

Article

Getting Under the Hood: How and for Whom Does Increasing Course Structure Work?

Sarah L. Eddy* and Kelly A. Hogan[†]

*Department of Biology, University of Washington, Seattle, WA 98195; [†]Department of Biology, University of North Carolina, Chapel Hill, NC 27599

Submitted March 17, 2014; Revised May 20, 2014; Accepted May 27, 2014
Monitoring Editor: Hannah Sevian

At the college level, the effectiveness of active-learning interventions is typically measured at the broadest scales: the achievement or retention of all students in a course. Coarse-grained measures like these cannot inform instructors about an intervention's relative effectiveness for the different student populations in their classrooms or about the proximate factors responsible for the observed changes in student achievement. In this study, we disaggregate student data by racial/ethnic groups and first-generation status to identify whether a particular intervention—increased course structure—works better for particular populations of students. We also explore possible factors that may mediate the observed changes in student achievement. We found that a “moderate-structure” intervention increased course performance for all student populations, but worked disproportionately well for black students—halving the black–white achievement gap—and first-generation students—closing the achievement gap with continuing-generation students. We also found that students consistently reported completing the assigned readings more frequently, spending more time studying for class, and feeling an increased sense of community in the moderate-structure course. These changes imply that increased course structure improves student achievement at least partially through increasing student use of distributed learning and creating a more interdependent classroom community.

INTRODUCTION

Studies across the many disciplines in science, technology, engineering, and mathematics (STEM) at the college level have shown that active learning is a more effective classroom strategy than lecture alone (reviewed in Freeman *et al.*, 2014). Given this extensive evidence, a recent synthesis of discipline-

based education research (DBER; Singer *et al.*, 2012) suggests that it is time to move beyond simply asking whether or not active learning works to more focused questions, including how and for whom these classroom interventions work. This type of research is being referred to as second-generation education research (Eddy *et al.*, 2013; Freeman *et al.*, 2014) and will help refine and optimize active-learning interventions by identifying the critical elements that make an intervention effective. Identifying these elements is crucial for successful transfer of classroom strategies between instructors and institutions (Borrego *et al.*, 2013).

Using these DBER recommendations as a guide, we have replicated a course intervention (increased course structure; Freeman *et al.*, 2011) that has been demonstrated to increase student achievement at an R1 university and explored its effectiveness when transferred to a different university with a different instructor and student population. Specifically, we expanded on the original intervention studies by exploring 1) how different student subpopulations respond to the treatment in terms of achievement and 2) course-related behaviors and perceptions. These two forms of assessment will help us both elucidate how this intervention achieves the observed increases in student achievement and identify the elements critical for the intervention's success.

DOI: 10.1187/cbe.14-03-0050

Address correspondence to: Kelly Hogan (Kelly_Hogan@unc.edu).
Conflict of interest statement: Kelly A. Hogan, a coauthor for Pearson's *Campbell Biology: Concepts and Connections*, 8th ed., and its associated Mastering Biology online tools (which were used in this study) was not affiliated with the products at the time of the course intervention. No promotion of Mastering Biology to the exclusion of other similar products should be construed.

© 2014 S. L. Eddy and K. A. Hogan. *CBE—Life Sciences Education*
© 2014 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.

Are Active-Learning Interventions Transferable?

The transferability of active-learning interventions into novel educational contexts is critical to the successful spread of active learning across universities (National Science Foundation, 2013). Unfortunately, transferability of an intervention across contexts cannot be assumed, as there is some evidence that the success of classroom interventions depends on the student populations in the classroom (Brownell *et al.*, 2013), instructor classroom management style (Borrego *et al.*, 2013), and the topics being taught (Andrews *et al.*, 2011). Thus, interventions that work with one instructor at one institution in one class may not necessarily transfer into novel contexts. Yet the majority of published active-learning interventions at the college level have been tested with at best one or two instructors who are usually at the same institution.

We test the transferability of the increased course structure intervention (Freeman *et al.*, 2011), which was effective at a Pacific Northwest R1 university with a predominately white and Asian student body, in a Southern R1 university with a different instructor (who had no contact with the original authors) and a more diverse student body. Additionally, the original study was an introductory biology course for aspiring majors, while the current implementation included mostly nonmajors in a mixed-majors general education course. Thus, in this study, we test the transferability of the increased course structure intervention across three contexts: 1) different instructors, 2) different student body, and 3) different courses (majors vs. nonmajors).

Do Course Interventions Differentially Impact Achievement in Some Student Subpopulations?

There is emerging evidence that classroom interventions could have different impacts on students from different cultural contexts. For example, Asian-American students learn less when they are told to talk through problems out loud compared with when they think through them silently. White students, on the other hand, performed just as well, and in some cases better, when allowed to talk through problems (Kim, 2002, 2008). This finding has implications for a differential impact of peer instruction on Asian students relative to their white classmates. In addition to different cultural norms for learning, students from different subpopulations bring different value sets into the classroom that can influence how they learn in different classroom environments. For example, one study found that when a setting is perceived as interdependent (rather than independent) first-generation students perform better, but continuing-generation students do not differ (Stephens *et al.*, 2012). Positive interpersonal feelings also increased the performance of Mexicans but not European Americans on a learning task (Savani *et al.*, 2013). Thus, the classroom environment itself could have differential impacts on different students. Findings like these begin to call into question whether “one-size-fits-all” classrooms interventions are possible and encourage researchers to disaggregate student response data by subpopulations (Singer *et al.*, 2012).

Up until now, the majority of college-level program evaluations that have disaggregated student groups have done so broadly based on their historical presence in science (underrepresented minority [URM] vs. majority students). Also, most of these studies have explored the impact of supplement-

tal instruction outside an actual science course on student achievement (reviewed in Tsui, 2007; Fox *et al.*, 2009). Only a few STEM course-based curricular interventions have disaggregated student performance (physics: Etkina *et al.*, 1999; Hitt *et al.*, 2013; math: Hooker, 2010; physical science: Poelzer and Zeng, 2008). In biology, two course-based active-learning interventions have been shown to reduce achievement gaps between historically underrepresented students and majority students. Preszler (2009) replaced a traditional course (3 h of lecture each week) with a reformed course that combined 2 h of lecture with 1 h of peer-led workshop. This change in class format increased the grades of all participating students, and the performance of URM students and females increased disproportionately. The second intervention was the increased course structure intervention (Haak *et al.*, 2011). This intervention decreased the achievement gap between students in the Educational Opportunities Program (students from educational or economically disadvantaged backgrounds) and those not in the program by 45% (Haak *et al.*, 2011).

Studies that cluster students into two categories (URM vs. majority) assume that students within these clusters respond in the same way to classroom interventions. Yet the URM label includes black, Latin@,¹ Native American, Hawaiian and Pacific Islander students, and the majority designation is often both white and Asian students. The consequence of clustering leads to conclusions that are too generalized; for example, that black students will respond in a similar way to a treatment as do Latin@ students (Carpenter *et al.*, 2006). Yet the different racial and ethnic groups that are included in the URM designation have very different cultures, histories, and exposure to college culture that could impact whether a particular classroom strategy is effective for them (Delpit, 2006). National trends in K–12 education, revealing different achievement patterns and trajectories for black and Latin@ students, also challenge the assumption that URMs are a homogeneous group (Reardon and Galindo, 2009).

To our knowledge, only two college-level curricular interventions in STEM, and none in biology, have subdivided the URM category into more fine-grained groups to explore the effectiveness of classroom interventions for these different student populations. In these studies, students of different racial/ethnic groups responded differently to the classroom interventions (Etkina *et al.*, 1999; Beichner *et al.*, 2007). This was demonstrated most dramatically by Beichner *et al.* (2007), in whose study white and black students were the only groups to benefit significantly from an active-learning intervention. These findings highlight the need for more studies to analyze college course performance by racial/ethnic groups. These smaller categories can still be problematic, as they still combine students with very different cultural backgrounds and experiences into broad categories such as white, Asian, Native American, and Latin@ (Lee, 2011; Carpenter *et al.*, 2006), but disaggregating students to this level will provide a finer-grained picture of the classroom than has been previously reported.

A second population of students of concern is first-generation students. These students have limited exposure to the culture of college and are often from working-class

¹Latin@ is a gender inclusive way of describing people of Latin American descent (Demby, 2013). The term is being increasingly used in the Latin@ community including many national organizations.

backgrounds that may be at odds with the middle-class cultural norms of universities (e.g., the emphasis on abstract over practical knowledge and independence over interdependence; Stephens *et al.*, 2012; Wilson and Kittleson, 2013). The differences between first- and continuing-generation students have been shown to change how they respond to “best-practices” in teaching at the college level, sometimes to the extent that they respond oppositionally (Padgett *et al.*, 2012). In biology, we are not aware of any studies that have explored the response of this population to an active-learning intervention, although there has been promising work with a psychology intervention (Harackiewicz *et al.*, 2014).

In our study, we explored whether racial (black, white, Native American, Asian) and/or ethnic (Latin@) identity and first-generation versus continuing-generation status influence a student’s response to the increased course structure. We hypothesized that different student groups would vary in the extent to which an active-learning intervention would influence their exam performance.

How Do Active-Learning Interventions Change Course-Related Behaviors and Attitudes of Students?

Understanding how interventions change course-related behaviors and attitudes is an important next step in education research, as these behaviors and attitudes mediate how the course structure influences performance (Singer *et al.*, 2012). Some work has already described how active learning increases achievement at the college level, although this work is lacking in the STEM disciplines and usually only looks at the student body as a whole. Courses with more active learning are positively correlated with increased student self-reported motivation and self-efficacy (van Wyk, 2012) and a deeper approach to learning (Eley, 1992). Unfortunately, this work is only done in active-learning classrooms, and either there is no control group (cf. Keeler and Steinhorst, 1995; Cavanagh, 2011) or the study asks students to compare their experience with a different course with a different instructor and content in which they are currently enrolled (cf. Sharma *et al.*, 2005). In our study, we examine how student attitudes and course-related behaviors change between a traditionally taught and an increased-structure course with the same content and instructor.

Reviewing the elements of successful classroom interventions suggests possible factors that could contribute to the increase in student achievement. For example, the increased course structure intervention involves the addition of three elements: graded preparatory assignments, extensive student in-class engagement, and graded review assignments (Table 1). Proponents of the increased course structure inter-

vention have hypothesized that the additional practice led to the rise in student performance (Freeman *et al.*, 2011). Yet providing opportunities for practice might not be enough. When and what students practice, as well as the context of and their perceptions of the practice may influence to the impact of the extra practice on learning.

There are many possible factors that change with the implementation of increased course structure. We focus on three candidate factors, but it is important to recognize that these factors are not mutually exclusive or exhaustive.

Factor 1. Time allocation: Increasing course structure will encourage students to spend more time each week on the course, particularly on preparation. How students allocate their out-of-class study time can greatly influence their learning and course achievement. Many students adopt the strategy of massing their study time and cramming just before exams (Michaels and Miethe, 1989; McIntyre and Munson, 2008). Yet distributed practice is a more effective method for learning, particularly for long-term retention of knowledge (Dunlosky *et al.*, 2013). The increased course structure helps students distribute their study time for the class by assigning daily or weekly preparatory and review assignments. These assignments 1) spread out the time students spend on the course throughout the quarter (distributed practice, rather than cramming just before exams) and 2) encourage students to engage with a topic before class (preparatory assignment) and then again in class (in-class activities) and again after class (review assignments). In addition, the preparatory assignments not only encourage students to read the book before class, but also have students answer questions related to the reading, which is a more effective method for learning new material than simply highlighting a text (Dunlosky *et al.*, 2013). We believe that the outside assignments scaffold how students spend time on the course and are one of the primary factors by which increased course structure impacts student performance. However, this idea has never been explicitly tested. In this study, we asked students to report how much time they spent outside of class on the course weekly and what they spent that time doing. We predicted that students would spend more time each week on the course and would spend more time on the parts associated with course points. These results would imply an increase in distributed practice and demonstrate that the instructor can successfully guide what students spend time on outside of class.

Factor 2. Classroom culture: Increasing course structure will encourage students to perceive the class as a community. To learn, students must feel comfortable enough to be

Table 1. The elements of a low-, moderate-, and high-structure course

	Graded preparatory assignments (example: reading quiz)	Student in-class engagement (example: clicker questions, worksheets, case studies)	Graded review assignments (example: practice exam problems)
Low (traditional lecture)	None or <1 per week	Talk <15% of course time	None or <1 per week
Moderate	Optional ^a : 1 per week	Talk 15–40% of course time	Optional ^a : 1 per week
High	≥1 per week	Talk >40% of course time	≥1 per week

^aNeed either a preparatory or review assignment once per week, but not both.

willing to take risks and engage in challenging thinking and problem solving (Ellis, 2004). High-stakes competitive classrooms dominated by a few student voices are not environments in which many students feel safe taking risks to learn (Johnson, 2007). The increased-structure format has students work in small groups, which may help students develop a more collaborative sense of the classroom. Collaborative learning in college has been shown to increase a sense of social support in the classroom as well as the sense that students like each other (Johnson *et al.*, 1998). This more interdependent environment also decreases anxiety and leads to increased participation in class (Fassinger, 2000) and critical thinking (Tsui, 2002). Increased participation on in-class practice alone could lead to increased performance on exams. In addition, a more interdependent environment has been shown to be particularly important for the performance of first-generation students and Mexican students (Stephens *et al.*, 2012; Savani *et al.*, 2013). Finally, feeling like they are part of a community increases both performance and motivation, especially for historically underrepresented groups (Walton and Cohen, 2007; Walton *et al.*, 2012). We predicted that students in an increased-structure course would change how they viewed the classroom, specifically, that they would feel an increased sense of community relative to students in low-structure courses.

Factor 3. Course value: Increasing course structure will increase the perceived value of the course to students. In the increased-structure course, students come to class having read the book, or at least worked through the preparatory assignment, and thus have begun the knowledge acquisition stage of learning. This shift of content acquisition from in class to before class opens up time in the classroom for the instructor to help students develop higher-order cognitive skills (Freeman *et al.*, 2011), providing opportunities to encourage students to make connections between course content and real-world impacts and to work through challenging problems. These opportunities for practice and real-world connections are thought to be more engaging to students than traditional lecture (Handelsman *et al.*, 2006). Thus, through increased engagement with the material (because of increased interest in it) student performance will increase (Carini *et al.*, 2006). We predicted students in the increased-structure course would feel more engaged by the material and thus would value the course more.

We considered these three factors—time allocation, classroom culture, and course value—when surveying students about their perceptions and behaviors. We analyzed student survey responses in both the traditional and increased-structure course to identify patterns in responses that support the impact of these three factors on student performance.

In summary, we test the transferability of one active-learning intervention (increased course structure; Freeman *et al.*, 2011) into a novel educational context. We expand upon the initial studies by 1) disaggregating student performance to test the hypothesis that student subpopulations respond differently to educational interventions and 2) using student self-reported data to identify possible factors (time allocation, classroom culture, course value) through which the intervention could be influencing student achievement.

METHODS AND RESULTS

The Course and the Students

The course, offered at a large research institution in the Southeast that qualifies as a more selective, full-time, 4-yr institution with a low transfer-in rate on the Carnegie scale, is a one-semester general introduction to biology serving a mixed-majors student population. The course is offered in both Fall and Spring semesters. Course topics include general introductions to the nature of science, cell biology, genetics, evolution and ecology, and animal physiology. The class met three times a week for 50 min each period. An optional laboratory course is associated with the lecture course, but lab grades are not linked to lecture grade. Although multiple instructors teach this course in a year, the data used in this study all come from six terms taught by the same instructor (K.A.H.). The instructor holds a PhD in pathology and laboratory medicine and had 6 yr of experience teaching this course before any of the terms used in this study.

The majority of students enrolled in the course were in their first year of college (69%), but the course is open to all students. The class size for each of the six terms of the study averaged 393 students. The most common majors in the course include biology, exercise and sports science, and psychology. The combined student demographics in this course during the years of this study were: 59% white, 13.9% black, 10.3% Latin@, 7.4% Asian, 1.1% Native American, and 8% of either undeclared race, mixed descent, or international origin. In addition, 66.3% of the students identified as female, 32.1% male, and 1.6% unspecified gender, and 24% of these students were first-generation college students.

The Intervention: Increasing Course Structure

Throughout our analyses, we compared the same course during three terms of low structure and three terms of moderate structure (Table 1). How these designations—low and moderate—were determined is explained later in the section *Determining the Structure Level of the Intervention*.

During the low-structure terms of this study (Spring 2009, Fall 2009, Spring 2010), the course was taught in a traditional lecture format in which students participated very little in class. In addition, only three homework assignments were completed outside the classroom to help students prepare for four high-stakes exams (three semester exams and one cumulative final).

In the reformed terms (Fall 2010, Spring 2011, Fall 2011), a moderate-structure format was used with both in-class and out of class activities added. The elements added—guided-reading questions, preparatory homework, and in-class activities—are detailed below, and Table 2 gives some specific examples across one topic.

Guided-Reading Questions. Twice a week, students were given ungraded, instructor-designed guided-reading questions to complete while reading their textbook before class. These questions helped to teach active reading (beyond highlighting) and to reinforce practice study skills, such as drawing, using the content in each chapter (Table 2; Supplemental Material, section 1). While these were not graded, the expectation set by the instructor was that the daily activities built from this content and referred to them, without covering them in the same format. Keys were not posted.

Table 2. Sample question types associated with the three assignment types added during the moderate-structure terms

Example learning objective: Determine the possible combinations of characteristics produced through independent assortment and correlate this to illustrations of metaphase I of meiosis		
Preclass (ungraded)	Preclass (graded)	In-class (extra credit)
<p>Example guided-reading questions</p> <ol style="list-style-type: none"> 1. Examine Figure 8.14, why are the chromosomes colored red and blue in this figure? What does red or blue represent? 2. Describe in words and draw how independent orientation of homologues at metaphase I produces variation. 	<p>Example preparatory homework question</p> <p>Independent orientation of chromosomes at metaphase I results in an increase in the number of:</p> <ol style="list-style-type: none"> a) Sex chromosomes b) Homologous chromosomes c) Points of crossing over d) Possible combinations of characteristics e) Gametes 	<p>Example in-class questions</p> <p>Students were shown an illustration of a diploid cell in metaphase I with the genotype AaBbDd.</p> <p>For all questions, students were told to “ignore crossing over.”</p> <ol style="list-style-type: none"> 1. For this cell, what is $n = ?$ 2. How many unique gametes can form? That is, how many unique combinations of chromosomes can form? 3. How many different ways in total can we draw metaphase I for this cell? 4. How many different combinations of chromosomes can you make in one of your gametes?

Preparatory Homework. Students were required to complete online graded homework associated with assigned readings before coming to class (Mastering Biology for Pearson’s *Campbell Biology: Concepts and Connections*). The instructor used settings for the program to coach the students and help them assess their own knowledge before class. Students were given multiple opportunities to answer each question (between two and six attempts, depending on question structure) and were allowed to access hints and immediate correct/incorrect answer feedback. The questions were typically at the knowledge and comprehension levels in Bloom’s taxonomy (Table 2).

In-Class Activities. As course content previously covered by lecture was moved into the guided-reading questions and preparatory homework, on average 34.5% of each class session was now devoted to activities that reinforced major concepts, study skills, and higher-order thinking skills. Students often worked in informal groups, answering questions similar to exam questions by using classroom-response software (www.poll Everywhere.com) on their laptops and cell phones. Thirty-six percent of these questions required a student to apply higher-order cognitive skills such as application of concepts to novel scenarios or analysis (see Supplemental Material, section 2, for methods). Although responses to in-class questions were not graded, students received 1–2 percentage points of extra credit on each of four exams if they participated in a defined number of in-class questions. The remaining 65.5% of class time involved the instructor setting up the activities, delivering content, and course logistics. These percentages are based on the observation of videos from four randomly chosen class session videos. The course was videotaped routinely, so the instructor did not know in advance which class sessions would be scored.

Determining the Structure Level of the Intervention

Using the data from two articles by Freeman and colleagues (Freeman *et al.*, 2007, 2011) and consulting with Scott Freeman (personal communication) and the Biology Educa-

tion Research Group at the University of Washington, we identified the critical elements of low, moderate, and high structure (Table 1). Based on these elements, our intervention was a “moderate” structure course: we had weekly graded preparatory homework, students were talking on average 35% of class time, and there were no graded review assignments.

Study 1: Does the Increased Course Structure Intervention Transfer to a Novel Environment?

Total Exam Points by Course Structure. Our measure of achievement was total exam points. We chose this measure over final grade, because the six terms of this course differed in the total points coming from homework (3 vs. 10%) and the opportunity for bonus points could inflate the final grade in the reformed class. Instead, we compared the total exam points earned out of the possible exam points. As total exam points varied across the six terms by 5 points (145–150), all terms were scaled to be out of 145 points in the final data set.

As this study took place over 4 years, we were concerned that term-to-term variation in student academic ability and exam difficulty could confound our survey and achievement results. To be confident that any gains we observed were due to the intervention and not these other sources of variation, we controlled for both exam cognitive level (cf. Crowe *et al.*, 2008) and student prior academic achievement (for more details see Supplemental Material, section 2). We found that exams were similar across all six terms and that the best control for prior academic achievement was a student’s combined SAT math and SAT verbal score (Table 3; Supplemental Material, section 2). We therefore used SAT scores as a control for student-level variation in our analyses and did not further control for exams.

Course and Exam Failure Rates by Course Structure. To become a biology major, students must earn a minimum of a “C–” in this course. Thus, for the purpose of this study, we considered a grade below 72.9% to be failing, because the

Table 3. Regression models used to determine whether 1) increased structure can be transferred to a novel environment (study1) and 2) student subpopulations vary in their response to increased course structure (study 2)^a

Base model: Student performance influenced by course structure	Outcome \sim Term + Combined SAT scores + Gender + Course Structure
Model 2: Impact of course structure on student performance varies by race/ethnicity/nationality.	Outcome \sim Term + SAT scores + Gender + Course Structure + Race + Race \times Course Structure
Model 3: Impact of course structure on student performance varies by first-generation status.	Outcome \sim Term + SAT scores + Gender + Course Structure + First-generation + First-generation \times Course Structure

^aBolded terms in models 2 and 3 are the new additions that test the specific hypotheses that the impact of course structure will vary by student populations. The outcome variable is either student achievement on exams or student failure rates.

student earning this would not be able to move on to the next biology course. We measured failure rates in two ways: 1) final grade and 2) total exam points. Although the components contributing to final course grade changed across the study, this “C–” cutoff for entering the biology major remained consistent. This measure may be more pertinent to students than overall exam performance, because it determines whether or not they can continue in the major.

To look more closely at whether increased student learning was occurring due to the intervention, we looked at failure rates on the exams themselves. This measure avoids the conflation of any boost in performance due to extra credit or homework points or deviations from a traditional grading scale but is not as pertinent to retention in the major as course grade.

The statistical analysis for this study is paired with that of study 2 and is described later.

Study 2. Does the Effectiveness of Increased Course Structure Vary across Different Student Populations?

In addition to identifying whether an overall increase in achievement occurred during the moderate-structure terms, we included categorical variables in our analyses to determine whether student subpopulations respond differently to the treatment. We focused on two designations: 1) student ethnic, racial, or national origin, which included the designations of Asian American, black, Latin@, mixed race/ethnicity, Native American, white, and international students; and 2) student generational status (first-generation vs. continuing-generation college student). Both of these factors were determined from student self-reported data from an in-class survey collected at the end of the term.

Statistical Analyses: Studies 1 and 2

Total Exam Points Earned by Course Structure and Student Populations. We modeled total exam points as continuous response and used a linear regression model to determine whether moderate course structure was correlated with increased exam performance (Table 3). In our baseline model, we included student combined SAT scores, gender identity (in this case, a binary factor: 0 = male, 1 = female), and the term a student was in the course (Fall vs. Spring) as control variables. Term was included, because the instructor has historically observed that students in the Spring term perform better than students in the Fall term.

To test our first hypothesis, that increasing the course structure would increase performance (study 1), we included treatment (0 = low structure, 1 = moderate structure) as a binary

explanatory variable. To test our second hypothesis, that students from distinct populations may differ in their response to the classroom intervention, we ran two models (Table 3) that included the four variables described above and either 1) student racial and ethnic group (a seven-level factor) or 2) student first-generation status (a binary factor: 1 = first generation, 0 = continuing generation). If any of these demographic descriptors were not available for a student, that student was not included in the study.

We ran separate regression models for race/ethnicity and generation status, because we found these terms were correlated in an initial test of correlations between our possible explanatory variables (Kruskal-Wallis $\chi^2 = 68.1$, $df = 5$, $p < 0.0001$). Thus, to avoid any confounds due to multicollinearity (correlation between two explanatory variables), we ran each term in a separate model.

Course and Exam Failure Rates for Student Populations. We also explored whether the failure rate in the course decreased with the implementation of moderate course structure and whether different populations of students responded to the treatment differently. Our response variable was either 1) passing or failing the class or 2) passing or failing the exams (with <72.9% of possible points considered failing). We used a logistic regression (Table 3) to determine whether race/ethnicity, first-generation status, gender identity, and/or treatment were significant predictors of the response variable after we controlled for combined SAT scores, gender identity, and term (Fall or Spring).

Results: Studies 1 and 2

Total Exam Points Earned by Student Populations: Performance Increased for All Students but Increased Disproportionately for Black and First-Generation Students

Exam Performance by Course Structure and Student Race/Ethnicity/Nationality. In the low-structure terms—after we accounted for differences in SAT math and reading scores, gender identity, and differences between term—Asian, Native American, and white students had the highest achievement. Black ($\beta = -8.1 \pm 1.6$ SE, $p < 0.0001$) and Latin@ ($\beta = -3.4 \pm 1.7$ SE, $p = 0.044$) students scored significantly fewer exam points (6 and 2% fewer points, respectively; Table 4). It is important to note that the Native American category in our analysis contains very few students, and these results may be due to a lack of statistical power, rather than there being no real difference between white students and Native American student performance.

In the moderate-structure term, after we controlled for SAT scores, gender identity, and term, the classroom intervention

Table 4. Students of all racial/ethnic/national populations benefit from increased course structure and black students benefit disproportionately^a

Regression coefficients	Estimate ± SE	p Value
<i>Model intercept</i>	4.6 ± 4.52	0.310
Exam performance patterns under low structure		
<i>Race/Ethnicity/Nationality:</i> (reference level: White)		
Native American	-2.4 ± 4.38	0.569
Asian	0.1 ± 2.19	0.951
Black	-8.10 ± 1.56	<0.0001
Latin@	-3.4 ± 1.67	0.044
Mixed Race	0.8 ± 3.85	0.826
International	-7.0 ± 4.91	0.157
Exam performance patterns under moderate structure		
<i>Class Structure:</i> (reference level: Low Structure)		
Moderate Structure	4.6 ± 1.01	<0.0001
<i>Class Structure*Race/Ethn./Nat.:</i> (reference level: Moderate*White)		
Moderate*Native American	-2.3 ± 6.56	0.726
Moderate*Asian	0.2 ± 2.74	0.948
Moderate*Black	4.5 ± 2.08	0.031
Moderate*Latin@	2.4 ± 2.40	0.317
Moderate*Mix Race	-2.1 ± 4.66	0.657
Moderate*International	5.4 ± 6.96	0.440
Controls for student characteristics and term		
<i>Term:</i> (reference level: Fall)		
Spring	4.0 ± 0.77	<0.0001
SAT.Combined	0.08 ± 0.0033	<0.0001
<i>Gender:</i> (reference level: Male)		
Female	1.8 ± 1.13	0.022

^aCoefficients from the regression model of impact of race and course structure on exam points earned (out of 145). Coefficient estimates are in terms of raw exam points. The categorical variable Race/Ethnicity/Nationality represents achievement by racial group under low structure relative to the achievement of white students under low structure. The Class Structure term represents the gain all students see in the moderate-structure course. The interaction term between class structure and race (Class Structure*Race/Ethn./Nat.) represents the gains each racial group sees relative to the gains white students see under moderate structure. Significant terms for Class Structure*Race/Ethn./Nat. indicate a disproportionate impact (relative to white students) of moderate structure on that particular student group.

increased the exam performance of all students by 3.2% ($\beta = 4.6 \pm 1.01$ SE, $p < 0.0001$). There was also an additional significant interaction term between treatment and black students ($\beta = 4.5 \pm 2.08$ SE, $p = 0.031$; Table 4) that increased their predicted exam grade by an additional 3.1% (for a total increase of 6.3%). Under moderate structure, Native American, Asian, and white students still had the highest achievement, but the gap in scores between these students and black students was halved. The gap between white and Latin@ students was not significantly impacted by the intervention.

Exam Performance by Course Structure and First-Generation Status. The regression model using first-generation status as a predictor rather than ethnicity/race/nationality deviated slightly from the predictors of the previous model. First, there was no significant difference between males and female after

Table 5. All students benefit from increased course structure and first-generation students benefit disproportionately^a

Regression coefficients	Estimate ± SE	p Value
<i>Model intercept</i>	-6.1 ± 4.00	0.128
Exam performance patterns under low structure		
<i>Generation status:</i> (reference level: Continuing-generation)		
First-generation	-3.9 ± 1.19	0.0012
Exam performance patterns under moderate structure		
<i>Class Structure:</i> (reference level: Low)		
Moderate	5.4 ± 0.87	0.003
<i>Class Struc*Gen. Status:</i>		
Moderate*First Gen	3.5 ± 1.64	0.032
Controls for student characteristics and term		
SAT.Combined	0.08 ± 0.003	<0.0001
<i>Gender:</i> (reference level: Male)		
Female	1.6 ± 0.78	0.680
<i>Term:</i> (reference level: Fall)		
Spring	4.1 ± 0.76	<0.0001

^aCoefficients for regression model of generation status and course structure on exam points earned (out of 145). Coefficient estimates are in terms of exam points. The categorical variable Generation Status represents achievement by first-generation students under low structure relative to the achievement of continuing-generation students under low structure. The Class Structure term represents the gain all students see in the moderate-structure course. The interaction term between class structure and generation status (Class Struc*Gen. Status) represents the gains first-generation students experience relative to the gains of continuing-generation students under moderate structure. Significant terms for Course Structure*Gen. Status indicate a disproportionate impact of moderate structure on first-generation students.

we controlled for term, SAT math and reading scores, and first-generation status ($\beta = 1.6 \pm 0.78$ SE, $p = 0.68$; Table 5). Second, increasing the course structure provided all students, regardless of first-generation status, a slightly larger boost (3.7%, $\beta = 5.4 \pm 0.87$ SE, $p = 0.003$; Table 5).

Our main focus, however, is whether first-generation students responded differently to increased course structure relative to continuing-generation students. Under low structure, there was a 2.5% difference in exam points earned between first-generation students and continuing-generation students ($\beta = -3.9 \pm 1.2$ SE, $p = 0.001$; Table 5). With increased course structure, the performance of all students increased by 3.7%, and first-generation students experienced an additional 2.4% increase in exam performance for a total 6.1% increase ($\beta = 3.5 \pm 1.6$ SE, $p = 0.032$; Table 5). This disproportionate increase in first-generation student performance closes the achievement gap between first- and continuing-generation students.

Overall, the major pattern associated with student achievement was that exam points earned increased under moderate structure relative to low structure and increased disproportionately for black and first-generation students (Figure 1).

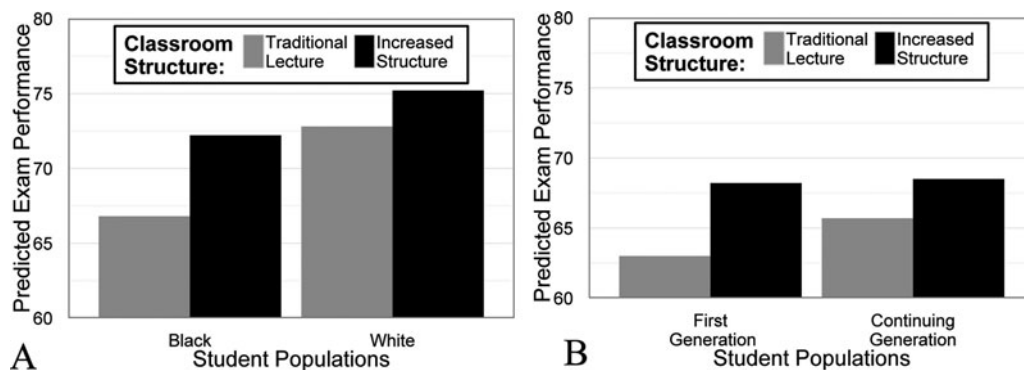


Figure 1. Some student populations (black and first-generation students) respond more strongly to increased structure than others. These figures are point estimates for exam performance (% correct) based on the regression models that included (A) race and ethnicity (Table 4) and (B) first-generation status (Table 5). The bars are the regression model predictions of performance for four hypothetical students who are in the Fall term of the course, are male, and have a combined SAT math and reading score of 1257 (the mean across the six terms). Thus, these students differ from each other in only two ways in each figure: whether they are in the low- (gray bars) or moderate- (black bars) structure course and (A) their racial identity (white vs. black) or (B) their generation status (first generation vs. continuing generation).

Course and Exam Failure Rates: Failure Rates Decreased for All Students

Raw failure rates in the course also dropped when comparing low with moderate structure. Overall, without controlling for any aspects of the course or student ability, 26.6% of the class earned a course grade of a “C–” or lower during the low-structure terms. With moderate structure, that failure rate dropped by 41.3% to 15.6%. Exam failure (earning <72.9% of the possible exam points) also decreased by 17.3% under moderate structure (from 52.6 to 43.5%).

After we controlled for student combined SAT score and the term, students experiencing moderate structure were 2.3 times more likely to earn above a “C–” in the course ($\beta = 0.828 \pm 0.19$ SE, $p < 0.0001$) and were 1.6 times more likely to earn more than 72.9% of the possible exam points ($\beta = 0.486 \pm 0.130$ SE, $p = 0.0001$). There was not a significant interaction between student race, gender, or first-generation status and course structure on this coarse scale.

Study 3: What Factors Might Influence Student Achievement in the Course with Increased Structure?

Methods. A 30-question survey was given to students immediately after the final exam in one term of the low-structure course and in three terms of the moderate-structure course. The survey questions focused on student course-related behaviors (Likert scale), student perception of the course, (Likert scale), and self-reported demographic variables (Supplemental Material, section 4). Specifically, we were interested examining three aspects of increased structure that might influence student learning. Survey questions had four or five response options. The questions associated with each factor are detailed in the following sections.

Factor 1. Time Allocation. We predicted that increased course structure would not only increase the time students spent on class material outside of class each week (thus distributing their practice throughout the term) but also the time they spent on behaviors associated with graded assignments. To test this prediction, we had students in both the low- and moderate-structure courses report how many hours they

spent studying a week and the frequency of behaviors related to preparing for and reviewing after class (Table 6, Supplemental Material, section 4). With our survey questions, it is not possible to parse out whether the increased hours of practice each week also led to an overall increase in the amount of time students spent practicing.

Factor 2. Classroom Culture. We predicted that students would feel a greater sense of interdependence in the moderate-structure course. To test this, we asked students to report how frequently they participated in class and how frequently they studied in groups outside of class. We also asked them three questions related to how interdependent they perceived the students in the class to be: how well they thought students in the class knew each other, if they believed students in the class tried to help one another, and whether they felt the class was a community (Table 6; Supplemental Material, section 4).

Factor 3. Course Value. We predicted that students would value the class and skills they acquired through the class more under moderate structure, because more higher-order skills were incorporated into the class. An assumption of this prediction is that students recognized that this class required higher-order thinking. To test this assumption, we asked students in both the low- and moderate-structure courses to identify how much of the course involved memorization. To assess the value students place in the course, we asked them to report lecture attendance, the importance of lecture for their learning, and the usefulness of the skills learned in the course for their future classes (Table 6; Supplemental Material, section 4).

In addition to looking at general impacts of active learning, we explored whether there were differences between student populations in their survey responses. These differences could help us understand why some student populations benefit more than others from the increased course structure intervention. The populations we focused on were identified through study 1: black and first-generation students performed disproportionately better in the increased-structure course relative to other student populations.

Table 6. Changes in student behaviors and perceptions with increased course structure^a

Characteristic	Low-structure term (raw median response)	Moderate-structure terms (raw median response)	Odds ratio: likelihood to increase with moderate structure (95% CI)	Odds ratio: likelihood to increase with increase in SAT scores (95% CI)
Factor 1. Time allocation: Increasing course structure will encourage students to spend more time each week on the course, particularly on preparation.				
Hours spent studying/week (0, 1–3, 4–7, 7–10, >10 h)	1–3 h	4–7 h	2.60 (2.02–3.35)	0.982 (0.974–0.990)
Complete readings before class (Never, Rarely, Sometimes, Often)	Rarely	Sometimes	1.97 (1.54–2.52)	0.994 (0.985–1.00)
Preparatory homework importance (Not at all, Somewhat, Important, Very)	Somewhat	Important	4.6 (3.56–5.85)	0.98 (0.97–0.98)
Review notes after class (Never, Rarely, Sometimes, Often)	Sometimes	Sometimes	0.738 (0.583–0.933)	0.972 (0.965–0.980)
Complete textbook review questions (Never, Rarely, Sometimes, Often)	Rarely	Rarely	0.50 (0.400–0.645)	0.98 (0.972–0.99)
Factor 2. Classroom culture: Increasing course structure will encourage students to perceive the class as more of a community.				
Contribute to classroom discussions (Never, Rarely, Sometimes, Often)	Never	Rarely	1.13 (0.890–1.44)	0.99 (0.988–1.00)
Work with a classmate outside of class (Never, Rarely, Sometimes, Often)	Sometimes	Sometimes	0.83 (0.664–1.06)	0.984 (0.0977–0.991)
Believe students in class know each other (Strongly disagree, Disagree, Neutral, Agree, Strongly agree)	Neutral	Neutral	2.4 (1.92–3.09)	0.996 (0.989–1.00)
Believe students in class help each other (Strongly disagree, Disagree, Neutral, Agree, Strongly agree)	Agree	Agree	1.22 (0.948–1.57)	1.01 (0.999–1.02)
Perceive class as a community (Strongly disagree, Disagree, Neutral, Agree, Strongly agree)	Neutral	Neutral	1.99 (1.57–2.52)	0.986 (0.979–0.993)
Factor 3. Course value: Increasing course structure will increase the value of the course to students.				
Amount of memorization (Most, Quite a bit, Some, Very Little, None)	Some	Some	1.07 (0.84–1.35)	0.98 (0.982–0.997)
Attend lecture (Never, Rarely, Sometimes, Often)	Often	Often	0.72 (0.471–1.09)	0.984 (0.971–0.997)
Use of skills learned (Strongly disagree, Disagree, Neutral, Agree, Strongly agree)	Agree	Agree	0.909 (0.720–1.15)	0.991 (0.983–0.998)
Lecture importance (Not at all, Somewhat, Important, Very)	Very Important	Important	0.57 (0.448–0.730)	0.998 (0.991–1.01)

^aThe second and third columns are the raw median responses under each structure. The fourth and fifth columns are the odds ratios from the log-odds regression including course structure and SAT scores as explanatory variables (> 1 = students more likely to report a higher value; < 1 = students more likely to report a lower value). SAT scores were divided by 10 to increase size of regression coefficients, so confidence intervals would be interpretable. Bolded odds ratios are significant.

Statistical Analysis: Study 3

General Patterns. We compared student responses to 14 survey questions concerning student course-related behaviors and perceptions of the classroom environment between one term of low course structure and three terms of moderate structure. Survey responses were ordered categorical responses (with four to five levels per question), so we used proportional log-odds regression models (implemented with the MASS package in R; Venables and Ripley, 2002). The proportional log-odds model works well for tightly bounded or ordinal data. The model compares the levels in the response variable by running a series of dichotomous comparisons (“never” vs. “rarely,” “rarely” vs. “sometimes,” etc.). Thus, the output of the log-odds regression is the effect that a change

in the explanatory variable (e.g., presence or absence of the classroom intervention) has on the odds that a student will report a higher rather than a lower response (i.e., “strongly agree” vs. “agree”) averaged across all possible levels of response (Antoine and Harrell, 2000).

Predictor variables for each model included: SAT combined score and course structure. We included SAT scores as a predictor, because we believed that students with higher levels of academic preparedness might differ from those with lower preparedness in terms of their course-related behaviors and attitudes. Thus, the model used for each question was: Survey response = Intercept + β*SATI.Comb + β*Treat. *p* Values were adjusted to account for false discovery rates due to multiple comparisons (Pike, 2011).

Population-Specific Patterns. On the basis of study 1, we identified two target populations who responded most strongly to the increased course structure: black and first-generation students. To test whether these two groups responded differently to any of the survey questions, we used forward stepwise model selection in the sequence that follows. We started with the base model, SAT math and verbal scores and treatment (Survey response = Intercept + β *SATI.Comb + β *Treat), and then added a main effect of first-generation status or black racial identity (i.e., Survey response = Intercept + β *SATI.Comb + β *Treat + β *First.Gen). We compared these two models to determine whether adding the new variable significantly increased the fit of the model to the data using standard model selection techniques. If the model with the population variable was significant, we then added an interaction between treatment and that variable (Survey response = Intercept + β *SATI.Comb + β *Treat + β *First.Gen + β *First.Gen*Treat) and tested the fit of this new model to the data. The significant difference in model fit (if present) was calculated using a type II analysis of variance (implemented with the car package in R; Fox and Weisberg, 2010). Again, p values were adjusted to account for false discovery rates due to multiple comparisons (Pike, 2011).

Correlation of Student Study Strategies/Perceptions with Total Exam Points. We also used survey responses to determine how student behaviors and attitudes correlated with their exam scores and whether this relationship with exam points was mediated by course structure.

Initially, we used the gamma rank correlation to explore whether responses on any of the 14 survey questions were correlated with responses on the other questions. This analysis was implemented in R using the rococo package (Bodenhofer and Klawonn, 2008). Even after correcting for the false discovery rate due to multiple comparisons, there were many moderate correlations between response in one survey question and response in another possibly leading to multicollinearity (for specific results, see Supplemental Material, section 3).

To be conservative, we ran 14 regression models, each linking a single survey question, the control for prior student academic achievement, and course structure to exam performance. This allowed us to determine whether the behavior or attitude in question was correlated with exam performance and whether this relationship was mediated by classroom structure. It did not allow us to identify which behaviors and attitudes relative to each other were most important.

Results: Study 3

General Patterns: Student Behaviors and Perceptions Changed with Increased Course Structure. Students reported employing different study strategies and perceiving the components of the course differently in low- and moderate-structure courses (Table 6).

Factor 1. Time Allocation. All five of the questions related to time allocation varied significantly with course structure (Table 6). As predicted, students spent more time each week preparing for class in the moderate-structure course. Specifically, students were 2.6 times more likely to report spending more hours a week studying for biology ($\beta = 0.95 \pm 0.13$ SE, $p < 0.0001$). Students focused more on preparing for class

in the moderate-structure course versus the low-structure course: after we controlled for SAT math and reading scores, students were 2.0 times as likely to complete reading assignments before class ($\beta = 0.68 \pm 0.12$ SE, $p < 0.0001$) and were 4.7 times more likely to report that the homework assignments were important for their understanding of course material in the moderate-structure term ($\beta = 1.5 \pm 0.12$ SE, $p > 0.0001$). Interestingly, even with the additional investment of hours each week, a focus on preparation seemed to represent a trade-off with time spent reviewing; after we controlled for SAT math and reading scores, students were 1.4 times less likely to review their notes after class as frequently ($\beta = -0.30 \pm 0.12$ SE, $p = 0.011$) and 1.9 times less likely to complete the practice questions at the end of each book chapter ($\beta = -0.68 \pm 0.12$ SE, $p < 0.0001$).

Factor 2. Classroom Culture. Of the five questions focused on the class climate, only two changed significantly with course structure (Table 6). As predicted, student in the moderate-structure terms were 2.0 times more likely to report a stronger sense of classroom community ($\beta = 0.69 \pm 0.12$ SE, $p < 0.0001$) and 2.4 times more likely to agree with the statement that “students in this class know each other” ($\beta = 0.89 \pm 0.12$ SE, $p < 0.0001$). The other three outcomes, which we expected would increase but did not, were: how strongly students believed that students in the class helped one another ($\beta = 0.20 \pm 0.13$ SE, $p = 0.12$), the frequency at which students worked with a partner outside of class ($\beta = -0.18 \pm 0.12$ SE, $p = 0.14$), and the frequency at which students participated in class ($\beta = 0.12 \pm 0.12$ SE, $p = 0.32$).

Factor 3. Course Value. Although we predicted outcomes that would suggest students valued the course more, we actually saw a decline (Table 6): students in the moderate-structure terms were 1.7 times less likely to rate the lecture component as important as students in the low-structure term ($\beta = -0.56 \pm 0.12$ SE, $p > 0.0001$). After we controlled for SAT math and reading scores, students also did not vary in their frequency of lecture attendance (although this could be because it was high to begin with; $\beta = -0.32 \pm 0.21$ SE, $p = 0.13$). Student perception of the importance of the skills they learned in the class did not vary between course structures ($\beta = -0.09 \pm 0.12$ SE, $p = 0.42$) nor did they perceive that the moderate-structure course involved more cognitive skills other than memorization ($\beta = 0.07 \pm 0.12$ SE, $p = 0.58$).

Population-Specific Patterns

Black Students Demonstrate Differences in Behaviors and Perceptions among Student Populations. On the basis of the results in study 1, which demonstrated that increased course structure was most effective for black and first-generation students, we explored student survey responses to determine whether we could document what was different for these populations of students.

We identified one behavior and three perception questions for which adding a binomial variable identifying whether a student was part of the black population or not increased the fit of the log-odds regression to the data. These differential responses may help us elucidate why this population responded so strongly to the increased-structure treatment.

The one behavior that changed disproportionately for black students relative to other students in the class was speaking in class. Under low structure, black students were 2.3 times more

likely to report a lower level of in-class participation than students of other ethnicities ($\beta = -0.84 \pm 0.35$ SE, $p = 0.012$). The significant interaction between being black and being enrolled in the moderate-structure course ($\beta = 0.89 \pm 0.38$ SE, $p = 0.019$) means this difference in participation completely disappears in the modified course.

Perception of the course also differed for black students compared with the rest of the students in three ways. First, black students were more likely to report that the homework was important for their understanding relative to other students in the class under both low and moderate structure. ($\beta = 1.06 \pm 0.31$ SE, $p = 0.0006$). The significant interaction term between course structure and black racial identity indicates the difference between black students and other students in the class decreases under moderate structure (Table 5; $\beta = 1.06 \pm 0.31$ SE, $p = 0.0006$), but this seems to be due to all students reporting higher value for the homework under moderate structure. In addition, black students perceived that there were less memorization and more higher-order skills in the class relative to other students in the class ($\beta = -0.39 \pm 0.59$ SE, $p = 0.024$) under both low and moderate structures. Finally, there was a trend for black students to be 1.3 times more likely to report that the skills they learned in this course would be useful for them ($\beta = 0.29 \pm 0.16$ SE, $p = 0.07$).

Unlike the clear patterns with black students, we found no significant differences in survey responses based on first-generation status.

Behaviors and Perceptions That Correlate with Success Are More Numerous under Moderate Structure. During the low-structure term, only lecture attendance impacted exam performance (i.e., significantly improved the fit of the models to the exam performance data after we controlled for student SAT scores; $F = 9.59$, $p < 0.0001$). Specifically, students who reported attending fewer lectures performed worse on exams. Students who reported accessing the textbook website more tended to perform better on exams ($F = 2.48$, $p = 0.060$), but this difference did not significantly improve the fit of the model.

In the moderate-structure terms, attending class ($F = 9.59$, $p < 0.0001$), speaking in class ($F = 9.03$, $p < 0.0001$), and hours spent studying ($F = 10.6$, $p < 0.0001$), reviewing notes ($F = 3.19$, $p = 0.023$), and seeking extra help ($F = 5.94$, $p < 0.0001$) all impacted student performance on exams. Additionally, one perception changed significantly: students with a higher sense of community performed better ($F = 4.14$, $p = 0.0025$).

DISCUSSION

With large foundation grants working toward improving STEM education, there has been a push for determining the transferability of specific educational innovations to “increase substantially the scale of these improvements within and across the higher education sector” (NSF, 2013). In this study, we provide evidence that one course intervention, increased course structure (Freeman *et al.*, 2011), can be transferred from one university context to another. In addition to replicating the increase in student achievement across all students, we were able to elaborate on the results of prior research on increased course structure by 1) identifying which student populations benefited the most from the increased course structure and 2) beginning to tease out the factors that may lead to these increases.

The Increased-Structure Intervention Can Transfer across Different Instructors, Different Student Bodies, and Different Courses (Majors vs. Nonmajors)

One of the concerns of any classroom intervention is that the results depend on the instructor teaching the course (i.e., the intervention will work for only one person) and the students in it. We can test the independence of the intervention by replicating it with a different instructor and student body and measuring whether similar impacts on student achievement occur. The university at which this study took place is quite different from the university where the increased course structure intervention was developed (Freeman *et al.*, 2011). Both universities are R1 institutions, but one is in the Southeast (and has a large black and Latin@ population), whereas the original university was in the Pacific Northwest (and has a high Asian population). Yet we find very similar results: in the original implementation of moderate structure in the Pacific Northwest course, the failure rate (defined as a course grade that would not allow a student to continue into the next grade in the biology series) dropped from 18.2% to an average of 12.8% (a 29.7% reduction; Freeman *et al.*, 2011). In our implementation of moderate structure, the failure rate dropped by a similar magnitude: from 26.6% to 15.6% (a 41.3% reduction). This result indicates that the impact of this increased-structure intervention may be independent of instructor and that the intervention could work with many different types of students.

Some Students Benefit More Than Others from Increased Course Structure

We found that transforming a classroom from low to moderate structure increased the exam performance of all students by 3.2%, and black students experienced an additional 3.1% increase (Figure 1A), and first-generation students experienced an additional 2.5% increase relative to continuing-generation students (Figure 1B). These results align with the small body of literature at the college level that indicates classroom interventions differ in the impact they have on student subpopulations (Kim, 2002; Preszler, 2009; Haak *et al.*, 2011). Our study is novel in that we control for both student past academic achievement and disaggregate student racial/ethnic groups beyond the URM/non-URM binary. Our approach provides a more nuanced picture of how course structure impacts students of diverse demographic characteristics (independent of academic ability).

One of the most exciting aspects of our results is that we confirm that active-learning interventions influence the achievement of student subpopulations differentially. This finding is supported by both work in physics (Beichner *et al.*, 2007), which found an intervention only worked for black and white students, and work in psychology, which revealed Asian-American students do not learn as well when they are told to talk through problems out loud (Kim, 2002). These studies highlight how important it is for us to disaggregate our results by student characteristics whenever possible, as overall positive results can mask actual differential outcomes present in the science classroom. Students come from a range of educational, cultural, and historical backgrounds and face different challenges in the classroom. It is not surprising that in the face of this diversity one intervention type does not fit all students equally.

Table 7. Changes in achievement and failure rate for SI programs in the first term of their implementation^a

Study	Classroom	Failure rate			Achievement		
		Non-SI	SI	% Change: failure rate	Non-SI	SI	% Change: achievement
Fullilove and Treisman, 1990	Calculus I	41%	7%	77	NA	NA	NA
Wischusen and Wischusen, 2007	Biology I	18.6%	6.9%	62.9	~85%	~87%	2.4
Rath <i>et al.</i> , 2007	Biology I	27%	15%	44.4	~75%	~79%	5.3
Peterfreund <i>et al.</i> , 2007	Biology I	27%	15%	44	~75%	~79%	5.3
Minchella <i>et al.</i> , 2002	Biology I and II	30.2%	16.9%	44	~75%	~78%	4
Barlow and Villarejo, 2004	General Chemistry	44%	28%	36.3	~80%	~83%	3.8
Dirks and Cunningham, 2006	Biology I	NA	NA	NA	~81%	~84%	3.7

^aMost achievement data were reported on the 4.0 scale, and the percentage of points earned was approximated using a conversion scale. In comparison, in the current student population, we saw a 41.3% reduction in the failure rate and a 3.2–6.3% increase in achievement, depending on which student subpopulation was the focus.

Comparing our results with published studies in STEM focused on historically underrepresented groups, we see that our achievement results are of a similar magnitude to other interventions. Unlike our intervention, previous interventions generally are not implemented within an existing course but are either run as separate initiatives or separate courses or are associated with a series of courses (i.e., involved supplemental instruction [SI]; cf. Maton *et al.*, 2000; Matsui *et al.*, 2003). These SI programs are effective, but can be costly (Barlow and Villarejo, 2004), and because of the cost, they are often not sustainable. Of seven SI programs that report data on achievement and retention in the first term or first two terms of the program, and thus are directly comparable to our study results, failure rate reductions ranged from 36.3 to 77%, and achievement increased by 2.4–5.3% (Table 7). In our study, the failure rate reduction was 41.3%, and overall exam performance increased by 3.2% (6.2% for black students and 6.1% for first-generation students), which is within the range of variation for the short-term results of the SI studies. These short-term results may be an underestimate of the effectiveness of the SI programs, as some studies have shown that their effectiveness increases with time (Born *et al.*, 2002). Yet the comparison still reveals promising results: one instructor in one course, without a large influx of money, can make a difference for students as large in magnitude as some supplemental instruction programs.

Exploring How Increased Course Structure Increases Student Performance

Survey data allowed us to explore how student course-related behaviors and attitudes changed with increased course structure. We focused on three specific factors and found evidence that changes in time allocation contributed to increased performance and some support for changes in classroom culture also impacting learning. We did not find evidence to support the idea that the value students found in the course influenced their performance.

Factor 1. Time Allocation. Under low structure, students on average spent only 1–3 h on the course outside of class, rarely came to class having read the assigned readings, and were highly dependent on the lecture for their learning. Students also placed little value on the occasional preparatory

homework assignments. With the implementation of moderate structure, students increased the amount of time they spent on the course each week to 4–7 h, were twice as likely to come to class having read the assigned readings, and saw the preparatory assignments as being equally as important for their learning as the actual lecture component. These shifts in behaviors and perceptions support our hypothesis that increased course structure encourages students both to distribute their studying throughout the term and to spend more time on behaviors related to graded assignments.

We believe that these changes in student behaviors and perceptions occurred because of the structure of accountability built into the moderate-structure course. Students reading before class is an outcome almost all instructors desire (based on the ubiquitous syllabus reading lists), but it is evident from our study and others that, under low structure, students were on average “rarely” meeting this expectation (see also Burchfield and Sappington, 2000). We found the dual method of assigning preparatory homework and making the reading more approachable with ungraded guided-reading questions increased the frequency of students reading before class. It seemed that course points (accountability) were necessary to invoke this change in student behavior, because we did not see a similar increase in the frequency with which students reviewed notes after class. It is possible that moving to high structure (Freeman *et al.*, 2011), with its weekly graded review assignments, could increase the achievement of our students even more, because they would be held accountable for reviewing their notes more frequently.

Factor 2. Classroom Culture. We found some evidence to support the hypothesis that increased course structure creates a community environment rather than a competitive environment. Under low structure, students did not seem to get to know the other students in the class and did not positively view the class as a community (although they did believe that students in the class tried to help one another). With increased structure, students were two times more likely to view the class as a community and 2.4 times more likely to say students in the class knew each other.

This result is a critical outcome of our study, arguably as important as increased performance, because a sense of being part of a community (belonging) is crucial for retention (Hurtado and Carter, 1997; Hoffman *et al.*, 2002) and has been

correlated with increased performance for first-generation students (Stephens *et al.*, 2012). When discussing reasons for leaving STEM, many students, particularly students of color and women, describe feelings of isolation and lack of belonging (Hewlett *et al.*, 2008; Cheryan *et al.*, 2009; Strayhorn, 2011). Because introductory courses are some of the first experiences students have in their major, these could potentially play a role in increasing retention simply by facilitating connections between students through small-group work in class.

Factor 3. Course Value. We did not find support for the hypothesis that students in the moderate-structure class found the course to be more valuable than students in the low-structure course. First, there was no difference in how much students valued the skills they learned in the course, but this could be because they did not recognize that the low- and moderate-structure terms were asking them to do different things. Across both terms, students on average believed that they were doing the same amount of memorizing versus higher-order skills such as application and analysis, even though the instructor emphasized higher-order skills more in the moderate-structure terms. In addition, behaviorally, we did not see any evidence of a higher value associated with the course in terms of increased attendance. In fact there was no difference in attendance across treatments. The attendance result was surprising to us, because increased attendance has been shown to be a common result of making a classroom more active (Caldwell, 2007; Freeman *et al.*, 2007); however, these previous interventions all assigned course points to in-class participation, whereas our interventions only gave students bonus points for participation. In a comparison of in-class attendance with and without points assigned to class participation, Freeman *et al.* (2007) found that attendance dropped in the class in which no points were assigned. Thus, it is possible that attendance in these classes could be increased in the future if points rather than extra credit were assigned for participation. This idea is supported by our data that it is actually the students with the highest predicted achievement (i.e., highest SAT scores) who are more likely to miss lecture. Because these students already were doing well in the course, it may be that the motivation of receiving a few bonus points for attending class was not enough encouragement.

Additional evidence that changes in time allocation and classroom culture contribute to achievement comes from the correlation between survey responses and exam performance. Under moderate structure, the number of hours a student spent studying per week and a higher sense of community were both positively correlated with exam performance.

The support for these two factors, time allocation and classroom culture, helps us identify potential critical elements for the implementation of the increased-structure intervention. First, students need to be made accountable for preparing before attending class. This can take multiple forms, including guided-reading questions, homework, and/or reading quizzes before class or at the start of class, but the key is that they need to be graded. Without this accountability in the low-structure terms, students were not doing the reading and were likely cramming the week before the exam instead of distributing their study time. The second critical element seems to be encouraging the students in the class to view themselves as a community through small-group work in

class. Further research could explore how best to approach in-class work to develop this sense of community rather than competition.

Changes in Achievement, Behaviors, and Perceptions Vary among Student Populations

In addition to looking at overall patterns in student behaviors and perceptions, we can also disaggregate these data to begin to understand why some groups might benefit more from the intervention. From the achievement data, we identified black and first-generation students as populations who responded most strongly to the treatment. Patterns in behaviors and attitudes were apparent for one of these populations (black students) and not the other (first-generation students).

The response of black students on our survey questions differed from other students in the class in three ways. First, under both classroom structures, black students were more likely to report that the homework contributed to their learning in the course, and there was a trend for black students more than any other student groups to report that they valued the skills they developed from this class more than other students. Second, black students perceived the class to require more higher-order skills. These results imply that these students had a greater need for the kind of guidance provided by instructor-designed assignments. Thus, the addition of more homework and more explicit practice may have had a disproportionate impact on these students' achievement. Third, black students were significantly less likely than other students to speak up in class, but this disparity disappeared under moderate structure. We suspect that the increased sense of the classroom as a community may have contributed to this increased participation.

Although first-generation students did not differ in how they responded to survey questions versus continuing-generation students, they could still differ in how valuable the changes in the course were to them. In particular, the increased sense of community that seemed to correlate with the implementation of moderate structure could have helped them disproportionately, as has been demonstrated in a previous study (Stephens *et al.*, 2012). In addition, although students grouped in the category first generation share some characteristics, they are also very different from one another in terms of culture, background, and the barriers they face in the classroom (Orbe, 2004; Prospero *et al.*, 2012). For example, in our university setting, 55% of first-generation students have parents with low socioeconomic status and 50% transfer in from community colleges. The variation in students could thus obscure any patterns in their responses. Future analyses will attempt to distinguish subpopulations to identify patterns potentially hidden in our analysis.

Limitations of This Work

One of the major purposes of this article is to recognize that classroom interventions that work in one classroom may not work in others because 1) student populations differ in how they respond to classroom treatments, and 2) instructors do not always implement the critical elements of an active-learning intervention. Thus, it is important for us to note that, although we have shown that increased structure can work with both majors and nonmajors and with students from a

range of racial and ethnic groups, we are still working in an R1 setting. More work needs to be done to establish the effectiveness of the increased course structure intervention in community college or comprehensive university settings (although the evidence that it works well for first-generation students is a good sign that it could transfer). In addition, this study was with one instructor, thus we can now say increased course structure has worked for two independent instructors (the instructor of the current course and the instructor of the original course; Freeman *et al.*, 2011), but further work is necessary to establish its general transferability.

In addition, this study has suggested two factors by which increased course structure seems to be working by 1) encouraging distributed practice with a focus on class preparation and 2) helping students view the class as more of a community. Yet these are only two of many possible hypotheses for how this intervention works. It is possible that assigned preparatory assignments and small-group work to encourage community are not the only elements critical for this intervention's success. Further studies could explore how to best implement activities in class or the impact of adding graded review assignments on achievement.

Implications for Instructor and Researcher Best Practices

As a result of implementing an increased course structure and examining student achievement and survey results, we identified the following elements critical for student success and the success of future implementations:

1. Students are not a monolithic group. This result is not surprising. Students vary in many ways, but currently we do not know much about the impact of these differences on their experience with and approach to a college-level course. Future studies on student learning should disaggregate the students involved in the study (if possible), so instructors looking to implement an intervention can determine whether, and potentially how well, a particular intervention will work for their population of students.
2. Accountability is essential for changing student behaviors and possibly grades. We found that without accountability, students were not reading or spending many hours each week on the course. With weekly graded preparatory homework, students increased the frequency of both behaviors. We did not provide them credit for reviewing each week, and we found the overall frequency of this behavior decreased (even though our results demonstrate that students who did review notes performed better).
3. Survey questions are a useful method of identifying what behaviors an instructor might target to increase student performance. From our survey results, it seems that creating weekly review assignments might increase the frequency that students review their notes and thus increase their grades. Without the survey, we would not have known which behaviors to target.

Overall, this work has contributed to our understanding of who is most impacted by a classroom intervention and how those impacts are achieved. By looking at the achievement of particular populations, we can begin to change our teach-

ing methods to accommodate diverse students and possibly increase the effectiveness of active-learning interventions.

ACKNOWLEDGMENTS

This work was supported by NSF DUE 1118890 and DUE 0942215 and a Large Course Redesign Grant/Lenovo Instructional Innovation Grant from the Center for Faculty Excellence at University of North Carolina (UNC). This study was conducted under the guidelines of UNC IRB Study 11-1752. We acknowledge support from the Center for Faculty Excellence at UNC, specifically Erika Bagley and Andrea Reubens, for getting the project off the ground with initial statistical analysis, and Bob Henshaw for his help with the collection of survey data. We also thank Mercedes Converse, Alison Crowe, Scott Freeman, Chris Lenn, Kathrin Stanger-Hall, and Mary Pat Wenderoth for their thoughtful contributions to early versions of the manuscript.

REFERENCES

- 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.
- Antoine G, Harrell F (2000). Ordinal response regression models in ecology. *J Veg Sci* 11, 617–626.
- Barlow AE, Villarejo M (2004). Making a difference for minorities: evaluation of an educational enrichment program. *J Res Sci Teach* 41, 861–881.
- Beichner RJ, Saul JM, Abbott 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: *Research-based Reform of Introductory Physics*, ed. E Redish and P Cooney. www.compadre.org/per (accessed 10 August 2013).
- Bodenhofer U, Klawonn F (2008). Robust rank correlation coefficients on the basis of fuzzy orderings: initial steps. *Mathware Soft Comput* 15, 5–20.
- Born WK, Revelle W, Pinto LH (2002). Improving biology performance with workshop groups. *J Sci Educ Technol* 11, 347–365.
- Borrego M, Culter S, Prince M, Henderson C, Froyd JE (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses. *J Eng Educ* 102, 294–425.
- Brownell SE, Kloser MJ, Fukami T, Shavelson RJ (2013). Context matters: volunteer bias, small sample size and the value of comparison groups in the assessment of research-based undergraduate introductory biology lab courses. *J Microbiol Biol Educ* 14, 176–182.
- Burchfield CM, Sappington J (2000). Compliance with required reading assignments. *Teach Psychol* 17, 58–60.
- Caldwell J (2007). Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ* 6, 9–20.
- Carpenter D, Ramirez A, Severn L (2006). Gap or gaps: challenging the singular definition of the achievement gap. *Educ Urban Soc* 39, 113–127.
- Carini RM, Kuh GD, Klein SP (2006). Student engagement and student learning: testing the linkages. *Res High Educ* 47, 1–32.
- Cavanagh M (2011). Students' experiences of active engagement through cooperative learning activities in lectures. *Act Learn High Educ* 12, 23–33.
- Cheryan S, Plaut V, Davies P, Steele C (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *J Pers Soci Psychol* 97, 1045–1060.
- Crowe A, Dirks C, Wenderoth M (2008). Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.

- Delpit L (2006). *Other People's Children: Cultural Conflict in the Classroom*, New York: New Press.
- Demby G (2013). Latin@ offers a gender neutral choice; But how to pronounce it? www.npr.org/blogs/thetwo-way/2013/01/07/168818064/latin-offers-a-gender-neutral-choice-but-how-to-pronounce-it (accessed 25 May 2014).
- Dirks C, Cunningham M (2006). Enhancing diversity in science: is teaching science process skills the answer? *Cell Biol Educ* 5, 218–226.
- Dunlosky J, Rawson K, Marsh EJ, Mitchell JN, Willingham DT (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychol Sci Public Interest* 14, 4–58.
- Eddy SL, Crowe AJ, Wenderoth MP, Freeman S (2013). How should we teach tree-thinking? An experimental test of two hypotheses. *Evol Educ Outreach* 6, 13.
- Eley MG (1992). Differential adoption of study approaches within individual students. *High Educ* 23, 231–254.
- Ellis K (2004). The impact of perceived teacher confirmation on receiver apprehension, motivation and learning. *Comm Educ* 53, 1–20.
- Etkina E, Gibbons K, Holton BL, Horton GK (1999). Lessons learned: a case study of an integrated way of teaching introductory physics to at-risk students at Rutgers University. *Am J Phys* 67, 810–818.
- Fassinger PA (2000). How classes influence students' participation in college classrooms. *J Class Inter* 35, 38–47.
- Fox J, Weisberg S (2010). *An R Companion to Applied Regression*, Los Angeles: Sage.
- Fox M, Sonnert G, Niiforova I (2009). Successful programs for undergraduate women in science and engineering: adapting versus adopting the institutional environment. *Res High Educ* 50, 333–353.
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). End of lecture: active learning increases student performance across the STEM disciplines. *Proc Natl Acad Sci USA* 111, 8410–8415.
- Freeman S, Haak D, Wenderoth M (2011). Increased course structure improves performance in introductory biology. *CBE Life Sci Educ* 10, 175–186.
- Freeman S, O'Connor E, Parks JW, Cunningham M, Hurley D, Haak D, Dirks C, Wenderoth MP (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* 6, 132–139.
- Fullilove RE, Treisman PU (1990). Mathematics achievement among African American undergraduates at the University of California, Berkeley: an evaluation of the mathematics workshop program. *J Negro Educ* 59, 463–478.
- Haak DC, HilleRisLambers J, Pitre E, Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science* 332, 1213–1216.
- Handelsman J, Miller S, Pfund C (2007). *Scientific Teaching*, New York: W.H. Freeman.
- Harackiewicz JM, Canning EA, Tibbetts Y, Giffen CJ, Blair SS, Rouse DI, Hyde JS (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *J Educ Psychol* 106, 375–389.
- Hewlett SA, Buck Luce C, Servon LJ, Sherbin L, Shiller P, Sosnovich E, Sumburg K (2008). *The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology*, New York: Center for Work-Life Policy.
- Hitt GW, Isakovic AF, Fawwaz O, Bawa'aneh MS, El-Kork N, Makkiyil S, Qattan IA (2013). Secondary implementation of interactive engagement teaching techniques: choices and challenges in a Gulf Arab context. <http://arxiv.org/abs/1302.7093> (accessed 14 March 2014).
- Hoffman M, Richmond J, Morrow JA, Salomone K (2002). Investigating "sense of belonging" in first-year college students. *J Coll Stud Retent* 4, 227–256.
- Hooker DDT (2010). A study of the effects of the implementation of small peer led collaborative group learning on students in developmental mathematics courses at a Tribal Community College. PhD dissertation, Bozeman: Montana State University.
- Hurtado S, Carter DF (1997). Effects of college transition and perceptions of the campus racial climate on Latino college students' sense of belonging. *Sociol Educ* 70, 324–345.
- Johnson AC (2007). Unintended consequences: how science professors discourage women of color. *Sci Educ* 91, 805–821.
- Johnson DW, Johnson RT, Smith KA (1998). Cooperative learning returns to college: what evidence is there that it works? *Change* 30, 26–35.
- Keeler CM, Steinhorst RK (1995). Using small groups to promote active learning in the introductory statistics course: a report from the field. *J Stats Educ* 3, 1–8.
- Kim HS (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *J Pers Soc Psychol* 83, 828–842.
- Kim HS (2008). Culture and the cognitive and neuroendocrine responses to speech. *J Pers Soc Psychol* 94, 32–47.
- Lee P (2011). The "Asian" category in MCAS achievement gap tracking: time for a change. *Harvard Kennedy School's Asian American Policy Review* 21, 19–27.
- Maton KI, Hrabowski FA, Schmitt CL (2000). African American college students excelling in the sciences: college and postcollege outcomes in the Meyerhoff Scholars Program. *J Res Sci Teach* 37, 629–654.
- Matsui J, Liu R, Kane CM (2003). Evaluating a science diversity program at UC Berkeley: more questions than answers. *Cell Biol Educ* 2, 117–121.
- McIntyre SH, Munson JM (2008). Exploring cramming: student behaviors, beliefs, and learning retention in the principles of marketing course. *J Market Educ* 3, 226–243.
- Michaels JW, Miethe TD (1989). Academic effort and college grades. *Soc Forces* 68, 309–319.
- Minchella DJ, Yazvac CW, Fodrea RA, Ball G (2002). Biology resource seminar: first aid for the first year. *Am Biol Teach* 64, 352–357.
- National Science Foundation (2013). Widening implementation and demonstration of evidence-based reforms (WIDER). Division of Undergraduate Education NSF, 13–552.
- Orbe MP (2004). Negotiating multiple identities within multiple frames: an analysis of first-generation college students. *Comm Educ* 53, 131–149.
- Padgett RD, Johnson MP, Pascarella ET (2012). First-generation undergraduate students and the impacts of the first year of college: additional evidence. *J Coll Stud Dev* 53, 243–266.
- Peterfreund AR, Rath KA, Xenos SP, Bayliss F (2007). The impact of supplemental instruction on students in STEM courses: results from San Francisco State University. *J Coll Stud Retent* 9, 487–503.
- Pike N (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods Ecol Evol* 2, 278–282.
- Poelzer GH, Zeng L (2008). A teaching intervention to increase achievement of Hispanic nonscience majors taking physical science courses. *J Coll Sci Teach* 38, 59–64.
- Preszler RW (2009). Replacing lecture with peer-led workshops improves student learning. *CBE Life Sci Educ* 8, 182–192.
- Prospero M, Russell AC, Vohra-Gupta S (2012). Effects of motivation on educational attainment: ethnic and developmental

- differences among first-generation students. *J Hisp High Educ* 11, 100–119.
- Rath KA, Peterfreund AR, Xenos SP, Bayliss F, Carnal N (2007). Supplemental instruction in introductory biology I: enhancing the performance and retention of underrepresented minority students. *CBE Life Sci Educ* 6, 203–216.
- Reardon SF, Galindo C (2009). The Hispanic-white achievement gap in math and reading in the elementary grades. *Am Educ Res J* 46, 853–891.
- Savani K, Alvarez A, Mesquita B, Markus HR (2013). Feeling close and doing well: the prevalence and motivational effects of interpersonally engaging emotions in Mexican and European American cultural contexts. *Int J Psychol* 48, 682–694.
- Sharma MD, Khachan J, Chan B, O’Byrne J (2005). An investigation of the effectiveness of electronic classroom communication systems in large lecture classes. *Austral J Educ Tech* 21, 137–154.
- Singer SR, Nielsen NR, Schweingruber HA (2012). *Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Washington, DC: National Academies Press.
- Stephens NM, Fryberg SA, Markus HR, Johnson CS (2012). Unseen disadvantage: how American universities’ focus on independence undermines the academic performance of first-generation college students. *J Pers Soc Psychol* 102, 1178–1197.
- Strayhorn TL (2011). Traits, commitments, and college satisfaction among black American community college students. *Community Coll J Res Pr* 35, 437–453.
- Tsui L (2002). Fostering critical thinking through effective pedagogy: evidence from four institutional case studies. *J High Educ* 73, 740–763.
- Tsui L (2007). Effective strategies to increase diversity in STEM fields: a review of the research literature. *J Negro Educ* 76, 555–581.
- van Wyk MM (2012). The Effects of the STAD-Cooperative Learning Method on student achievement, attitude and motivation in economics education. *J Soc Sci* 33, 261–270.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, 4th ed., New York: Springer.
- Walton GM, Cohen GL (2007). A question of belonging: race, social fit, and achievement. *J Pers Soc Psychol* 92, 82–96.
- Walton GM, Cohen GL, Cwir D, Spenser SJ (2012). Mere belonging: the power of social connections. *J Pers Soc Psychol* 102, 513–532.
- Wilson RE, Kittleson J (2013). Science as a classed and gendered endeavor: persistence of two white female first-generation college students within an undergraduate science context. *J Res Sci Teach* 7, 802–825.
- Wischusen SM, Wischusen EW (2007). Biology intensive orientation for students (BIOS): a biology “boot camp.” *CBE Life Sci Educ* 6, 172–178.