (CLASS-BIO; Semsar *et al.*, 2011). The degree to which students put forth effort in the course was measured using a line item on the IDEA survey instrument, "As a rule, I

TABLE 2. Summary of learning objectives, measurements, and implementation approaches

Category	Goals	Assessment and implementation
Conceptual mastery	<ol style="list-style-type: none"> <li>1. Explain foundational concepts and theories</li> <li>2. Apply knowledge to solve problems and case studies</li> </ol>	Preclass assignments: weekly IMCA Concept Inventory: beginning and end of semester In-class assessments: daily Unit assessments: five per semester
Scientific reasoning and process skills	<ol style="list-style-type: none"> <li>1. Contextualize research and experimental goals</li> <li>2. Make observations; formulate and test hypotheses</li> <li>3. Design and execute experiments; organize data</li> <li>4. Analyze and interpret experimental data; apply quantitative reasoning and statistics</li> <li>5. Draw conclusions from data; construct scientific arguments and models</li> <li>6. Communicate arguments orally and in writing</li> <li>7. Work collaboratively as a member of a team</li> </ol>	Prelab assignments: weekly Laboratory notebook (rubric): weekly Oral presentation (rubric): two per semester Poster presentation (rubric): end of semester



**FIGURE 3. Comparison of equivalence for student populations in SCALE-UP and Mock-up classrooms. (A) Bar chart showing the relative distribution of class standing in the SCALE-UP and Mock-up classrooms. There were significantly more freshmen in the Mock-up classroom (Mann-Whitney  $U$ ,  $p < 0.05$ ). (B) Summary distribution of response to survey item “As a rule, I put forth more effort in study relative to my peers” (1 = strongly disagree; 5 = strongly agree). The shapes of the distributions for SCALE-UP and Mock-up classes were similar, and effort was comparable (Mann-Whitney  $U$ ,  $p > 0.05$ ). (C) Box-plot summary of favorable score measures from the precourse CLASS-BIO survey of attitudes in biology (top), ACT score (middle), and college GPA (bottom). Center lines indicate median; + indicates mean; box limits indicate the 25th and 75th percentiles. Students in both sections had similarly favorable attitudes about biology and comparable ACT scores and GPAs (Student’s  $t$  and Mann-Whitney  $U$ , all  $p > 0.05$ )**

put forth more effort in study relative to my peers,” implemented at the conclusion of the course (IDEA Center: <http://ideaedu.org>).

In both the SCALE-UP and Mock-up sections of the course, all students were retained (i.e., no withdrawals), and there were no earned grades in the “D” to “F” category (unpublished data). The Mock-up section enrolled significantly more freshmen (83%) compared with the SCALE-UP classroom (5% freshman). In contrast, the SCALE-UP classroom was predominantly sophomore students (57%) and upperclassmen (Figure 3A; Mann-Whitney  $U$ ,  $p < 0.05$ ).

Despite differences in class standing, students in both sections of the course reported applying comparable levels of effort to college course work. The frequency distribution of responses to the end-of-semester survey question “As a rule I put forth more effort in study relative to my peers” showed an overall similar shape (Figure 3B; Mann-Whitney  $U$ ,  $p > 0.05$ ). At the beginning of the course, students in both sections had similarly favorable attitudes about biology based on responses to the CLASS-BIO instrument (Figure 3C, top; SCALE-UP =  $70.9 \pm 2.9$  [mean  $\pm$  SE], Mock-up =  $72.3 \pm 5.1$ ; Student’s  $t = 0.23$ ;  $p = 0.41$ ). In addition to similar effort and attitudes, incoming ACT scores and GPA were comparable for both sections of the course (Figure 3C, middle and bottom; ACT =  $26.5 \pm 0.71$  and  $27.1 \pm 0.56$  for SCALE-UP and Mock-up, respectively;  $t$  test  $p = 0.30$ ; GPA =  $3.37 \pm 0.08$  and  $3.33 \pm 0.15$  for

SCALE-up and Mock-up, respectively,  $t$  test  $p = 0.34$ ).

In summary, student populations were comparable in terms of college preparation, prior academic achievement, attitudes toward the discipline of biology, and effort in college course work. However, because the students enrolled in the SCALE-UP classroom had more college-level learning experience compared with those in the Mock-up environment, these differences in class standing were used as covariates when performing regression analyses for comparison of learning gains and student satisfaction. In addition, we considered the impact of class standing in the interpretation of data described herein.

### Analysis of Learning Gains Using Multivariate Regression Analysis

Direct comparisons of student learning were made using regression analysis, coded in R (R Core Development Team, 2013), between students enrolled in the SCALE-UP versus Mock-up sections and controlled for class standing as a covariate. Assessments of conceptual student learning used as outcome variables included: unit exams (five total), in-class assignments (including clicker responses, modeling, construction and labeling of figures), and preclass homework activities. Further, comparisons of student con-

ceptual understanding were made using the Introduction to Molecular and Cellular Biology (IMCA) Concept Inventory (Shi *et al.*, 2010). In-class student artifacts were procured by means of photographing wall or whiteboard work or scanning paper-based documents. Differences in student scores not explained by class standing differences as a covariate were attributed to differences in the SCALE-UP versus Mock-up learning environment.

We used three linear regression models to explore the variables necessary and sufficient for predicting learning gains in both classrooms. We modeled assessment percentage points as a continuous outcome response and included input covariables as follows: GPA (continuous), biology major (binary factor: 0 = no, 1 = yes), enrollment in classroom (binary factor: SCALE-UP = 1, Mock-up = 0), and class standing (a series of dummy variables [1 = freshman, 2 = sophomore, 3 = junior, 4 = senior] as covariables; Table 3). The R code excluded the dummy variable for being a freshman, while including the variables for being a sophomore, junior, or senior. Thus, these variables were interpreted in relationship to the intercept and show the mean differences for each class in relationship to being a freshman. Our simplest model tested the hypothesis that enrollment in a SCALE-UP classroom would increase assessment performance and treated the classroom as a single binary explanatory variable (model set 3). Model set 2 tests the impact of enrollment in the SCALE-UP classroom and class standing on assessment

**TABLE 3. Regression models used to determine whether the SCALE-UP or Mock-up environments contributed to learning gains**

Base model: conceptual mastery influenced by the SCALE-UP classroom	Outcome ~ SCALE-UP
Model 1: Impact of SCALE-UP classroom on student performance varies with class standing, GPA, being a biology major.	Outcome ~ ClassStanding + GPA + BioMajor + SCALE-UP
Model 2: Impact of SCALE-UP classroom on student performance varies with class standing.	Outcome ~ ClassStanding + SCALE-UP
Model 3: Impact of SCALE-UP classroom on student performance depends only on classroom.	Outcome ~ SCALE-UP

performance. Model set 1 is the most complex and uses GPA and major as additional covariates (Table 3). For completeness, we ran 23 models in total and calculated regression coefficients, SE, and *p* values.

#### Analysis of Student Satisfaction Using Logistic Regression Models

We compared SCALE-UP and Mock-up responses to end-of-semester survey questions related to student satisfaction. Students rated the extent to which they agreed with statements “Overall, I rate this course as excellent” and “I have more positive feelings about the subject,” and to evaluate equivalency, “As a rule, I put forth more effort than my peers.” Because survey responses were ordered categorical responses in a bounded scale (1 = strongly disagree; 5 = strongly agree), we used logistic regression models in which a dichotomous outcome is predicted by one or more proportional variables (Venables and Ripley, 2002). In this case, our explanatory variable was SCALE-UP (0) or Mock-up (1), and the outcome variable was agreement with the given statement (i.e., electing 4 or 5 on the Likert scale; Antoine and Harrell, 2000). Coefficients of all prediction formulas (with standard errors and significance levels) and odds ratios (with confidence intervals) were calculated using the MASS package in R (R Core Development Team, 2013). Thus, the model used for each question was: Survey

response (agree) = Intercept +  $\beta$ \*classroom type. We were not able to use class standing as a predictor, because these data were deidentified and unmatched.

#### Qualitative Analysis of Open-Ended Student Comments

To assess student satisfaction in the low- and high-tech SCALE-UP classrooms, we administered an end-of-semester survey to students enrolled in both sections of the course with the prompt “Would you take another course in a SCALE-UP/Mock-up room, why/why not?” Qualitative coding methods (Bogdan and Biklen, 1998) were implemented to discern patterns in the data set. Emerging themes were summarized into a coding rubric (Table 4) that was subsequently used by two trained raters with a high interrater reliability (Cohen’s kappa = 0.93; Glaser and Strauss, 1967) to code each transcription. Areas of disagreement were discussed until consensus was reached; 54 responses (93% response rate) were analyzed. We also coded comments from the end-of-semester course evaluations (IDEA Center: <http://ideaedu.org>) into categories encompassing positive and negative comments about the instructor, class, and classroom. The frequency for each response code was tabulated, and the Chi-square test was used to compare the patterns in response distributions.

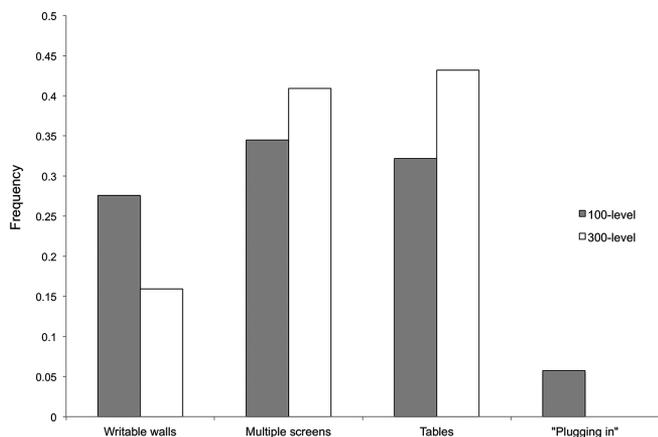
## RESULTS

### Phase 1. Effectiveness of Individual Features of the SCALE-UP Classroom

To gauge initial reactions to teaching in the new SCALE-UP space, focus groups with faculty participants identified three thematic attributes of the classroom space that helped their teaching: 1) collaborative table groups, 2) writable wall space, and 3) the ability for students to display their work using technology. Faculty reported that the writable walls and ability to plug in provided accountability for students to stay on task and make their thinking and learning visible and allowed students to participate in collaborative learning. Faculty also appreciated these formats for visualizing learning as an easy way to spot and address incomplete conceptions of course concepts. The pods increased student engagement and facilitated instructor access to students so that individualized formative feedback could be provided. Faculty noted that the layout of the room increased their perceived pressure to have the students “doing

**TABLE 4. Coding rubric used for analysis of open-ended student comments “Would you take another course in a SCALE-UP classroom again? Why or why not?”**

Yes	No
1. Collaborative learning (e.g., discussion, groups, pods, learning together)	1. Sight lines did not work (e.g., student could not see or multiple screens were a distraction)
2. Visualize learning (e.g., fosters active engagement, writing on walls or board to display thinking)	2. Dispersal of power (e.g., teacher not the center of the space)
3. Multiple sight lines (e.g., easier viewing for visual learners, great for viewing technology)	3. Slow pace (e.g., lecture learning “covers” more material)
4. Community and atmosphere (e.g., the “feel” of the course—smaller or “laid back”)	
5. Interaction and hands-on nature (e.g., manipulating objects and models at the table)	
6. Variety (e.g., change of pace, different from norm)	
7. Easier learning (e.g., retain concepts longer)	
8. Style of learning (e.g., conducive to personal learning preference, enjoyment of active-learning approaches)	



**FIGURE 4.** Content analysis distributions for 100- and 300-level students to open-ended question about SCALE-UP classroom. The 100-level students report the multiple screens, tables, writable walls, and technology (plugging in) as helpful to their learning;  $n = 83$ ; data from Fall 2013. The 300-level students reported table groups and multiple screens as most helpful, followed by writable walls. In contrast, plugging in was not cited as helpful ( $n = 34$ ; data from Fall 2013).

something” and for “me to stop talking,” and adjusted their practices accordingly. For several faculty members, these changes invigorated their teaching, while for others the lack of a clear “front” coupled with viewing the backs of a large proportion of students was at first disconcerting but eventually manageable.

Student responses to an open-ended survey administered at the conclusion of the semester asking “Which component(s) of the SCALE-UP classroom helped your learning?” indicated attributes that were similar to those noted by faculty, notably writable walls, table groups (pods), and the ability to use technology (i.e., plugging in) to share their work. Interestingly, students identified an additional feature not commented on by faculty—multiple sight lines—as being helpful to their learning. Multiple sight lines and pod tables, followed by writable walls, were the most frequently cited helpful attributes of the classroom for 100-level students in open-response surveys (Figure 4). These trends were similar in subsequent forced-response surveys ( $n = 37$ ; Spring 2014), in which 100-level students ranked features in order of most (1) to least (4) helpful (Table 5, phase 1). In these forced-response surveys, multiple sight lines

likewise ranked highest, followed by writable walls and table pods.

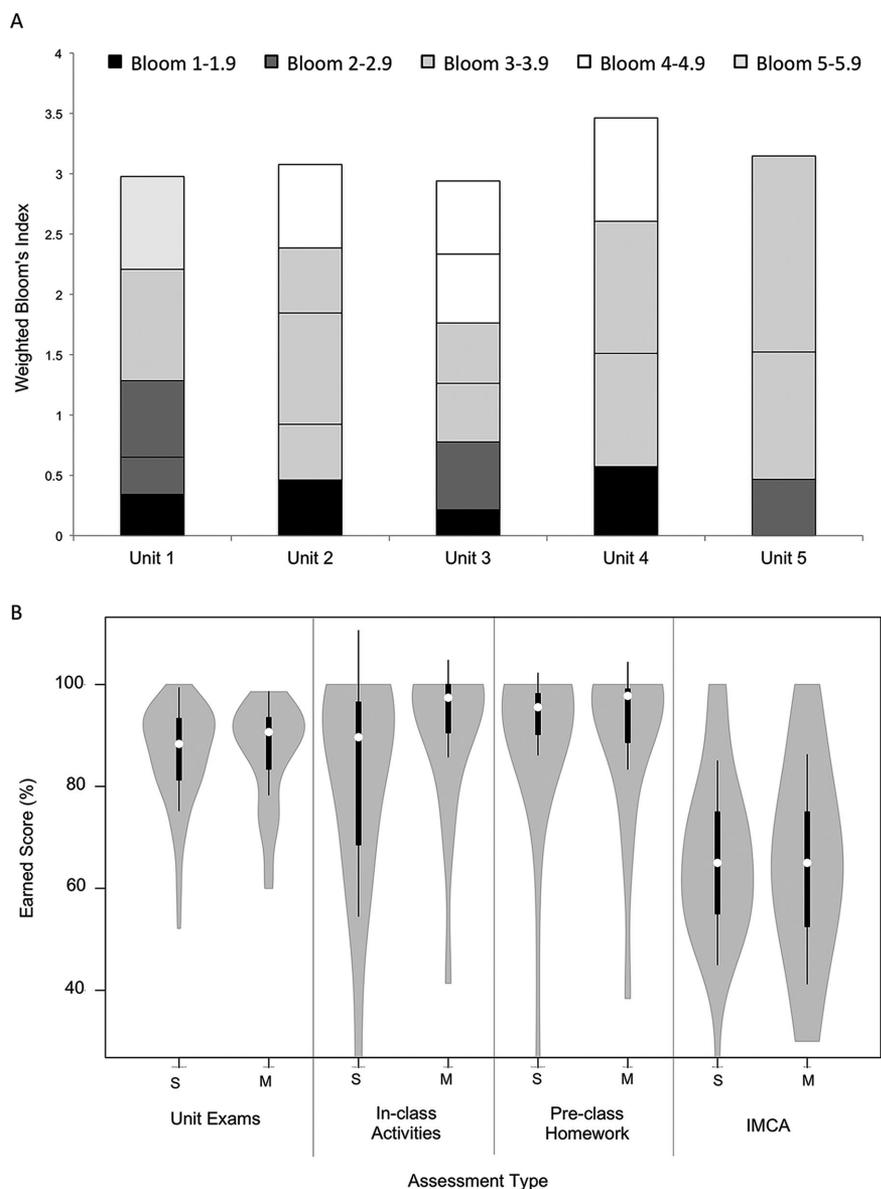
When 300-level students were asked the identical series of questions, subtle differences consistent with developmental learning emerged. For example, upper-level students valued table groups and multiple sight lines the most, placing less value on the writable wall space (and concomitant instructor formative feedback) compared with 100-level students. Interestingly, both groups of students ranked the technological feature of plugging in as the least helpful attribute of the space (Table 5). Notably, this least-valued feature of the classroom is the most expensive to construct. Taken together, these survey data from faculty and 100- and 300-level learners suggest that the most salient attributes of the SCALE-UP classroom are those that foster collaboration (pods), ease of viewing content (multiple screens, lines of sight), and peer and instructor feedback and formative assessment (writable walls). The expensive high-tech feature of plugging in was the least valuable for student learners.

### Phase 2. Quasi-experimental Comparison of Student Learning in SCALE-UP and Mock-Up Environments

Informed by the phase 1 faculty and student perceptions of the most valuable attributes of the SCALE-UP classroom—pod seating and collaborative writing space, *not* plugging in—we tested the relative impact of these high- and low-tech classroom features in a quasi-experimental design (Figure 2 and Table 1). First, to characterize the relationship between the Bloom’s level of assessment items relative to student performance in the SCALE-UP versus Mock-up environments, unit exams administered throughout the course of the semester were evaluated using a weighted Bloom’s index. Unit assessments were case based and open response, containing four to six multipart questions spanning all levels of Bloom’s taxonomy. Overall, the cognitive loads for the assessments were consistent and converged to the mid-cognitive range of Bloom’s 3–4 (Figure 5A). There was no significant difference in exam performance for low- and high-order Bloom’s items, and the overall distributions and averages for unit exams were comparable (Figure 5B; 84.8% SCALE-UP and 86.6% Mock-up; Student’s  $t = 1.6$ ;  $p = 0.06$ ). Similarly, performance on daily in-class activities spanning a range of Bloom’s categories was also similar for each section (Figure 5B; Student’s  $t = -1.4$ ;  $p = 0.07$ ), as was performance on preclass assignments targeting Bloom’s 1–2 knowledge and comprehension (Figure 5B; Student’s  $t = -0.17$ ;  $p = 0.43$ ). Finally, scores from the postinstruction IMCA Concept Inventory

**TABLE 5.** Summary of forced-response survey student rankings of SCALE-UP and Mock-up classroom features for phases 1 and 2

Classroom features	Phase 1 SCALE-UP student level			Phase 2			
	100	300	Mean	Mode	Rank	SCALE-UP rank	Mock-up rank
Plugging in	100	300	$3.2 \pm 1.0$	4	4	4	4
			$3.2 \pm 1.1$	4	4		
Walls	100	300	$1.86 \pm 1.1$	1	2	2	2
			$2.4 \pm 1.3$	2	3		
Multiple screens	100	300	$1.7 \pm 1.1$	1	1	1	3
			$2.8 \pm 1.3$	2	2		
Tables (pods)	100	300	$1.97 \pm 1.2$	1	3	3	1
			$2.4 \pm 1.5$	1	1		



**FIGURE 5.** Assessment and concept mastery in SCALE-UP and Mock-up classrooms. (A) Distribution of weighted Bloom's index for unit-assessment items. The height of each bar represents the computed weighted Bloom's index for each unit assessment; boxes represent the relative proportion of points associated with each question item. Assessment items spanned all levels of Bloom's taxonomy, and the average Bloom's ranking for individual unit assessments overall was  $3.0 \pm 0.9$  (light gray and white). (B) Violin-plot summary of assessment performance in SCALE-UP (S) and Mock-up (M) classrooms. White circles indicate medians; box limits indicate the 25th and 75th percentiles; polygons represent density estimates of data and extend to extreme values. Score distributions for all assessments are comparable for the SCALE-UP and Mock-up learning environments (see also Table 6 for regression coefficients and significance).

(Shi *et al.*, 2010) also revealed comparable distributions (Figure 5B; Student's  $t = 0.18$ ;  $p = 0.43$ ).

To test whether the high- and low- tech classrooms predicted student learning while analytically controlling for potentially confounding variables (e.g., class standing; Figure 3A), we ran a series of multivariate linear regression models (see Table 3 for summary of input and output variables). Regression coefficients

(with SE) and significance are summarized in Table 6. Model set 1, which contained the largest number of covariates (input class standing, GPA, biology major, enrollment in high-tech SCALE-UP) produced the most robust  $R^2$  values ( $p < 0.001$ ), explaining the variance in the data sets to a greater extent compared with the models containing fewer variables. Models 2 and 3 had small  $R^2$  values, signifying a poor fit to the data but nonetheless serving as useful reductionist control models. These models, particularly the most robust (model 1), indicate that the coefficients for assessment performance were not statistically different between the SCALE-UP and Mock-up environments. Unsurprisingly, the only statistically significant input variable was incoming college GPA in model 1 for all assessments except the IMCA, indicating that excellent performance on previous college course work is associated with earning higher scores for all course assessments. Previous GPA, however, did not significantly correlate with performance on the IMCA Concept Inventory, perhaps due to lack of direct alignment of certain items with course content and pedagogy. Although class standing differed in the SCALE-UP versus Mock-up classrooms, this variable was not a significant predictor of assessment performance—students in both classrooms performed equally well on assessments of learning, both high and low stakes, and the IMCA. Taken together, these data illustrate that the SCALE-UP and Mock-up environments contributed to equal learning gains.

### Student Attitudes and Satisfaction in the SCALE-UP and Mock-Up Environments

Because scientific collaboration is an essential attribute of the discipline of biology, and the learners in this study highly valued collaborative pods and writing spaces for their learning (Figure 4), we measured the impact of the SCALE-UP versus Mock-up space on student attitudes toward the discipline of biology using the CLASS-BIO survey instrument (Semsar *et al.*, 2011). Students in both the SCALE-UP and Mock-up learning environments experienced a significant progression from novice to expert-like thinking (Figure 6, A and B; Mann-Whitney  $U$ ,  $p = 0.03$  for SCALE-UP and  $p = 0.04$  for Mock-up). The overall precourse scores were the same between the SCALE-UP and Mock-up classes (Figure 3C), as were the postcourse scores overall (Figure 6, A and B;  $t$  test  $p = 0.37$ ). Favorable shifts parsed and evaluated in terms of individual factor loadings demonstrated similar trends. Interestingly,

**TABLE 6. Regression coefficients for multiple regression models testing relationships among input variables (GPA, class standing, major, and enrollment in SCALE-UP or Mock-up environment) with performance on assessments<sup>a</sup>**

Output	Unit 1 $\beta$ (SE)	Unit 2 $\beta$ (SE)	Unit 3 $\beta$ (SE)	Unit 4 $\beta$ (SE)	Unit 5 $\beta$ (SE)	IMCA $\beta$ (SE)	Pre-HW $\beta$ (SE)	Class $\beta$ (SE)
<b>Model set 1</b>								
Intercept	49.0 (4.2)	51.8 (4.8)	43.3 (6.1)	36.0 (7.0)	58.4 (5.4)	37.8 (15.5)	25.7 (14.5)	34.8 (12.5)
<i>p</i> Value	1.2 e-15	1.2e-14	5.0e-9	4.2e-6	1.2e-14	0.02	0.08	0.008
Sophomore	0.58 (2.1)	-2.2 (2.3)	-4.3 (3.0)	-5.2 (3.4)	-4.0 (2.6)	-9.1 (7.7)	11.4 (7.0)	-0.6 (6.1)
Junior	2.1 (2.4)	-4.2 (2.7)	-2.8 (3.5)	-7.1 (3.9)	-3.9 (3.1)	-11.2 (8.9)	14.6 (8.2)	-0.97 (7.1)
Senior	-0.7 (3.2)	-1.4 (3.6)	1.5 (4.6)	-7.1 (5.2)	-0.6 (4.0)	-0.8 (11.3)	11.5 (10.8)	-7.3 (9.4)
GPA	<b>11.4 (1.2)</b>	<b>10.3 (1.4)</b>	<b>12.6 (1.8)</b>	<b>15.8 (2.0)</b>	<b>8.5 (1.6)</b>	<b>7.6 (4.6)</b>	<b>18.4 (4.2)</b>	<b>16.2 (3.6)</b>
Bio major	1.8 (1.4)	3.2 (1.6)	2.3 (2.0)	-1.3 (2.3)	2.8 (1.8)	8.2 (5.6)	2.6 (4.8)	4.9 (4.2)
SCALE-UP	1.4 (2.0)	0.5 (2.3)	0.82 (2.9)	0.4 (0.69)	1.2 (0.2)	5.4 (7.7)	-9.2 (6.9)	-5.0 (6.0)
Multiple <i>R</i> <sup>2</sup>	0.6	0.59	0.57	0.58	0.45	0.17	0.35	0.35
<i>F</i> ( <i>df</i> = 6)	16.3	12.0	11.1	11.6	6.7	1.4	4.4	4.6
<i>p</i> Value	2.6e-10	2.5e-8	7.8e-8	4.4e-8	3.0e-5	0.22	0.001	0.00091
<b>Model set 2</b>								
Intercept	87.6 (2.0)	87.3 (2.0)	85.5 (2.6)	86.3 (3.0)	86.9 (2.1)	66.1 (4.8)	88.0 (4.7)	91.1 (4.2)
<i>p</i> Value	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16
Sophomore	2.6 (3.5)	-0.5 (3.5)	-1.9 (4.4)	-2.1 (5.1)	-2.1 (3.7)	-6.1 (8.4)	14.4 (8.1)	2.1 (7.2)
Junior	3.4 (4.1)	-3.5 (4.1)	-1.3 (5.1)	-4.3 (5.9)	-2.9 (4.3)	-10.6 (9.6)	16.3 (9.4)	0.2 (8.3)
Senior	5.5 (5.0)	3.1 (5.0)	8.4 (6.2)	4.3 (7.2)	3.4 (5.2)	-0.4 (11.2)	21.5 (11.5)	-0.1 (10.2)
SCALE-UP	-0.51 (3.4)	-1.2 (3.4)	-1.0 (4.3)	-1.2 (5.0)	2.4 (3.6)	5.3 (8.2)	-12.7 (7.9)	-7.9 (7.0)
Multiple <i>R</i> <sup>2</sup>	0.04	0.05	0.07	0.04	0.05	0.04	0.08	0.04
<i>F</i> ( <i>df</i> = 6)	0.5	0.65	1.1	0.6	0.63	0.42	1.2	0.56
<i>p</i> Value	0.74	0.63	0.38	0.7	0.63	0.79	0.34	0.69
<b>Model set 3</b>								
Intercept	88.1 (1.9)	87.1 (1.9)	85.2 (2.5)	85.8 (2.8)	86.5 (2.0)	65.0 (4.6)	90.5 (4.6)	91.4 (3.9)
<i>p</i> Value	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16
SCALE-UP	1.9 (2.3)	-1.6 (2.3)	-1.0 (3.0)	-2.2 (3.4)	1.4 (2.4)	1.0 (5.5)	-1.0 (5.5)	-6.9 (4.7)
Multiple <i>R</i> <sup>2</sup>	0.01	0.008	0.002	0.008	0.006	0.0007	0.0005	0.04
<i>F</i> ( <i>df</i> = 6)	0.67	0.45	0.12	0.43	0.32	0.03	0.03	2.2
<i>p</i> Value	0.42	0.51	0.73	0.51	0.57	0.86	0.86	0.14

<sup>a</sup>  $\beta$  = regression coefficient; *df* = degrees of freedom; SEM = standard error of the mean. Positive, statistically significant coefficients for input variables associated with an increased assessment score (output) are highlighted in bold. Negative coefficient for SCALE-UP indicates an effect for the Mock-up environment.

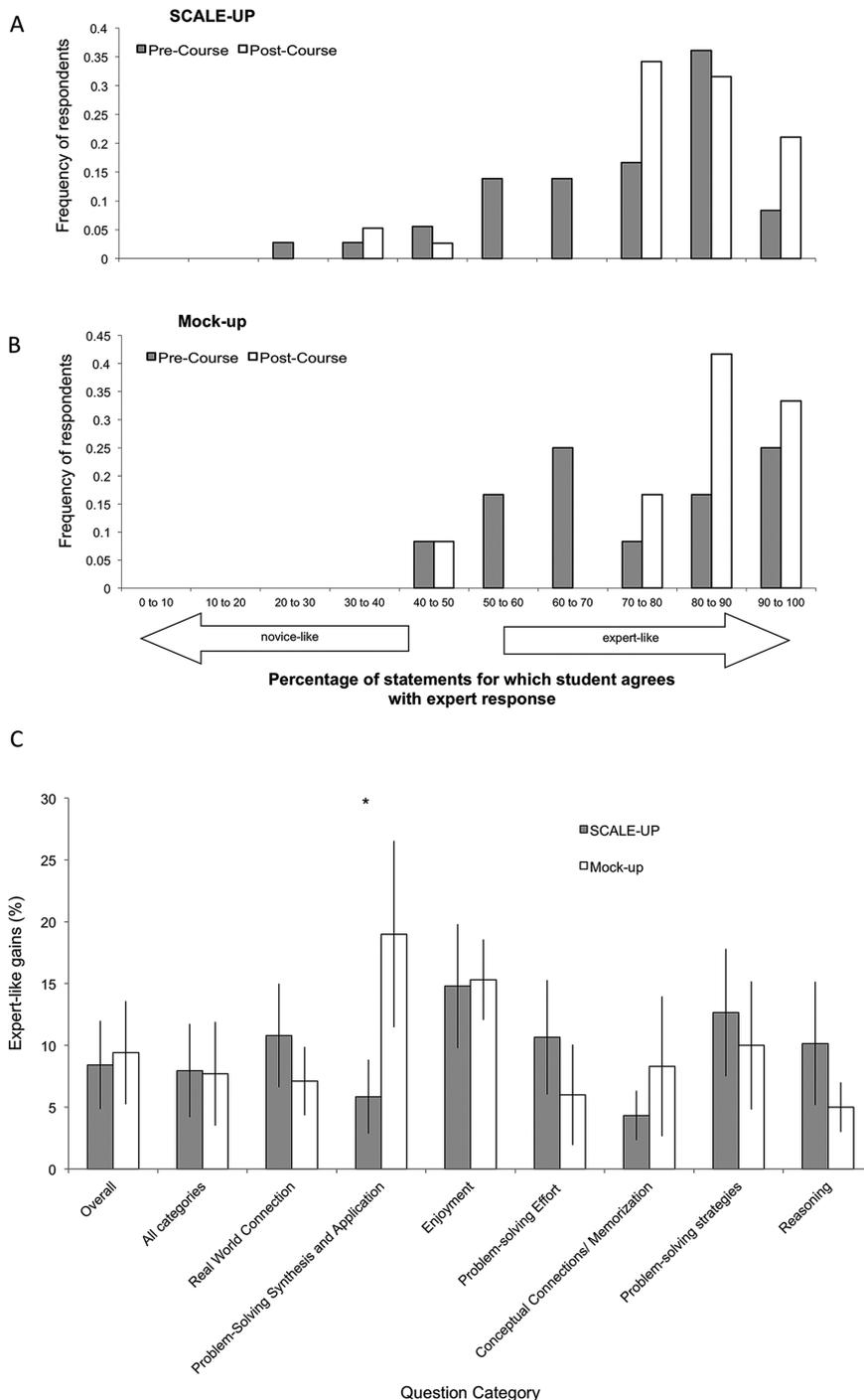
students in the Mock-up classroom showed significantly greater gains in perceptions of their problem-solving abilities at the synthesis and application level (>2 SEM). In summary, the overall progress made by the students in their novice-to-expert perceptions of biology were not due to differences in the SCALE-UP or Mock-up learning environments.

At the conclusion of the semester, students enrolled in the SCALE-UP and Mock-up spaces were asked to rank the order of relative helpfulness of classroom features in a forced-choice response survey and to reflect about their overall course experience in an open-response format. Students identified writable walls and tables as the most effective to their learning in the Mock-up room, whereas multiple sight lines and writable walls took precedence over table pods in the high-tech SCALE-UP classroom. Consistent with phase 1 focus groups and survey data (Figure 4 and Table 5), the technology feature plugging in was not deemed particularly useful for either group in this quasi-experimental phase 2 study (Table 5).

To determine the degree to which enrollment in the SCALE-UP and Mock-up classrooms influenced student satisfaction in

the course and toward the subject, we asked students to rate their course experience at the end of the semester. Frequency distributions of responses to the survey question "Overall, I rate this course as excellent" are shown in Figure 7A. Students in the Mock-up classroom showed a slightly more robust tendency to agree with this statement. Although regression modeling suggested that students in the Mock-up classroom were 3.5 times more likely to agree with this statement, this trend was not statistically significant at the 95% confidence interval (Table 7;  $\beta = 1.25 \pm 1.12$  SE,  $p = 0.26$ ). Similar patterns were observed for the survey question "I have more positive feelings about the subject" (Figure 7B and Table 7;  $\beta = 1.06 \pm 1.1$  SE,  $p = 0.34$ ). Thus, we conclude that students in both the SCALE-UP and Mock-up learning environments had equally satisfying experiences in the course as an introduction to the discipline.

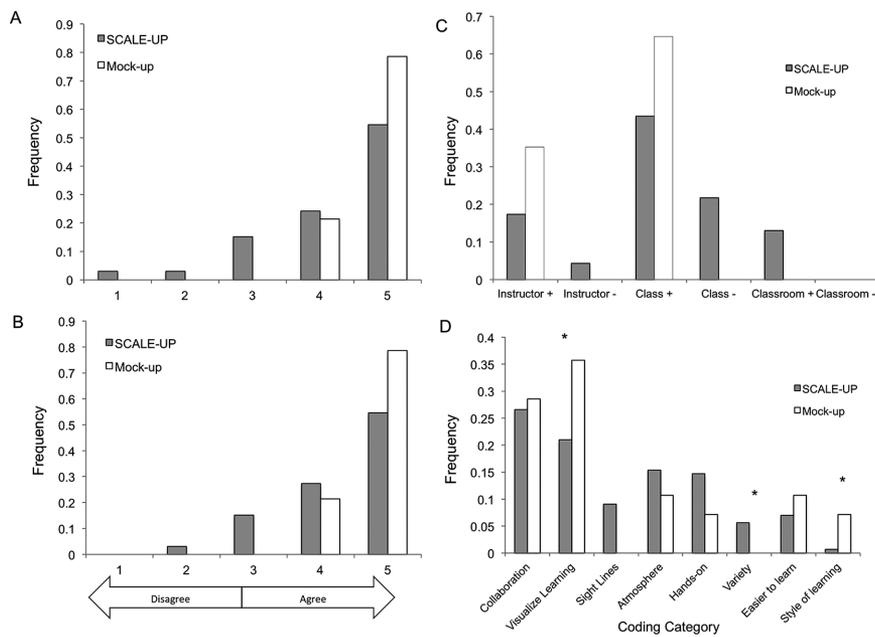
In addition to these survey items, content analysis of open-ended comments from the end-of-semester course evaluations were coded into categories encompassing positive and negative comments about the instructor, class, and classroom. There were significantly fewer negative comments in the Mock-up



**FIGURE 6.** Progression of novice to expert-like attitudes about biology in the SCALE-UP and Mock-up environments. Summary of the percentage of students agreeing with expert statements about the discipline of biology in the SCALE-UP (A) and Mock-up (B) classrooms. The overall shift to the right (white bars) indicates developmental progress during the course of the semester for matched data (Mann-Whitney  $U$ ,  $p < 0.05$  for both). Pre and post gains were comparable between SCALE-UP and Mock-up environments overall (Student's  $t$   $p > 0.05$  for pre and post). (C) Breakdown of favorable gains by factor loading. The progress made in certain categories differed between SCALE-UP and Mock-up populations (error bars = SEM; asterisk indicates significant at  $>2$  SEM), with students in the Mock-up group making significant progress in problem solving at the application and synthesis level relative to the SCALE-UP group.

classroom compared with the SCALE-UP classroom (Figure 7C; Chi-square,  $\chi^2 = 9.5$ ,  $df = 4$ ,  $p = 0.04$ ;  $n = 23$ ). Although there were no negative comments about the classroom per se, the negative comments about the class conveyed resistance to certain aspects of active learning as a whole (slower pace of content, resistance to group work and flipped homework structure), while other aspects of active learning were strongly valued in the positive comments (cooperative learning, personal responsibility for learning, sense of community). Such resistance to active learning was not evident in the Mock-up section, perhaps because this section enrolled predominantly freshman with fewer lecture-based preconceptions and expectations for their learning compared with the more experienced students in the SCALE-UP classroom. Taken together, student satisfaction from course evaluation ratings and comments indicated an equal, if not better, overall experience in the Mock-up classroom.

To further unpack student satisfaction in the SCALE-UP and Mock-up experiences, we coded written responses to “Would you take another course in a SCALE-UP classroom? Why or why not?” The majority (95%) of our respondents indicated “yes.” The minority of students who indicated “no” cited a few problems with their experience in the high-tech learning space: 1) sight lines (e.g., student couldn't see or multiple screens were a distraction), 2) decentralized space (e.g., not having a clear front with a teacher at the center), and 3) slower pace compared with lecture learning. Notably, these responses were consistent with the resistance to active learning and preference for lecture pedagogies in Figure 7C. Interestingly, no students enrolled in the Mock-up section indicated “no” to this prompt. Nonetheless, for the vast majority of our learners, both spaces were worthy of a repeat course experience for several reasons. Principal among the most often cited reasons were the value of peer-peer collaboration (28% of coded responses) and the sense of community (15%) fostered by pod seating. Particularly for the Mock-up students, being able to visualize thinking and learning using whiteboards, even as a proxy for writable walls, was more important (36%) than for students in the high-tech SCALE-UP space (21%; Figure 7D;



**FIGURE 7. Summary of student satisfaction in SCALE-UP and Mock-up experiences.** (A) Frequency distributions of responses (1 = strongly disagree; 5 = strongly agree) to the survey question “Overall, I rate this course as excellent.” Students in the Mock-up and the SCALE-UP classrooms perceived the course as equally excellent ( $\beta = 1.25 \pm 1.12$  SE,  $p = 0.26$ ). (B) Frequency distributions for responses to the survey question “I have more positive feelings about the subject.” Students in the Mock-up classroom had positive feelings toward biology similar to students in the SCALE-UP room ( $\beta = 1.06 \pm 1.1$  SE,  $p = 0.34$ ). (C) Summary of comments from open-ended survey questions. Open-ended comments from the end-of-semester course evaluations were coded into categories encompassing positive and negative comments about the instructor, class, and classroom. There were fewer negative comments in the Mock-up classroom compared with the SCALE-UP classroom (Chi-square,  $\chi^2 = 9.5$ ,  $df = 4$ ,  $p = 0.04$ ). (D) Frequency distribution of yes-coded responses to “Would you take another course in a SCALE-UP classroom? Why or why not?” The distributions between SCALE-UP and Mock-up responses were comparable overall (Chi-square,  $\chi^2 = 9.5$ ,  $df = 7$ ,  $p = 0.44$ ), but students in the Mock-up classroom valued visualization using low-tech whiteboards more than students using writable walls in the high-tech room (Chi-square,  $\chi^2 = 5.5$ ,  $df = 1$ ,  $p = 0.02$ ). Asterisks indicate significance at  $p < 0.05$  between SCALE-UP and Mock-up.

Chi-square,  $\chi^2 = 5.5$ ,  $df = 1$ ,  $p = 0.02$ ). Notably, more students in the Mock-up classroom stated that the room favored their learning style (Figure 7D; Chi-square,  $\chi^2 = 4.7$ ,  $df = 1$ ,  $p = 0.03$ ), and SCALE-UP students preferentially valued the variety or “change of pace” from traditional spaces (Figure 7D; Chi-square,  $\chi^2 = 6.2$ ,  $df = 1$ ,  $p = 0.01$ ). Relatively few responses highlighted multiple sight lines (9%) as a reason for retaking a class in the high-tech room. Consistent with our phase 1 findings, 0% of the respondents indicated that the costly, high-tech feature of plugging in was a reason to take another course in a SCALE-UP classroom.

## DISCUSSION

### Is the Cost of Enhanced Technology in the SCALE-UP Classroom Worth the Investment for Smaller Institutions?

Results from this study show that three classroom components emerged as critical for faculty and students: 1) whiteboard space, 2) collaborative seating, and 3) multiple sight lines for viewing. These findings are consistent with a recently published paper (Knaub *et al.*, 2016) that looked at 21 secondary adopters of SCALE-UP classrooms. Knaub *et al.* (2016) found that having collaborative seating (e.g., round tables), technology (e.g., multiple screens, instructor station, student computer use) and student whiteboard spaces were the top-mentioned features of the SCALE-UP classrooms, respectively. Of these features, the seating arrangement and whiteboards were mentioned as being helpful, whereas technology received mixed reviews. Likewise, participants in our study mentioned whiteboards and seating (with associated sight lines) as the top features of the classroom, a pattern consistent between 100- and 300-level biology students. However, the order of these features differed slightly between the lower- and upper-level student groups (Figure 4). These subtle differences were consistent with attitudes measured in pilot studies by the University of Minnesota, where first- and second-year students rated the new learning spaces significantly higher in terms of engagement, enrichment, effectiveness, flexibility, fit, and instructor use compared with their upper-division peers (Whiteside *et al.*, 2010).

In addition, faculty reactions to the space were enthusiastic and consistent with other pilot studies (Whiteside *et al.*, 2010; Walker *et al.*, 2011; Cotner *et al.*, 2013), especially with respect to efficiencies in providing formative feedback in session. While faculty expressed technology as a helpful feature of the room, they explained the usefulness in terms of being able to visualize student work as a means to increase accountability, noting that whiteboards also enabled them to see student work and increase participation. These results also suggest that the more costly component—connecting all screens to the main teaching station—is least valuable to faculty and students.

**TABLE 7. Comparison of perceived course excellence in SCALE-UP and Mock-up classrooms<sup>a</sup>**

Question	Odds ratio (95% CI) <sup>b</sup>	$\beta$ (SE)	$p$ Value	Intercept (SE)	$p$ Value
Overall, I rate this course as excellent.	3.5 (0.38–31.5)	1.25 (1.1)	0.26	1.3 (0.42)	0.002
I have more positive feelings about the subject.	1.1 (0.31–26.5)	2 (1.1)	0.34	1.5 (0.45)	0.0009

<sup>a</sup>  $\beta$  = regression coefficient; CI = confidence interval.

<sup>b</sup> The odds ratio reflects the likelihood of agreement with the given statement.

Interestingly, when asked at the end of the semester whether or not they would take another course in a SCALE-UP room and why/why not, no students commented about the technology feature of plugging in (Figure 7). Rather, students focused on collaborative features such as sitting with one another and visualizing their thinking on whiteboard spaces. When technology was valued, it was in reference to sight lines (i.e., multiple monitors from which to view projected material), a necessary feature for students seated in angular lines associated with circular table pods. While the lower preference for plugging in is consistent across courses and class level, it is possible that the impact of the instructor's pedagogical choices influenced the value students placed on the high-tech features of the classroom. For example, while the instructor used the plugging-in feature once per instructional unit, the writable spaces were used several times per session. Perhaps this feature would have been rated more highly if the sharing technology was used more frequently.

Despite the reduction in technology-enhanced features, learners in the Mock-up classroom did not differ from those in the SCALE-UP room in terms of achievement (Figure 5 and Table 6). We note that some coefficients from the models for in-class work and preclass homework showed a more favorable score for students enrolled in the Mock-up room, whereas exam scores and IMCA performance favored the SCALE-UP room (Table 6). Even though none of these trends were statistically significant, it is possible that differences in class standing and/or class size contributed to these trends. For example, the combination of majority freshman students in a smaller class may have led to increased pre- and in-class accountability in the Mock-up population. Similarly, class standing and increased experience in college-level biology may explain the favorable exam and IMCA scores in the SCALE-UP population. Overall, however, our results on student achievement demonstrate that no learning is lost in the Mock-up learning environment. Because SCALE-UP classrooms have notable achievement benefits (for a review, see Knaub *et al.*, 2016), it is encouraging to see similar benefits in a Mock-up classroom.

We also note that the effect of class size differences between the SCALE-UP and Mock-up sections were not specifically tested as predictors of performance in the regression models. In the analysis reported here, the regression accounts for the class size variable as part of the definition of "SCALE-UP versus Mock-up" and was coded as a 0 or 1. When class size was pulled out as a predictive variable, the resulting model was not able to compute the coefficient because of too few degrees of freedom for class size, returning an "NA" for the variable of class size (unpublished data). Given these limitations, the models reported in this study capture the variation in class size within the coded "SCALE-UP versus Mock-up" variables to the best extent possible.

Learners in the Mock-up room had certain attitudinal and affective gains that exceeded those of learners in the SCALE-UP room (Figures 6 and 7). In terms of attitudes toward the discipline of biology as a whole, students in both sections showed similar overall growth in their novice-to-expert perceptions, but the Mock-up students particularly excelled in application and synthesis-level problem solving (Figure 6). We interpret this finding in light of ceiling effects associated with class standing previously observed with the CLASS-BIO instrument

(Semsar *et al.*, 2011). In other words, the majority of freshmen enrolled in the Mock-up classroom may have more growth potential within the discriminating range of the CLASS-BIO instrument for this factor compared with their more experienced counterparts in the SCALE-UP classroom. In addition, the freshman gains in problem-solving ability may be connected primarily to the laboratory experience, which used a course-based undergraduate laboratory experience (CURE) emphasizing high-level Bloom's activities and scientific process skills ([www.smallworldinitiative.org](http://www.smallworldinitiative.org)). For many of these freshman students in the Mock-up classroom, this course was their first experience with a CURE, thus explaining the unexpectedly high gain in this area.

### Student Satisfaction

We observed slightly increased levels of course satisfaction in the Mock-up classroom, potentially attributable to a variety of factors. First, the SCALE-UP session was offered very early in the morning, starting at 7:40 am, and was the first course of the day for both the students and the instructor. In comparison, the Mock-up class session was offered at 9:00 am, after the majority of students had already completed the first session of the day. Thus, levels of student engagement and overall optimism for learning may have been higher due to alertness concomitant with their natural circadian rhythms. Second, it is possible that small adjustments to instruction occurred between the first (SCALE-UP) and second (Mock-up) sections as part of the instructor's reflective practice, in spite of the fact that each class session was designed identically. Such adjustments would be a natural instinct for the instructor, and any increased student satisfaction in the Mock-up may be a reflection of these adjustments. Finally and notably, attributes of the class size and/or class standing distribution may have contributed to a more enjoyable course experience for students in the Mock-up section. For example, the majority of freshmen enrolled in the Mock-up section may have been slightly more open-minded toward active learning and collaboration than more seasoned college learners enrolled in the SCALE-UP section.

Although small class size is typical for this institution, the disparity in class size between the SCALE-UP and Mock-up sections may have influenced the attitudinal differences we observed. The higher student satisfaction observed in the Mock-up section support previous reports that smaller class sizes contribute positively to student satisfaction (McKeachie, 1980). However, studies of low-satisfaction ratings in larger classes stem from a number of factors that were not applicable here, such as low student engagement (i.e., reliance on lecture), little use of active-learning approaches, and limited interaction with the faculty member (Cuseo, 2007). Notably, the Mock-up section used smaller a pod size—six students compared with nine students in the SCALE-UP room—in addition to having a lower overall enrollment. Thus, it is possible that the smaller pod size likewise made a significant contribution to student satisfaction in addition to overall class size, perhaps by enabling students to achieve familiarity to a smaller subset of learning collaborators, contributing to fewer distraction opportunities compared with the larger-enrollment SCALE-UP environment. Such a distraction phenomenon would be further exacerbated in the layout of the SCALE-UP room, which contained multiple lines of sight to other nearby

Pods not present in the Mock-up layout. Future studies investigating the specific role of pod size and class size on student satisfaction and distraction in the SCALE-UP environment would be valuable for determining the extent to which our observations here are generalizable.

Student responses to why they would/would not take another class in a SCALE-UP room were particularly revealing about the benefits of individual features of the room (Figure 7). In both classrooms, collaboration and visualizing learning overwhelmingly emerged as the top two reasons to repeat-enroll in a SCALE-UP room. Interestingly, students in the Mock-up room more frequently commented on visualizing their learning compared with students in the SCALE-UP room. Perhaps this has to do with both the proximity and experiential access to the whiteboard, since these foam boards were positioned directly on top of student tables. Students could remain seated while passing, holding, or reorienting the board during discussion. This proximity and tactile interaction with the writable space may have contributed to increased ownership of learning. In contrast, the SCALE-UP room required students to stand up and migrate 2–3 feet away from the pod to write and create on walls. Often, students would delegate scribes to write on behalf of the nine-student pod, thus potentially decreasing individual ownership of learning.

It is especially important to note that no students in either section mentioned any aspect of technology as a reason to re-enroll in a SCALE-UP classroom. This was an especially surprising result, since the original intent of the classroom design emphasized the use of plugging in computers for collaboration. In agreement with our study, a more recent review of SCALE-UP implementation likewise noted that technology may not be as important as originally conceived, suggesting that the SCALE-UP experience may be implemented without the plugging-in technology. Our quasi-experimental study provides direct evidence to support this claim, heretofore only supported anecdotally (Knaub *et al.*, 2016).

Our results indicate that certain aspects of the SCALE-UP room design are essential to student success—collaborative seating and writable space. If classrooms are reconfigured to emphasize these features, technology and multiple sight-line screens are not needed, and a single line of sight (i.e., a clear “front”) enables students to see the projection screen. Although this arrangement does not decentralize the experience as in the original SCALE-UP design, we argue that the advantages of the SCALE-UP experience can still be achieved with a single sight line and by omitting the technology features (multiple sight lines and plugging in). Because each institution has unique circumstances and resources, we encourage similar institutions to consider these results in their classroom planning and design if cost savings is a priority. Furthermore, we encourage institutions to conduct similar studies to evaluate the impact of pilot SCALE-UP spaces on their unique student populations not only to test the generalizability of our results but also to inform the design of their classrooms according to specific end-user needs.

## ACKNOWLEDGMENTS

We are grateful to the many colleagues within the SABER community who provided feedback during the developmental stages of this project, as well as members of the Bethel University

Biology Department and University Classroom Oversight Committee. We especially thank Adam Wyse for assistance with multivariate regression. No external funding sources were used in the creation of this mock classroom experience. This study was conducted under the guidelines and approval of Bethel University’s Institutional Review Board.

## REFERENCES

- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.
- Anderson LW, Krathwohl DR, Bloom BS (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives*, Boston, MA: Allyn & Bacon.
- Antoine G, Harrell F (2000). Ordinal response regression models in ecology. *J Veg Sci* 11, 617–626.
- Beichner RJ, Saul JM, Allain RJ, Deardorff DL, Abbott DS (2000). Introduction to SCALE-UP: Student-Centered Activities for Large Enrollment University Physics. [www.ncsu.edu/ncsu/pams/physics/Physics\\_Ed/Articles/01ASEE\\_paper\\_S-UP.pdf](http://www.ncsu.edu/ncsu/pams/physics/Physics_Ed/Articles/01ASEE_paper_S-UP.pdf) (accessed 26 January 2017).
- Beichner RJ, Saul JM, Abbott DS, Morse JJ, Deardorff D, Allain RJ, Bonham SW, Dancy MH, Risley JS (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. In: *Research-Based Reform of University Physics*, ed. EF Redish and PJ Cooney, College Park, MD: American Association of Physics Teachers, 1(1), 2–39.
- Blanchard MR, Southerland SA, Osborne JW, Sampson VD, Annetta LA, Granger EM (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Sci Educ* 94, 577–616.
- Bligh DA (2000). *What’s the Use of Lectures?* San Francisco: Jossey-Bass.
- Bogdan RC, Biklen SK (1998). Foundations of qualitative research in education. In: *Qualitative Research in Education: An Introduction to Theory and Methods*, 1–48.
- Brooks DC (2011). Space matters: the impact of formal learning environments on student learning. *Br J Educ Technol* 42, 719–726.
- Cotner S, Loper J, Walker JD, Brooks DC (2013). “It’s not you, it’s the room” (or, are the high-tech, active learning classrooms worth it?). *J Coll Sci Teach* 42(6), 82–88.
- Cuseo J (2007). The empirical case against large class size: adverse effects on the teaching, learning and retention of first year students. *J Fac Dev* 1, 5–21.
- Dori YJ (2007). Educational reform at MIT: advancing and evaluating technology-based projects on- and off-campus. *J Sci Educ Tech* 16, 279–281.
- Dori YJ, Belcher J (2005). How does technology-enabled active learning affect undergraduate students’ understanding of electromagnetism concepts? *J Learn Sci* 14, 243–279.
- Erol M, Ozcan M, Luft JA (2016). Six reasons to teach undergraduate courses in SCALE-UP classrooms: suggestions for higher education authorities and instructors. *Educ Sci* 641(184), 27–48.
- 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.
- Gaffney JDH, Richards E, Kustusch MB, Ding L, Beichner R (2008). Scaling up educational reform. *J Coll Sci Teach* 37(5), 48–53.
- Glaser BG, Strauss A.L (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine.
- Hestenes D, Wells M, Swackhamer G (1992). Force Concept Inventory. *Phys Teach* 30, 141–158.
- Hmelo-Silver CE (2004). Problem-based learning: what and how do students learn? *Educ Psych Rev* 16, 235–266.
- Johnson DW, Johnson RT, Smith K (1998). Cooperative learning returns to college: what evidence is there that it works? *Change* 30, 27–35.
- Knaub AV, Foote KT, Henderson C, Dancy M, Beichner RJ (2016). Get a room: the role of classroom space in sustained implementation of studio style instruction. *Intl J STEM Educ* 63, 8.

- Knight JK, Wise SB, Southard KM (2013). Understanding clicker discussions: student reasoning and the impact of instructional cues. *CBE Life Sci Educ* 12, 645–654.
- Knight JK, Wood WB (2005). Teaching more by lecturing less. *Cell Biol Educ* 4, 298–310.
- Lage MJ, Platt GJ, Treglia M (2000). Inverting the classroom: a gateway to creating an inclusive learning environment. *J Econ Educ* 31, 30–43.
- Lasry N, Dugdale M, Charles E (2014). Just in time to flip your classroom. *Phys Teach* 52, 34–37.
- Mazur E (1997). *Peer Instruction: A User's Manual*, Upper Saddle River, NJ: Prentice Hall.
- Mazur E (2009). Farewell, lecture? *Science* 323, 50–51.
- McKeachie W (1980). Class size, large classes and multiple sections. *Academe* 66, 24–27.
- Michael J (2006). Where's the evidence that active learning works? *Adv Phys Educ* 30, 159–167.
- Milne AJ (2006). Designing blended learning space to the student experience. In: *Learning Spaces*, ed. DG Oblinger, Washington, DC: Educause, 11.1–11.5.
- Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- National Research Council (NRC) (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.
- NRC (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Washington, DC: National Academies Press.
- Oblinger DG (ed.) (2006). *Learning Spaces*, Washington, DC: Educause.
- Prince M (2004). Does active learning work? A review of the research. *J Eng Educ* 93, 223–231.
- R Core Development Team (2013). *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [www.R-project.org](http://www.R-project.org) (accessed 26 June 2016).
- Savery JR (2006). Overview of problem-based learning: definitions and distinctions. *Interdisc J Problem-Based Learning* 1, ar3.
- Semsar K, Knight JK, Birol G, Smith MK (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE Life Sci Educ* 10, 268–278.
- Seymour E, Hewitt N (1997). *Talking about Leaving: Factors Contributing to High Attrition Rates among Science, Math and Engineering Undergraduate Majors*, Boulder, CO: Bureau of Sociological Research.
- 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.
- Toulmin SE (1958). *The Uses of Argument*, Cambridge, UK: Cambridge University Press.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, 4th ed., New York: Springer.
- Walker JD, Brooks DC, Baepler P (2011). Pedagogy and space: empirical research on new learning environments. *EDUCAUSE Q* 34(4). <http://er.educause.edu/articles/2011/12/pedagogy-and-space-empirical-research-on-new-learning-environments> (accessed 26 January 2017).
- Whiteside AW, Brooks DC, Walker JD (2010). Making the case for space: three years of empirical research on formal and informal learning environments. *EDUCAUSE Q* 33(3). <http://er.educause.edu/articles/2010/9/making-the-case-for-space-three-years-of-empirical-research-on-learning-environments> (accessed 26 January 2017).