The Journal of Experimental Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

Mind the gap: how a large-scale course re-design in economics reduced performance gaps

Dana Miller-Cotto & Christian Schunn

To cite this article: Dana Miller-Cotto & Christian Schunn (2020): Mind the gap: how a large-scale course re-design in economics reduced performance gaps, The Journal of Experimental Education, DOI: 10.1080/00220973.2020.1805717

To link to this article: https://doi.org/10.1080/00220973.2020.1805717

Published online: 12 Sep 2020.

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2020.1805717

Routledge
Taylor & Francis Group

LEARNING, INSTRUCTION, AND COGNITION

Mind the gap: how a large-scale course re-design in economics reduced performance gaps

Dana Miller-Cotto and Christian Schunn

University of Pittsburgh, Pittsburgh, PA

ABSTRACT

Large lecture classes in higher education continue to be a context in which large performance differences between underrepresented minorities and their White and Asian peers are observed. In the current study, we sought to develop a package of interventions that may reduce this gap in a multi-section Micro Economics course. The focus of this intervention was to re-design instruction in students' recitation sections, while also iteratively training teaching assistants on their instruction during recitation sections. Participants ($N = 2{,}679$) who were enrolled in Micro Economics worked together in groups where teaching assistants facilitated their learning in their recitation section. Results indicated that while all students demonstrated higher Micro Economics grades after the course transformation than their peers in prior semesters, this was particularly the case for underrepresented minorities, essentially eliminating the performance gap observed in prior semesters. Findings highlight the importance of instructional training for teaching assistants and employing teaching practices that can promote engagement and potentially promote inclusion during the learning process.

KEYWORDS

Active learning; achievement gap; economics education; large lecture classes; TA training

NATIONAL CENTER OF EDUCATIONAL Statistics indicates that underrepresented minorities (URMs), on average, are 17% less likely to complete college than their White and Asian peers, with Black students being 24% less likely and Latinx students being 10% less likely. Potential reasons for these findings can be attributed to large inequities in educational opportunities in earlier grades, leading to less preparation for college-level work, evidenced by study skills or lower mathematics performance, high school GPA or the SAT Math scores (Atuahene & Russell, 2016; Zwick & Sklar, 2005). However, such explanations are less relevant within more selective colleges because those institutions have often removed the least prepared students (Carnevale et al., 2018). Further, there are still noticeable performance differences in course grades and persistence in college between URMs and non-URMs even after controlling for these academic preparation factors (Carnevale et al., 2018). Therefore, scholars have now begun to attend to the learning environment at the university and the ways they may not be conducive to learning for all students (Aragón et al., 2017; Chesler & Young, 2013; Considine et al., 2014). To this end, we examined an intervention that substantially changed the learning environment and therefore had the potential to close course performance gaps within one mathematics-intensive, large-enrollment lecture course, Introduction to Micro-Economics.

CONTACT Miller-Cotto Dana dana.miller-cotto@ucsf.edu Learning Research Development Center, University of Pittsburgh, Pittsburgh, PA
Dana Miller-Cotto is now at the University of California San Francisco.

Understanding the U.S. Educational Landscape

In the United States, URMs generally do not perform as well as White and Asian students in science and mathematics, with URMs demonstrating lower proficiency in mathematics course grades, science course grades, and the SAT Math relative to their White and Asian peers. For example, at the end of high school, only 5% of Black students and 9% of Hispanic students were proficient in Science on the 2015 National Assessment of Educational Progress, in contrast to 38% of Asian students and 29% of White students (Snyder, & Dillow, 2015). Asian and White students are approximately 20% more likely to graduate with a Bachelors within 6-years than are Black and Hispanic students, and over 10% more likely to earn a STEM degree even among the graduates. Large differences in obtained courses grades are regularly seen (for an example, see Salehi et al., 2019 for a half-letter grade difference for URM students observed in introductory physics grades).

When students' course performance at both high school and university levels is lower than expected for URMs and students from lower socioeconomic backgrounds, this can often be explained by a lack of resources (Calarco, 2018; Darling-Hammond, 2006; Salehi et al., 2019) which may include both early learning opportunities and the opportunity to engage with challenging work (TNTP, 2018). These prior differences in experiences produce differences in foundational academic skills, often measured indirectly in terms of high school GPA or SAT-like measures, influence the extent to which students are fully prepared for college level work (Hodara & Lewis, 2017). But in more selective universities, admitted students generally have high levels of foundational academic skills (Carnevale et al., 2018), arguing against such explanations for observed course performance gaps in those contexts. Further, although sometimes significantly reduced (Salehi et al., 2019), there often remain course performance gaps by race/ethnicity within these selective universities even when minor differences in prior academic achievement are statistically controlled (Carnevale et al., 2018).

Despite efforts to remedy this profound problem, such as placement in remedial course work before entering mainstream courses (Kulik, Kulik, & Schwalk, 1983), counseling, study group, among numerous initiatives (Snyder, Fong, Painter, Pittard, Barr, & Pattall, 2019), the effect sizes for these interventions tend to be small, and these ethnic/racial performance gaps persist. Developing solutions to this problem requires large-scale changes to learning environments to better facilitate student learning for all, including making high-quality learning opportunities available to students from various racial and ethnic backgrounds that challenge them in numerous ways.

Techniques Used to Reduce Course Performance Gaps

Interventions that encourage students to engage in analytical thinking, adjust their misconceptions about the content, and encourage students to explain their answers or procedures to their peers may improve performance. Active learning, in particular, has been a central focus in reforms designed to increase engagement and inclusion and has demonstrated promising results for URMs (Theobald et al., 2020).

Active learning is defined as learning opportunity where students participate in meaningful activities that encourage them to engage with the learning (Slavich & Zimbardo, 2012). What makes active learning useful is that these activites occur in the classroom with instructors facilitating students learning process (Andrews et al., 2011; Slavich & Zimbardo, 2012). Active learning activities may include small group discussions, think-pair-share, or brief writing exercises used to preface another activity (DeLozier & Rhodes, 2017). Some argue that the active learning activities are particularly effective when done in a group because it forces students to employ self-explanation, which is believed to make their knowledge more explicit (Haak et al.,

2011; Roy & Chi, 2005; Slavich & Zimbardo, 2012) or to revise previously ill-conceived ideas (Allen, McNamara, McCrudden, 2015). These approaches may be particularly beneficial for engaging students across different backgrounds to connect with their peers to solve problems in groups.

Problem solving in groups during large lecture classes has been linked to improved exam performance, attendance, and overall course satisfaction (Stockwell et al., 2015). At a general level, meta-analysis findings (Freeman et al., 2014) suggest that students in active learning classrooms are 1.5 times less likely to fail than are students in a traditional lecture courses. Surprisingly, this meta-analysis did not examine whether particular subgroups benefited more or less from active learning. However, Haak and colleagues (2011) examined a large dataset of students' records from many thousands of students in large introductory biology courses at one large university. When comparing outcomes for traditional lecture courses to courses that employed daily problem solving, data analysis, and opportunities to practice reasoning skills, the gap in grades between students from lower socioeconomic backgrounds and higher socioeconomic backgrounds was reduced by 45%. This study differed from the current one in that it focused primarily on socio-economic status rather than relative representation for URMs in a given domain.

So why might active learning be especially useful for URMs? First, in a general sense, active learning often includes numerous learning strategies that are effective for everyone: having opportunities for practice, creating a classroom culture where students feel comfortable making mistakes, and discussing reasoning with peers. Ellis (2004) argued that changing classroom culture to being more community-oriented may allow students to take more risks, even despite the fear of confirming negative stereotypes about their group, and engage in challenging thinking and problem solving. Being in a smaller group may help students develop a collaborative perspective of their classroom and view their peers as a source of social support. Indeed, Eddy and Hogan (2014) found that working in small groups was particularly beneficial for Black students relative to other students. Thus, working in small groups may be a crucial component of active learning that drives URM students' engagement in the class. Further, another possibility is that new types of practice instead of drilling may be good learning opportunities for students to increase higher-order thinking that they did not have access to prior. Schools that serve mostly URM students often encourage rote memorization and drilling (TNTP, 2018), a strategy linked to poor reasoning ability and therefore poor preparation for college coursework.

Overall, it seems that learning in groups may be beneficial in courses that tend to be more challenging such as foundational economics courses, though where this group learning should occur is unclear. Interestingly, much of the prior research on addressing performance gaps in large introductory courses through active learning has focused on reforming the large lectures. Another place of reform could be recitation sections that often accompany the large lectures. On the one hand, recitation sections are an ideal place for reform because active learning and group work is easier to support in small class sizes (Freeman et al., 2014; Wright et al., 2019). Further, the interaction between students and a teaching assistant is more frequent/personal, potentially for good (e.g., through additional learning supports) or bad (e.g., via micro-aggressions). On the other hand, recitations usually occupy less time per week than do lectures, and require coordinating a large number of teaching assistants to implement a reform. Many departments do require graduate students to engage in training or teaching pedagogy courses prior to serving as recitation TAs. However, the amount of training typically provided may not be enough to address the complexities of active learning strategies, particularly in disciplines/departments where many TAs are themselves have little experience as students with active learning. Given the significant opportunities and challenges associated with reforming recitation sections, it is important to examine reforms that target recitations.

The Current Study

In the current study, we tested the effect of active and collaborative learning approaches in a large-scale recitation-section re-design in Economics. We compared the semester when the changes were first implemented to the prior semesters before the course transformation began to assess the effectiveness of the course transformation on students' grades in the Economics course. Economics is a mathematically-oriented discipline with relatively high rates of belief that brilliance is required for success (Leslie et al., 2015; Meyer, Cimpian, & Leslie et al., 2015). Unfortunately, there is a profound history of URMs being stereotyped as less competent at mathematics, often offering these stereotypes as an explanation for their lower performance than by their White and Asian peers (McGee, 2013, 2018; McGee & Martin, 2011) even though there are not inherent differences in mathematical competence between groups. Thus, it is possible that being in courses that require a fair amount of mathematics may make such stereotypes about a students' ethnic or racial group more apparent. It is also possible that for some URM students who are aware of this stereotype threat, a large enrollment Economics course would be a course where students can easily fade in to the background to avoid confirming negative stereotypes about their ethnic/racial group. We hypothesized that we would see an increase in grades for all students in the course transformation from prior semesters, and this increase would be particularly evident for URM students in Economics.

Method

The Micro Economics Course

Introduction to Microeconomic Theory (henceforth "Micro") is typically the most widely taught introductory course in Economics. This high enrollment course ($\sim$1,700 – 1,800 students per year) at the University of Pittsburgh fulfills a general education requirement for students across the college of Arts & Sciences and is also required for both Economics majors and Business Administration majors. Micro is offered every Fall and Spring semester, with four sections offered in the Fall and at least two offered in the Spring. Several instructors teach the various sections. The course includes a traditional lecture format with two lectures each week taught by faculty and a one recitation meeting per week taught by a graduate student teaching assistant (TA). Each course section has a typical capacity of 300 students with 10 recitations of 30 students. Instructors can be assigned anywhere from two to five TAs who are then each assigned multiple recitation sections.

The content focuses on how a free market system resolves the basic social questions of what goods and services to produce, how scarce resources are organized to produce goods and services, and to whom goods and services are distributed once they are produced. Relatedly, course objectives include students having a deeper understanding of how scarcity and incentives influence the actions taken by households, firms, and governments, with emphasis on how these actions are coordinated by markets through price adjustments that balance supply and demand. A variety of economic and social issues are examined related to the nature and impact of several forms of market failure and the potential impact of government intervention to address such failures.

The course includes homework assignments (both paper and electronic), short quizzes, and three exams which are typically a combination of multiple-choice questions. Through these assessments and activities, instructors seek to instill problem solving and analytical skills to prepare students for careers in industry, government, and nonprofits, or graduate programs in economics, law, medicine, and business. In the past, recitations were focused on helping students complete homework activities and answer questions about lecture content.

Recitation Re-Design

Over the summer of 2018, to prepare for the course transformation, a large group of faculty from the Department of Economics collaborated on the re-design of recitation sections. Recitations were re-designed to focus on conceptual understanding and problem solving, and pedagogically to use groupwork and guided discussion rather than focus on worksheets and homework help. The general motivation for the redesign, as shown in the internal grant proposal that funded this effort, from the perspective of the faculty involved in the transformation, was to improve overall learning outcomes (primary goal) and improve retention of women in the field of economics. Further, the faculty were generally unaware of the relative performance by race/ethnicity in these courses as those kinds of analyses were not usually done; the current study's authors, as independent researchers, chose the current research question based on their own prior research. It is also possible that if the Hawthorne effect were present, we would also have observed an improvement in course performance for all students; we did not observe this. Thus, it is unlikely that a Hawthorne effect (i.e., expectations regarding the effects of the intervention) would be the cause of any observed effects.

TAs, given their central role in implementing the recitation intervention, were given training by staff from the University's Center for Teaching and Learning (CTL). CTL Staff reviewed and provided feedback on revised recitations plans, as well as offered a series of workshops for Teaching Assistants at the beginning of the term and as part of the curriculum in the existing "Teaching Economics" course, taught by the director of graduate studies. Workshop sessions included "Teaching Interactively," "Dealing with Difficult Situations," and "Teaching with Case Studies." CTL staff conducted video-recordings and observations of recitation sections to assess consistency of implementation of recitations as planned and to provide follow-up consultations with teaching assistants; additionally, CTL staff attended weekly meetings of faculty with TAs to assess standardization of communication of instructions across sections. Throughout the term and at its conclusion, CTL staff consulted regularly with course directors to communicate results of these endeavors.

Fall 2018, the first semester of implementation, was treated as the treatment condition, and we compared students in this treatment to students in three prior semesters: Fall 2015, Fall 2016, and Fall 2017. Students who took Economics in the Fall as opposed to the Spring or Summer could be different in a number of ways (e.g., an additional semester of experience overall and with university level mathematics in particular), and thus the prior Fall semesters served as a better matched comparison group. There were a total of 10 recitation sections for Fall 2018 (treatment semester), similar to the number of recitations in the prior Fall semesters. Note that one instructor in Fall 2018, new to the university, did not implement the revised recitation plans, provided no information on how the recitations for this instructor were conducted, and was therefore excluded from analysis.

The typical structure of the revised recitations focused on conceptual understanding and problem solving, rather the prior recitation focus on fluent application of lecture content to many simple calculation or multiple-choice problems. Most recitations began with an interesting historical or recent example of a phenomenon, asking to students to think about possible explanations as a warm-up activity. For example, to introduce the concept of efficiency vs. equity, students were asked to consider the difference in table reservation policies between McDonalds, a moderately expensive café, and a very expensive restaurant. Sometimes there was a short reading related to one example of the phenomenon (e.g., an article about Uber surge pricing). Most recitation plans then presented two or three concrete data problems of growing complexity in the form of tables or graphs of data for students to analyze, using a sequence of scaffolded prompts related to basic calculations, more advanced calculations, and then conceptual interpretations.

Pedagogically the revision involved groupwork and guided discussion rather than using the time to TAs to demonstrate how to complete homework problems on the chalk board, or for

students to complete worksheets or ask for homework help on an individual basis. Thus, recitation sections involved working in groups and then engaging in whole-class discussion through a set of conceptually-rich micro economics problems. It was encouraged that students rely on one another in their group for hints on specific problems, and avoid asking the TA for help unless the group was stuck. Students were asked to work in the same groups each week so that they would form bonds with their peers that go beyond the recitation section, offering more opportunities for asking for help, motivation to be engaged during the recitation sections, and creating an affirming culture that may not always be inherent in active learning environments.

Because recitation participation and activities were graded, recitations were generally well attended. For example, in Fall 2018, average weekly attendance across recitation sections ranged between 86% and 100%; we determined attendance for recitations by response rates on course surveys assessing students' experiences in each recitation.

Sample

The sample in the current study consisted of 2,679 undergraduate students who were enrolled in Micro during the semester of the recitation re-design and the three fall semesters prior to the re-design (i.e., Fall semesters of 2015 to 2018) at the University of Pittsburgh. The University of Pittsburgh, according to the Carnegie Classification system, is a high research activity, doctoral degree granting university whose undergraduate profile is four-year, full-time, more selective, and lower transfer in, with an 59% acceptance rate. The dataset in the current study came from a larger project assessing academic outcomes of students enrolled at the university's main campus. Students in the recitation re-design semester were blind to the recitation changes prior to enrolling in the course, as well as during their participation in the course.

Measures

Demographics

Information regarding students' demographics were part of the student records at the University of Pittsburgh. This includes variables for gender, race/ethnicity, first-generation status. Each of these variables are indicator variables, with 1 indicating yes and 0 indicating no. URM was defined as whether a student self-identified as Black or Latinx; there were very small numbers of other race/ethnicities underrepresented in US universities.

In the current study, women made up 38% of the sample; in terms of race/ethnicity, 75% of the sample identified as White, 9% as Asian, 9% as URMs, and the remainder consisting of those either not listing a race/ethnicity (4%), listing another race/ethnicity (e.g., "other", 1%), or listing a combination of race/ethnicities (2%), and with 10% of the population having first-generation status. Samples across semesters have a comparable proportions of URM and non-URM students, as well as comparable proportions of women and men and for first-generation students.

Prior Achievement

The student records included several academic performance covariates, such as the students' scores in the three main components of the Scholastic Aptitude Test (SAT verbal, writing, and mathematics), as well as high-school grade point average (GPA). In the case of students who had taken the ACT instead of the SAT, scores were converted by the university into SAT units. We focused on our main analyses on SAT mathematics ($M = 660$, $SD = 67$) and high-school GPA ($M = 3.88$, $SD = 0.48$); preliminary analyses that suggested that the SAT mathematics and high-school GPA were strong positive predictors of Micro Economics grades. We constrained the sample to students with a high-school GPA on a scale from $1.6 - 5.0$, removing any outliers from the

Table 1. Sample demographics for each of the four cohorts, including number of students (and percent of cohort), Mean SAT Math, and Mean HS GPA.

	Control									Treatment		
	2015			2016			2017			2018		
	N (%)	SAT Math	HS GPA	*N* (%)	SAT Math	HS GPA	*N* (%)	SAT Math	HS GPA	*N* (%)	SAT Math	HS GPA
Women	287 (39%)	640	3.91	273 (37%)	641	3.96	242 (35%)	641	3.97	214 (39%)	648	4.04
Men	437 (61%)	671	3.79	455 (62%)	675	3.85	434 (64%)	670	3.83	336 (61%)	674	3.91
First-Gen	76 (10%)	660	3.83	70 (9%)	635	3.85	73(10%)	650	3.78	58 (10%)	630	3.93
URM	65 (9%)	624	3.58	56 (7%)	629	3.73	57 (8%)	632	3.73	59 (10%)	638	3.77
White	550 (76%)	657	3.87	550 (75%)	660	3.91	515 (76%)	659	3.91	411 (75%)	661	3.97
Asian	72 (10%)	703	3.84	80 (11%)	697	3.93	63 (9%)	698	3.89	59 (10%)	713	4.09

analysis ($N = 37$, 1.4%), as those scores were likely to be based on other scales (e.g., 1-10 scale used in India or 1-6 scale used in IB schools) and there were no indicators in the data for what GPA scale was being used.

Table 1 presents descriptive statistics for students enrolled in Micro, separated by those who enrolled during the years prior to the course transformation (e.g., 2015, 2016, 2017) and those from the year the course transformation took place, 2018. Table 1 also presents means in each cohort for these two key prior achievement variables. The students were relatively similar with regard to their SAT mathematics scores and HS GPAs across all semesters. However, preliminary analyses indicated that there was a significance difference on HS GPA (F [3, 1,444] = 7.64, $p <$ .00), indicating that HS GPA for students in the 2018 cohort was significantly different from the 2017 cohort (M_D = .07), the 2016 cohort (M_D = .07), and the 2015 cohort (M_D = .12). However, this pattern was similar within each student subgroup and thus could not explain differential effects by subgroup. We also did not observe a statistically significant difference on SAT Math across cohorts (F [3, 1,185] = .06, p = .587). Finally, we did observe a statistically significant difference between instructors (F [1, 757] = 10.54, $p <$ .01). However, instructors were the same pool of instructors who had taught in previous years, and thus there would not be a confound of instructor by condition.

Micro Economics Grade

To ascertain the effects of the recitation re-design, we used the academic record information to calculate students' grades in Micro Economics: A denotes a 4.0, B denotes a 3.0, C denotes a 2.0, D denotes a 1.0, and 0 denotes an F; a minus subtracts 0.25 and plus adds 0.25 (e.g., a C– is a 1.75). Grades that were not A – F (e.g., incomplete, withdrawal; $N = 58$, 2.2%) were treated as outliers and removed from the analysis. It should be noted that instructors made minimal changes to the course exams, since changing exams is a laborious task and was not part of the course re-design. Thus, despite the various changes to lesson plans and recitations, instructors kept their course exams the same from prior semesters, making the comparison before the course re-design and after particularly compelling. Here, we focus on course grade because it determines whether students may go on to more advanced coursework in Economics.

English Composition Grade

To assess whether students' enrolled in the Micro course part of the recitation demonstrated similar grade improvements and therefore suggestive of confounding cohort differences, we examined the grades obtained in English Composition (based on the same scoring system used for the Micro Grades, A - F), focusing only on the students who were concurrenty enrolled in both classes. In other words, only students enrolled in Micro in the target semesters were included in this analysis. We chose English Composition for comparison because English Composition is the only course taken by large numbers of students across all majors in their first year (although only

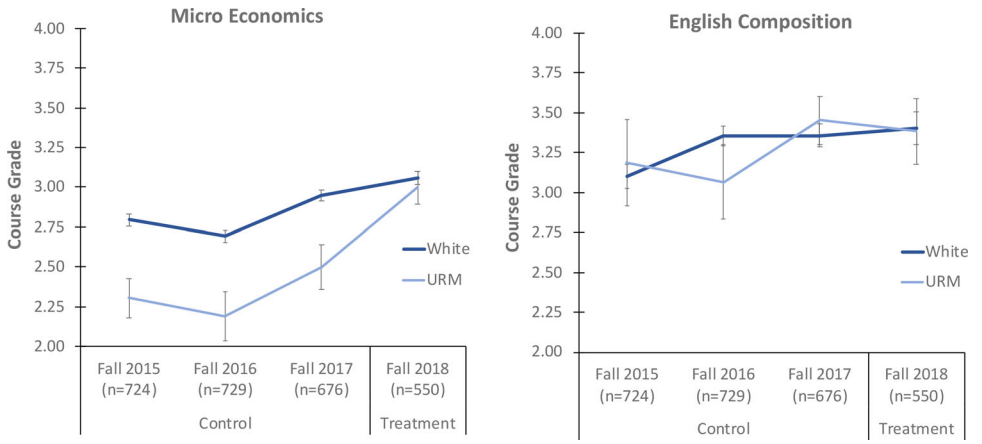

Figure 1. Mean (with SE Bars) Micro Economics and English Composition Grades (with SE bars and n information) by Semester and Race/Ethnicity. The treatment refers to the change in Micro Economics instruction; no change was implemented in English Composition instruction.

half doing so in the Fall), and thus not likely to be subject to further selective/interest-based enrollment confounds. Grades that were not A – F (e.g., incomplete, withdrawal; 1.9%) were treated as outliers and removed from the analysis.

Analysis

We first examined temporal trends across semesters by race/ethnicity in mean course grades (Micro and English Composition), SAT Math, and HS GPA. These descriptives were used to assess whether there were gradual temporal trends across cohorts or not, generally testing the plausibility of a cohort difference in 2018, and informing the choice of statistical modeling (e.g., regression discontinuity vs. simple control/treatment contrast). To foreshadow the results, a simple control/treatment contrast appeared to be the best choice due to an absence of prior temporal trends.

We then ran a series of ordinary least squares (OLS) regressions predicting Micro Grade performance to (1) ascertain the effect of the recitation re-design on students' end-of-semester grades when controlling for prior achievement (e.g., SAT Math and HS GPA), (2) to determine whether the effect of the recitation re-design was differentially beneficial for URMs. The first analysis involved a multiple linear regression predicting Economics Grade with URM (or not), condition (pre/post), the interaction of URM and condition, SAT Math, and HS GPA as predictors. Similar analyses were implemented for English Composition. Finally, we calculated the percentage of URM and non-URM students receiving particular letter grades before and during the course transformation semesters to determine whether any changes in performance were especially large at particular performance levels. For this analysis, minus and plus grades were rounded to the overall letter grade.

Results

Temporal Trends in Course Grades and Prior Academic Performance

Figure 1 presents course grades in Economics and English Composition over four semesters by ethnicity, focusing on the URM vs. White contrast, initially a conservative contrast given that Whites alone are the majority in this context and Asians can be a heterogenous group that can

Table 2. Prediction models (Unstandardized and Standardized Betas) by course transformation, URM status, HS GPA, SAT Math, and the interaction of course transformation and URM for micro grade.

	Model 1		Model 2	
	B	β	B	β
Course Transformation	0.21	.08**	0.17	.06*
URM	−0.08	-.02	−0.16	-.04*
High School GPA	0.67	.29**	0.67	.29**
SAT Math/100	0.003	.21**	0.003	.21**
URM x Course Transformation	*Not Tested*		0.36	.04*
R^2	19.1%		19.2%	

*= p < .05, **= p < .01.

include historically under-performing subgroups as well. It shows the pervasive racial/ethnic differences in performance in prior semesters in Micro Economics. Performance in English Composition is shown to examine the course-specificity of performance changes over time. There was no general trend for grades to gradually increase in either of the courses, suggesting that the intervention could be effectively modeled as a pre/post approach rather than a linear trend by year. Economics showed a pattern of change only in the intervention semester: the pervasive race/ethnicity gap is almost entirely erased. This pattern was not shown in English Composition (i.e., no pervasive White vs. URM gap prior to the intervention year, no general improvement in performance in the treatment year relative to the prior years), suggesting the difference was not a result of changes in cohorts of students or the general support programs offered by the university. We test these changes formally, as well as the interactions, in the next section. Further, remaining analyses focus on the URM vs. non-URM distinction since White and Asian students performed quite similarly in Micro-Economics.

Formal Statistical Tests of Changes in Grade

Table 2 displays the results of the multiple regression analyses of Micro grades. In the first model, testing the main effects of the course transformation and URM, controlling for high school GPA and SAT Math, the model revealed a significant positive main effect of the the course transformation on Micro grades ($\beta = .06$, $p = .04$). That is, students who participated in the the course transformation outperformed students from prior semesters before the course transformation was implemented. The second model, tests the moderation of the effect by URM status, finding that the main effect of the course transformation continues to be significant (i.e., overall benefits) and the significant moderation by URM status (i.e., significantly larger benefits for URM students), demonstrated in Figure 2. A similar multiple regression analysis for English Composition grades revealed no pre/post effects overall or interactions by race/ethnicity ($ps>.5$), arguing against cohort difference on other (unmeasured) background variables that affect academic performance. Gender and first-generation status were excluded from these analyses as there was no significant effect of these demographic variables on Micro Economics grades ($ps>.5$), and thus were also excluded from Table 2.

Figure 3 presents the changes in the percentage of students at each grade level, contrasting the course transformation semester with the prior Fall semesters. This direct contrast makes it easier to see relative change amounts by letter grade, by ethnicity, and their interaction. The figure also indicates whether the change in percentage for non-URMs and URMs significantly differed (for As χ^2 (1, $N=1{,}165$) = 4.64, $p < .05$; for Bs, χ^2 (1, $N=1{,}538$) = 2.74, $p < .1$; for Cs, χ^2 (1, $N=737$) = 4.78, $p < .05$; for Ds, χ^2 (1, $N=211$) < 1; for Fs, χ^2 (1, $N=218$) = 2.58, $p < .2$). As can be seen in this figure, there were sizable increases for both non-URMs and URMs, but noticeably greater performance improvements for URMs at both ends of the performance spectrum, although the noise levels are higher for the relatively rare Fs and Ds.

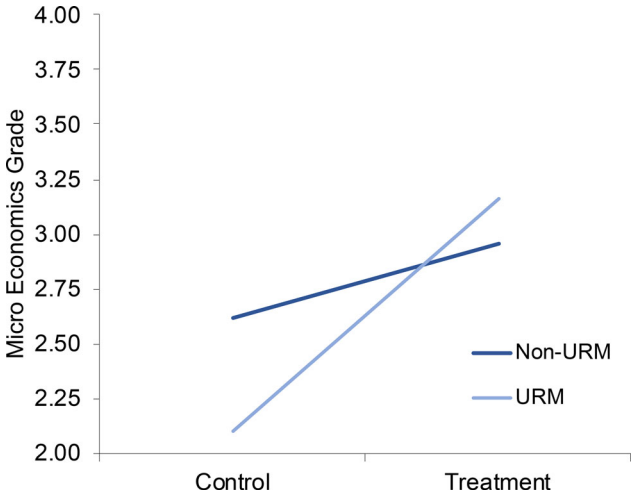

Figure 2. Adjusting Mean Final Micro Grade Controlling for Co-variates Before and During the Course Transformation for URMs and Non-URMs.

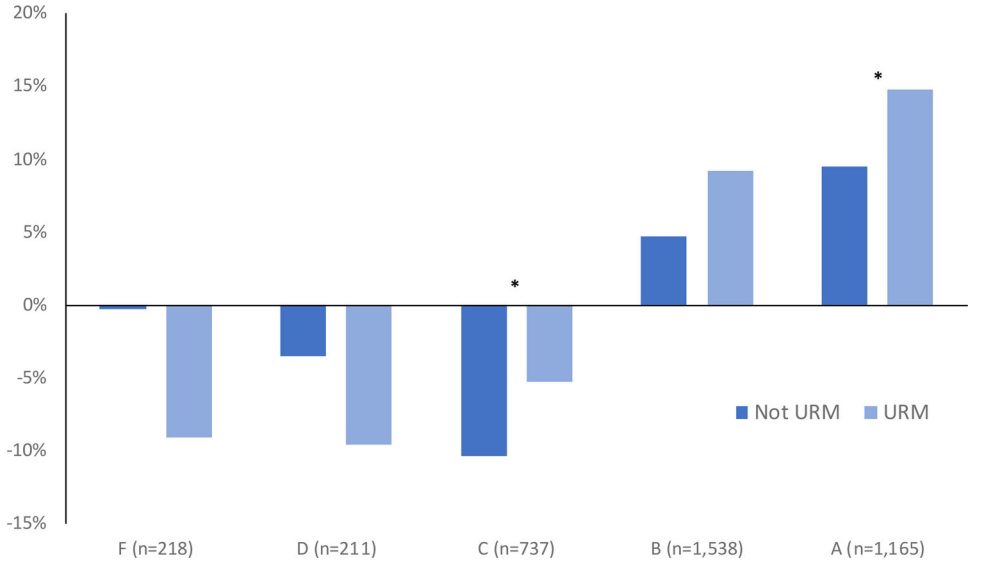

Figure 3. Percent Change in round Letter Grades (F through A) from the Course Transformation Relative to the Prior Semesters for Non-URM and URM Students (* = p <.05). Positive numbers indicate relatively higher percentages in the Course Transformation semester.

Discussion

Educators and education researchers have made many attempts to not only improve course grades in large lecture classes but have also attempted to increase inclusion for historically URMs in higher education. The present study suggests quite clearly that employing active learning strategies in recitation courses can improve overall grades while also greatly reducing performance differences by race/ethnicity. Even after taking into account SAT Math and HS GPA, we found that Micro grades were significantly better in the course transformation semester than they were in prior semesters before the course transformation was enacted. Students who identified as URMs demonstrated the most benefit from the course transformation, though a combination of

producing close to equal outcomes in the proportion of students receiving the highest grades and completely eliminating differences in the proportion of students receiving failing grades.

This study supports prior work showing positive effects of active learning on student learning and promoting learning gains (Slavich & Zimbardo, 2012). Some prior work suggests that working in small groups and creating a sense of community as opposed to competition increased inclusion for women in domains where they were historically under-represented (Boaler et al., 2011). Women are also often minoritized in similar various ways in mathematics-heavy disciplines; these prior findings related to how women benefit from group work may offer some insights as to why small group learning is also impactful for URMs in our study. Though little work has been done in the U.S. to promote learning and performance for URMs in domains where they are historically underrepresented, the current findings suggest that the package of interventions enacted here may be a hopeful route to improving inclusion as it related to race as well, not just gender. Our findings are also consistent with earlier recommendations for creating classroom communities where students work through problems together, critically analyze information, and work in teams where they can look to their peers for support (Ellis, 2004). These actions may potentially reduce the likelihood that URMs feel pressure to perform in certain ways while being aware of stereotypes about their race or ethnic group. Further, while most research has primarily focused on math and science and successfully promoted active learning strategies to produce favorable results (Freeman et al., 2014; Haak et al., 2011), two domains that also fail to promote inclusion and retain URMs, we have demonstrated that these strategies also generalize to a domain that also requires mathematical proficiency but is also often left out of the gap reducing and retention conversation.

We hypothesized that the combination of including small group work, teacher facilitation of group work and problem solving, and student building community with their groupmates would promote learning because these features may increase students' likelihood that they will interact with other students, especially for students who have structural reasons for not having a strong connection to other students within the classroom. Of course, more work must be done in this area to determine whether URMs truly felt like they belonged and it was indeed these features that were the source of the learning gains. Future studies in this space should measure belongingness or attitudes toward Economics to confirm these hypotheses and generally test attitudinal effects of the intervention, which can be equally important as grade differences since students can drop out of educational pathways if they do not feel competence even when receiving high grades (e.g., Witherspoon et al., 2019).

Among the strengths of this study were the large sample size and the variety of TAs employing the new recitation designs across multiple sections of Micro. Much prior work in this area attempted to make changes to classes that had much smaller sample sizes, attempted to employ active learning strategies within the large lecture section, or only employed these active learning strategies with only one instructor and in one class. Testing this kind of active learning in many recitation sections with many different TAs, the current study gives us greater confidence that this form of active learning will be particularly beneficial for URMs in others settings with TAs over varying backgrounds and teaching abilities.

Despite these strengths, the study is limited by the course transformation occurring during only the Fall semester. As noted earlier, Spring semester students tend to be inherently different from Fall semester students on a variety of factors (e.g. more time enrolled in college) and thus are not an analogous comparison group. Further, enrollment for Spring semester for this course tends to be about half the usual enrollment of Fall semesters, greatly reducing the power for examining differential effects of the intervention for URMs. Thus, combining data across years or institutions will be required to assess those effects.

We also acknowledge that our outcome measure, students' final Micro Economics course grade, has its challenges. In Economics, there were no available standardized assessments to

ascertain students' understanding of the material. Further, there was some variation in which exact content was covered in the lecture so instructors would be unwilling to use a shared exam even if there was one to be used. Course grade has challenges as a pure measure of learning, but it is a context-valid measure of performance: that which appears on the student's transcripts, and uniquely determines whether the students may go on to more advanced coursework in Economics.

We also did not collect information about implementation variations across different sections, which limits our ability to understand how such variation changes outcomes. Facilitated learning within groups was an important aspect of the recitation re-design, and prior research suggests that group size and heterogeneity are important features when assigning groups to complete work (Epsey, 2018; Yamarik, 2007). However, in the current study we did not collect information on how students were grouped within recitations. Although we generally found a benefit of group work, different effects may be observed depending upon such implementation details. Future work should examine such information to better understand the benefits of group assignment and the specific ways it might facilitate learning.

Another important issue is the treatment of Black and Latinx students as one homogeneous population. The historical and socio-political experiences of the two groups in the U.S. have many important differences, and others have found differential performance in some settings (Eddy & Hogan, 2014). However, these groups are regularly minoritized in similar ways in the U.S. that have similar negative effects, thus providing a rationale for combining them for analyses. A related consideration for future research based on larger samples is the intersection of race/ethnicity and gender. Prior research has sometimes found that URM women are especially likely to encounter adversity in the classroom (Johnson, 2011; Pittman, 2010). Unfortunately, the current study did not have the sample size to examine the intersection of race/ethnicity and gender, nor did we have the adequate sample size to break out Black and Latinx students in separate analyses for each, a common challenge in research conducted at selective universities.

A final concern about the current study is the use of a cohort comparison rather than random-assignment-to-condition design. In particular, the cohort design raises questions about equality of participants and other co-occurring supports. We were able to rule out possible differences in the strongest individual differences typically associated with performance in this course as well as show no general gains in another commonly taken course. It is also important to note that random assignment to condition is logistically challenging when implemented at sufficient scale to examine interactions by race/ethnicity.

Conclusion

Perhaps the most important lesson to be learned from the current study is that large lecture classes do not have to be completedly re-structured or downsized to small classes to make a big impact. Revision to the large lectures themselves is often challenged by problems of infrastructure: active learning depends upon having new classrooms which are expensive and rare in large universities. As we observed here, it is possible to make changes to large lecture classes through alternative routes, like changing the recitation sections. By doing this, we were able to create an environment where students could be more engaged and received direct feedback or direction from their TAs, consequently improving grades for students in the process. However, it is important to note that when so many eggs are placed in the TA basket, training for the TAs is likely to be very important (Allgood et al., 2018); we caution against implementing this kind of strategy without significant and sustained investment in TA training.

ORCID

Dana Miller-Cotto http://orcid.org/0000-0003-0577-4868
Christian Schunn http://orcid.org/0000-0003-3589-297X

References

Allen, L. K., McNamara, D. S., & McCrudden, M. (2015). Change your Mind: Investigating the Effects of Self-Explanation in the Resolution of Misconceptions. In CogSci.

Allgood, S., Hoyt, G., & McGoldrick, K. (2018). Teacher training for PhD students and new faculty in economics. *The Journal of Economic Education*, *49*(2), 209–219. https://doi.org/10.1080/00220485.2018.1438947

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, *10*(4), 394–405. https://doi.org/10.1187/cbe.11-07-0061

Aragón, O. R., Dovidio, J. F., & Graham, M. J. (2017). Colorblind and multicultural ideologies are associated with faculty adoption of inclusive teaching practices. *Journal of Diversity in Higher Education*, *10*(3), 201–215. https://doi.org/10.1037/dhe0000026

Atuahene, F., & Russell, T. A. (2016). Mathematics readiness of first-year university students. *Journal of Developmental Education*, *39*, 12.

Boaler, J., Altendorff, L., & Kent, G. (2011). Mathematics and science inequalities in the United Kingdom: when elitism, sexism and culture collide. *Oxford Review of Education*, *37*(4), 457–484. https://doi.org/10.1080/03054985.2011.595551

Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., Pianta, R., & NICHD Early Child Care Research Network (2011). Examining the Black-White achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development*, *82*(5), 1404–1420. https://doi.org/10.1111/j.1467-8624.2011.01620.x

Calarco, J. M. (2018). *Negotiating opportunities: How the middle class secures advantages in school*. Oxford University Press.

Carnevale, A. P., Van Der Werf, M., Quinn, M. C., Strohl, J., & Repnikov, D. (2018). Our Separate & Unequal Public Colleges: How Public Colleges Reinforce White Racial Privilege and Marginalize Black and Latino Students. Georgetown University Center on Education and the Workforce

Chesler, M.A. & Young, A. A. (Eds.). (2013). *Faculty identities and the challenge of diversity: Reflections on teaching in higher education*. Paradigm.

Considine, J. R., Mihalick, J. E., Mogi-Hein, Y. R., Penick-Parks, M. W., & Van Auken, P. M. (2014). Who am I to bring diversity into the classroom?" Learning communities wrestle with creating inclusive college classrooms. *Journal of the Scholarship of Teaching and Learning*, *14*(4), 18–30. https://doi.org/10.14434/v14i4.3895

Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher*, *35*(7), 13–24. https://doi.org/10.3102/0013189X035007013

DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational Psychology Review*, *29*(1), 141–151. https://doi.org/10.1007/s10648-015-9356-9

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE Life Sciences Education*, *13*(3), 453–468. https://doi.org/10.1187/cbe.14-03-0050

Ellis, K. (2004). The impact of perceived teacher confirmation on receiver apprehension, motivation, and learning. *Communication Education*, *53*(1) [Database] https://doi.org/10.1080/0363452032000135742

Epsey, M. (2018). Diversity, effort, and cooperation in team based learning. *The Journal of Economic Education*, *49*, 8–21.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *11*, 8410–8415.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science (New York, N.Y.)*, *332*(6034), 1213–1216. https://doi.org/10.1126/science.1204820

Hodara, M., & Lewis, K. (2017). How Well Does High School Grade Point Average Predict College Performance by Student Urbanicity and Timing of College Entry? REL 2017-250. Regional Educational Laboratory Northwest.

Johnson, D. R. (2011). Women of color in science, technology, engineering, and mathematics (STEM). *New Directions for Institutional Research*, 2011(152), 75–85. https://doi.org/10.1002/ir.410

Kulik, C-L., Kulik, J., & Schwalb, B. (1983). College programs for high risk and disadvantaged students: A meta-analysis of findings. *Review of Educational Research*, *53*, 397–414. https://doi.org/10.3102/00346543053003397

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science (New York, N.Y.)*, *347*(6219), 262–265. https://doi.org/10.1126/science.1261375

McGee, E. (2013). Young, Black, mathematically gifted, and stereotyped. *The High School Journal*, *96*(3), 253–263. https://doi.org/10.1353/hsj.2013.0011

McGee, E. (2018). Black genius, Asian fail": The detriment of stereotype lift and stereotype threat in high-achieving Asian and Black STEM students. *AERA Open*, *4*(4), 233285841881665. https://doi.org/10.1177/2332858418816658

McGee, E. O., & Martin, D. B. (2011). You would not believe what I have to go through to prove my intellectual value!" Stereotype management among academically successful Black mathematics and engineering students. *American Educational Research Journal*, *48*(6), 1347–1389. https://doi.org/10.3102/0002831211423972

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, *6*, 235. https://doi.org/10.3389/fpsyg.2015.00235

Oyserman, D., & Lewis, N. A. (2017). Seeing the destination AND the path: Using identity-based motivation to understand and reduce racial disparities in academic achievement. *Social Issues and Policy Review*, *11*(1), 159–194. https://doi.org/10.1111/sipr.12030

Pittman, C. T. (2010). Race and gender oppression in the classroom: The experiences of women faculty of color with white male students. *Teaching Sociology*, *38*(3), 183–196. https://doi.org/10.1177/0092055X10370120

Roy, M., & Chi, M. T. (2005). *The self-explanation principle in multimedia learning. The Cambridge handbook of multimedia learning*, 271–286.

Salehi, S., Burkholder, E., Lepage, G. P., Pollock, S., & Wieman, C. (2019). Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics. *Physical Review Physics Education Research*, *15*(2), 020114. https://doi.org/10.1103/PhysRevPhysEducRes.15.020114

Slavich, G. M., & Zimbardo, P. G. (2012). Transformational teaching: Theoretical underpinnings, basic principles, and core methods. *Educ Psychol Rev*, *24*(4), 569–608. https://doi.org/10.1007/s10648-012-9199-6

Snyder, T. D., & Dillow, S. A. (2015). Digest of Education Statistics 2013. NCES 2015-011. National Center for Education Statistics.

Snyder, K. E., Fong, C. J., Painter, J. K., Pittard, C. M., Barr, S. M., & Patall, E. A. (2019). Interventions for academically underachieving students: A systematic review and meta-analysis. *Educational Research Review*, *28*, 100294. https://doi.org/10.1016/j.edurev.2019.100294

Statistics 2013. NCES 2015-011. National Center for Education Statistics.

Stockwell, B. R., Stockwell, M. S., Cennamo, M., & Jiang, E. (2015). Blended learning improves science education. *Cell*, *162*(5), 933–936. https://doi.org/10.1016/j.cell.2015.08.009

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., … & Grummer, J. A. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences, 117*(12), 6476–6483.

TNTP (2018). The Opportunity Myth. The New Teachers Project.

Witherspoon, E. B., Vincent-Ruz, P., & Schunn, C. D. (2019). When making the grade isn't enough: The gendered nature of premed science course attrition. *Educational Researcher*, *48*(4), 193–204.

Wright, M. C., Bergom, I., & Bartholomew, T. (2019). Decreased class size, increased active learning? Intended and enacted teaching strategies in smaller classes. *Active Learning in Higher Education*, *20*(1), 51–62. https://doi.org/10.1177/1469787417735607

Yamarik, S. (2007). Does cooperative learning improve student learning outcomes?. *The Journal of Economic Education, 38*(3), 259–277.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, *42*(3), 439–464. https://doi.org/10.3102/00028312042003439