

Reimagining the Introductory Math Curriculum for Life Sciences Students

Erin Sanders O’Leary,^{†*} Hannah Whang Sayson,^{‡*} Casey Shapiro,^{§*}
 Alan Garfinkel,^{||} William J. Conley,[¶] Marc Levis-Fitzgerald,[§] M. Kevin Eagan,[#]
 and Blaire Van Valkenburgh^{@*}

[†]Center for the Advancement of Teaching Excellence, University of Illinois Chicago, Chicago, IL 60607; [‡]Center for Educational Assessment, Center for the Advancement of Teaching;

[§]Department of Integrative Biology & Physiology, [¶]Department of Mathematics, [#]Graduate School

of Education & Information Studies, and [@]Department of Ecology & Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095

ABSTRACT

Calculus is typically one of the first college courses encountered by science, technology, engineering, and mathematics (STEM) majors. Calculus often presents major challenges affecting STEM student persistence, particularly for students from groups historically underrepresented in STEM. For life sciences majors, calculus courses may not offer content that is relevant to biological systems or connect with students’ interests in biology. We developed a transformative approach to teaching college-level math, using a dynamical systems perspective that focuses first on demonstrating why students need math to understand living systems, followed by providing quantitative and computational skills, including concepts from calculus, that students need to build and analyze mathematical models representing these systems. We found that students who complete these new math courses perform better in subsequent science courses than their counterparts who take traditional calculus courses. We also provide evidence that the new math curriculum positively impacts students’ academic performance, with data that show narrowing of the achievement gap, based on students’ math grades, between student subgroups in the new math courses. Moreover, our results indicate that students’ interest in the concepts and skills critical to the quantitative preparation of 21st-century life sciences majors increases after completing the new contextualized math curriculum.

INTRODUCTION

First-year calculus courses have long presented challenges for recruitment and retention of students in science, technology, engineering, and mathematics (STEM) fields at the college level. Varying levels of prior knowledge, unwelcoming first-year calculus courses, and a lack of relevant real-world examples in those courses are major factors in students’ decisions to abandon a STEM major (Laursen *et al.*, 2011; President’s Council of Advisors on Science and Technology [PCAST], 2012). For life sciences students in particular, the quantitative and computational skills essential to modern biological research and biotechnology typically are not taught in first-year calculus courses (Marshall and Durán, 2018), and consequently students often view these classes as unpleasant and irrelevant hurdles to conquer in their quest for a degree in the biological sciences (Bialek and Botstein, 2004). These challenges are particularly problematic for students from disadvantaged backgrounds or social identity groups historically underrepresented in STEM, whose college-level calculus preparation may reflect inequitable access to high-quality mathematics instruction in high school (e.g., see Moore *et al.*, 2010; Benken *et al.*, 2015). These disparities can become exacerbated in first-year college math courses, manifested as gaps in academic performance outcomes, and can ultimately result in major barriers for such students to continue their studies in their intended STEM fields (PCAST, 2012).

Jennifer Knight, *Monitoring Editor*

Submitted Nov 4, 2020; Revised Aug 27, 2021;

Accepted Aug 31, 2021

CBE Life Sci Educ December 1, 2021 20:ar62

DOI:10.1187/cbe.20-11-0252

[†]These three authors contributed equally to this work.

Conflict of interest statement: A.G. is an author of the textbook *Modeling Life: The Mathematics of Biological Systems*, which supports the new math courses described in this study. No promotion of this product to the exclusion of other similar products should be construed.

*Address correspondence to: Blaire Van Valkenburgh (bvanval@ucla.edu).

© 2021 E. S. O’Leary, H. W. Sayson, C. Shapiro, *et al.* CBE—Life Sciences Education © 2021 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–Non-commercial–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.

Setting the Stage for a Transformational Approach to College-Level Math Instruction

It has long been recognized that mathematical approaches should be integral to undergraduate biology education (Gross, 2000), and multiple institutions have approached the development of quantitative competencies among life sciences majors in a variety of ways. For example, several institutions have implemented interdisciplinary courses with embedded activities that enable first-year students to practice their math skills in the context of learning biological concepts. These range from single quantitative science courses (Caudill *et al.*, 2010; Matthews *et al.*, 2010) to a multisequence series of courses in mixed lecture and laboratory settings (Depelteau *et al.*, 2010). Oftentimes, such courses necessitate multidisciplinary expertise and thus are team taught by biologists and mathematicians, collaborations that may not be possible across all institutions. In response to a call for students to engage in authentic coding experiences to enhance their quantitative skills in the context of biology (National Research Council [NRC], 2003), Matthews and colleagues incorporated computer programming into their first-year course (Matthews *et al.*, 2010). Survey data indicate students gained appreciation for math but not necessarily programming, which led the authors to question whether learning computational skills (e.g., coding) was too complex for first-year students (Robins *et al.*, 2003). This finding suggests additional research on effective approaches for introducing computer programming to first-year biology students is needed.

Another strategy that many institutions pursued in improving the quantitative curriculum for life sciences students was the development of biology-relevant calculus sections (biocalculus) within math departments already teaching traditional calculus for math, physical science, and engineering students (Duffus and Olfier, 2010; Usher *et al.*, 2010; Eaton and Highlander, 2017; Aikens *et al.*, 2021). For instance, Eaton and colleagues redesigned calculus to address the core competencies in *Vision and Change* (American Association for the Advancement of Science [AAAS], 2011) by including quantitative reasoning, modeling, simulation, and interdisciplinary collaborations, which they predicted would make the course more valuable to life sciences students. The authors compared biocalculus student academic outcomes with their peers in traditional calculus at two different institutions and showed similar levels of achievement for these two student populations at both institutions in content knowledge as measured by either the Calculus Concept Inventory (Epstein, 2013) or common quizzes assigned in all sections of the course (Eaton and Highlander, 2017). They also found that drop/fail/withdrawal (DFW) rates decreased over the duration of the study. Expanding upon this earlier study, Aikens and colleagues used pre–post surveys to examine attitudinal changes toward math as a function of its perceived value, or usefulness, in biology (Aikens *et al.*, 2021). Previous studies have shown that valuing the utility and relevance of course material contributes to students having increased subject matter interest (Hulleman *et al.*, 2010), a predictor of academic achievement (Schiefele *et al.*, 1992). Aikens and colleagues posit that redesigning calculus courses to emphasize the relevance of calculus to biological contexts is likely to increase the interest of life sciences students in math as well as enhance their performance in calculus courses. In support of this hypothesis, they find that 45% of post survey respondents

reported an affirming attitude toward math, and that understanding the relevance and utility of math to biology motivated these attitudinal changes.

The curricular initiative we describe here confronts the national problem with quantitative literacy and STEM retention and, based on the outcomes of this study, reflects a transformational approach to teaching college-level math that was developed at a large, public, research-intensive university. The new two-course series, called Mathematics for Life Scientists (courses Life Science 30A, or LS30A, and Life Science 30B, or LS30B), focuses on bridging the gap between the way math is *taught* and the way it is often *applied* in STEM fields (Marshall and Durán, 2018) and represents a paradigm shift in *how* mathematical concepts and skills are introduced to first-year college students in life sciences. Referred to hereafter as a contextualized math curriculum, these courses also provide a platform for teaching computer programming, a critical skill for the modern STEM workforce (NRC, 2003; Callier *et al.*, 2014; Rubinstein and Chor, 2014). Briefly, students are engaged in classic calculus topics, such as the derivative and the integral, but with a focus on the application of these concepts to dynamical systems, which are also used as a framework for teaching linear algebra and as a platform for students learning computer programming in Python so that they can numerically integrate nonlinear systems of differential equations. Perhaps most importantly, students learn how to think quantitatively about biological systems and practice constructing dynamical models for problems they have not previously encountered (i.e., engaging in knowledge transfer; Halpern, 1998).

Synopsis of College-Level Calculus Courses and Their Limitations

Most traditional first-year calculus courses for biology students devote many weeks of instruction to differentiation techniques, integration techniques, and in some cases, at the beginning of the course, computing limits. Some courses also cover various amounts of linear algebra, differential equations, and multivariable calculus, but only after an exhaustive treatment of single-variable differentiation and integration. Traditional “calculus for biology” courses typically use textbooks that teach some applications of calculus to biology (e.g., Comar, 2008; Rheinlander and Wallace, 2011), consistent with the approaches previously described for the development of biocalculus courses (Duffus and Olfier, 2010; Usher *et al.*, 2010; Eaton and Highlander, 2017; Aikens *et al.*, 2021). Applications of differentiation may include optimization, for example. However, this is primarily covered in the single-variable setting, with a focus on paper-and-pencil solutions, whereas numerical or other general techniques for optimizing functions of multiple variables would be much more applicable to modern biology. Applications of integration are generally limited to solving particularly simple (separable) single-variable differential equations and working with continuous random variables. Solutions to such simple differential equations have very limited use in the life sciences today, and treating the topic of random variables properly requires another whole course in probability theory. Furthermore, it is inevitably the case that these applications, the ones most relevant to modern biology, such as multivariable optimization techniques and numerically solving nonlinear differential equations, are covered very late in the course, if at all. The

end result is almost invariably a calculus course in which students spend the vast majority of their time learning math, with an occasional application added in, and the promise that if they spend months learning this material, they will eventually get to applications relevant to their life sciences major.

Overview of a Novel Contextualized Math Curriculum for Life Sciences Students

The contextualized math curriculum, Mathematics for Life Scientists (LS30A, LS30B), takes a fundamentally different approach to the quantitative preparation of biology students (Garfinkel *et al.*, 2017). The two courses that make up this curriculum focus on biological applications, and instructors develop the mathematics as needed to serve those applications. The topic of differential equations, and dynamical systems more generally, offers a vast wealth of applications to biology. Thus, the curriculum begins by introducing the concepts of positive and negative feedback in the language of differential equations, as well as techniques for using differential equations to model dynamical systems in ecology, epidemiology, physiology, chemistry, and other subjects. Simultaneously, students learn the fundamentals of computer programming, so that they can implement basic numerical solution techniques on a computer. After that introduction, the curriculum briefly detours into single-variable calculus to enable students to gain a conceptual understanding of the derivative. The students still learn differentiation rules, but primarily as a tool for analyzing the stability of equilibrium points of single-variable differential equations. To do the same for multivariable systems of differential equations, the curriculum transitions into linear algebra and basic multivariable calculus. While covering linear algebra, the crucially important concepts of eigenvalues and eigenvectors are introduced by considering the long-term behavior of matrix population models. As a result, students learn a number of useful mathematical tools, including selected topics from calculus and linear algebra, but always motivated by real scientific applications and always with a strong emphasis on the mathematical concepts rather than just paper-and-pencil calculations. Students gain the mathematical foundations needed to delve into several more advanced topics, including multivariable optimization techniques, limit cycles in dynamical systems, bifurcation theory, and chaotic behavior, all of which are covered in the two-course sequence.

A pilot of the contextualized math course, LS30A, was offered to 19 students in the Spring of 2013, and following that term, the two-course sequence was consistently offered throughout subsequent academic years. Support from a federal grant provided resources to expand the number of seats available in LS30 beginning in the Fall of 2014. As of Spring 2019, the majority of life sciences majors were completing the Mathematics for Life Scientists curriculum, with an average enrollment of about 1300 unique students per year, which corresponds to a little more than 70% of freshman undergraduates who enter the institution with a declared major in life sciences. Overall, LS30A has enrolled more than 4800 undergraduates since its pilot in Spring 2013. Approximately 4200 students have completed the second course in the sequence, LS30B, since it was first offered in Winter 2014. Here, we report the findings of our study of student outcomes in the contextualized LS30AB math curriculum. Our results demonstrate the efficacy of this innovative

instructional approach, and reinforce the long-standing national call to radically reform the undergraduate math curriculum for life sciences students (NRC, 2003, 2009; Edelstein-Keshet, 2005; Association of American Medical Colleges-Howard Hughes Medical Institute, 2009; AAAS, 2011).

CONTEXT AND MOTIVATION FOR RESEARCH STUDY

Because the contextualized math curriculum diverges significantly from traditional calculus in its approach to teaching mathematical concepts with applications to the biological sciences, its launch prompted a research study of student outcomes. The overarching goal of the study was to examine the longitudinal impacts on science course grades attributed to broad, course-level differences between the new contextualized curriculum (LS30A, LS30B) and a traditional “calculus for biology” series (Math 3A, 3B, 3C). We also documented student grades in the new math courses in comparison with their peers in traditional calculus, monitoring performance gaps between various student subgroups characterized by demographic characteristics such as sex, race/ethnicity, socioeconomic status (SES), and parent or legal guardian education status (i.e., first-generation status). Additionally, responses from prior surveys of life sciences majors suggested low levels of satisfaction with the traditional calculus sequence (B.V.V., 2013, unpublished data). Students expressed lower levels of confidence in their mathematical abilities after completing Math 3A, 3B, or 3C, with some students feeling discouraged from continuing their pursuit of a bachelor’s degree in the sciences. Thus, our study also incorporated findings from end-of-term student ratings of instruction (SRIs) in which we compared student interest in the subject matter at the beginning and end of each course. As noted previously, increased subject matter interest can be a positive predictor of academic achievement (Schiefele *et al.*, 1992).

Rationale for Measuring Student Performance in Subsequent Science Courses

In alignment with the overarching goal of the study, we focused on grade outcomes in courses that represented a shared, homogeneous curricular experience for life sciences students while also maximizing our study sample size. During the time frame of the study, life sciences students typically completed their math and chemistry course work during their first year of college and before beginning their introductory biology course work, which usually occurred in their sophomore year. Once the math, introductory biology, and chemistry curricula were completed, students could matriculate into their upper-division major course work, choosing from among 12 different life sciences majors spread across six departments and two interdisciplinary programs. Our sample size quickly diminished and the number of confounding factors increased as students’ curricular experiences diverged in upper-division course work; thus, we focused our course-level grade outcome analysis on introductory courses in biology and chemistry, as these represented a common curricular experience for all life sciences majors. In addition, most of these students also are required to complete a physics curriculum, the majority doing so in their junior and senior years of college. Because life sciences students often completed physics later in their course of study, we tracked grade outcomes in physics as a longitudinal marker of student academic performance in their STEM course work.

Substantiated as follows, we focused our grade analysis on the first course in the sequence comprising the curricula for biology, chemistry, and physics. In biology, we monitored student performance in an introductory cell biology and physiology course, Life Science 2, in which students were expected to apply the knowledge and skills learned in requisite math courses, exploring possible differences between those who took either the new contextualized math course LS30A or the traditional “calculus for biology” course Math 3A as their first math course. Life Science 2 covered various systems in human physiology. When studying the circulatory system, for instance, students calculated changes in flow and velocity through blood vessels, but to model the muscle system, students had to create diagrams demonstrating the relationship between power and velocity of a muscle from relaxation through contraction (D. Pires, personal communication). Thus, students progressed from calculating rates of change to building physiological models of dynamic, real-world systems, the latter of which were skills specifically emphasized in LS30A. During the time frame of the study, Life Science 2 was one of four introductory biology courses required of all life sciences STEM majors. It was the first in a three-course sequence; the fourth course could be taken in any order at any time in college.

Chemistry 14A was an enforced prerequisite for Life Science 2. As a result, Chemistry 14A effectively served as a gateway course for life sciences students, and its inclusion in our study was motivated, in part, by its critical position in the sequence of math and science courses comprising the overall introductory curriculum for life sciences majors. In addition, some faculty and administrators in the chemistry department were initially skeptical of an alternative instructional approach to teaching college-level math. There was no calculus pre- or co-requisite for Chemistry 14A when this study began, but this changed by year 4 of the study, because preliminary data from our evaluation indicated that students completing their math courses either concurrently with or before taking their introductory chemistry courses were performing better (e.g., achieving higher average chemistry grades) than those who waited to take their math courses until after starting chemistry. A recent review of math course syllabi, learning objectives, sample midterm exams, and final exams failed to reveal substantial conceptional or skill-based connections (J. Casey, personal communication), which suggests that improved performance outcomes in chemistry likely are more attributable to students developing more generalized cognitive skills that transfer well to science courses as opposed to enhanced learning of particular content.

There also was pushback from the physics department concerning the new math series, with skepticism as to whether the LS30A and LS30B courses would adequately prepare students for the physics curriculum. That said, a set of skills noted to be highly relevant to physics courses included modeling systems in real-world contexts, understanding how physical quantities get integrated into the system being modeled, and then visualizing and qualitatively analyzing the behaviors of these systems (J. Samani and S. Shaked, personal communication). These skills are emphasized in the new contextualized math courses but are not covered in traditional calculus for biology courses. Consequently, the overlap in concepts/skills between LS30A/B and the physics courses suggested to us that grades in the first phys-

ics course of the three-term series, Physics 6A, was an appropriate proxy for longer-term student performance outcomes.

To garner broad buy-in from skeptical science departments, gain approval from the college curricular committees for the new math courses, and reassure ourselves that we were helping rather than hindering life sciences students, we needed to demonstrate that student grade outcomes in chemistry and physics as well as biology would not be adversely affected by taking these new math courses in place of traditional calculus for biology. Consequently, our research study design was influenced heavily by our need to overcome this cultural barrier within the institution, so we focused on course-level performance outcomes in science classes (i.e., grade achieved) and not on analyses of actual learning in math courses (e.g., via concept inventories or common exams).

Research Study Predictions

Given the potential content connections between the new contextualized math courses and Life Sciences 2 and Physics 6A, we expected to observe grade improvements in these two science courses among students who took contextualized math relative to their peers who took traditional calculus for biology. Our study sample targeted students who had taken as their *first* math course at the institution either the contextualized math course (LS30A) offered by a department in life sciences or the corresponding calculus for biology course (Math 3A) offered by the math department. We did not necessarily predict any significant differences between these two student populations in Chemistry 14A grades, because obvious content connections to the math courses were not evident. However, if students were indeed acquiring more generalizable and transferrable cognitive or non-cognitive skills in the new math courses, it was possible that grade outcomes might also improve or at least be the same for both student populations. In gauging LS30A/B students' interest in the subject matter (i.e., learning math in the context of biology), we predicted a positive shift if LS30A/B was successful in establishing the relevance of math to biology in a curriculum designed for life sciences majors. Likewise, we hypothesized that this novel approach to learning math in a biology context might help to mitigate academic performance gaps historically observed in STEM gateway courses between students from disadvantaged backgrounds or social identity groups historically underrepresented in STEM and their counterparts with respect to background and social identity groups.

METHODS

Overview of Math Course Characteristics and Instructional Approaches

To better establish the classroom setting for the study, we provide information about the common and differentiating structures and instructional features of LS30A and Math 3A, the first two math courses of their respective series. Table 1 summarizes course characteristics in four categories: personnel, primary sections led by the instructor, instructional practices in primary sections, and secondary sections led by graduate student teaching assistants (TAs). Two authors on this paper (A.G. and W.J.C.) were instructors for LS30A, one of whom (W.J.C.) is full-time teaching faculty in the Department of Mathematics and also taught Math 3A. Together, they have firsthand knowledge of the math course information provided in the table and detailed as follows.

TABLE 1. Comparison of LS30A and Math 3A course characteristics

Course component	LS30A	Math 3A
Personnel		
Instructor departmental affiliation	Mix of life sciences (majority of instructors) and math	Math only
Graduate student teaching assistant (TA) appointment	50% time (20 hours/week)	25% time (10 hours/week)
Undergraduate student learning assistants (LAs)	Yes	No
Graduate student Readers	Yes	Yes
Primary sections		
Total enrolled during 4-year study	1879	3608
Average enrolled per year	470	902
Approx. enrolled per section per year	100–300	200
Instructional practices in primary sections		
Instructor coordination of pacing, curricular content, and classroom culture	Yes	No
General pedagogy	Primarily lecture with “chalk talk”	Primarily lecture with “chalk talk”
Grading practice	Criterion-based grading	Norm-referenced grading
Secondary sections		
Computational lab	Yes	No
Discussion section	Yes ^a	Yes
Average size	18.8	30.6
Number of secondary sections per TA	2	2

^aApproximately half of the 2-hour lab was dedicated to TA-led discussion in which students could ask questions about lecture and homework.

Personnel. During the study period, LS30A was taught by a team of instructors, mostly from the life sciences, and consisted of one tenured faculty member who conducts research in quantitative biology and several non-tenure track instructors with PhDs in either math or a quantitative subdiscipline of biology, most of whom were hired on a long-term or permanent basis, with the teaching of LS30A specified in their job descriptions. This structure helps with maintaining continuity of classroom culture, one that explicitly emphasizes a growth mindset (Dweck, 1999; Canning *et al.*, 2019), with instructors consistently messaging that everyone is capable of success (Cohen *et al.*, 1999). Math 3A has been taught by the occasional full-time instructor or ladder faculty member, but these courses are primarily taught by postdocs, who are temporary instructors hired on a 2- to 3-year contract, and all of the Math 3A instructors were from the math department. Besides the departmental affiliation of the instructors, perhaps the biggest instructor-level difference between LS30A and Math 3A is that there was always a strong degree of communication and coordination among LS30A instructors; this was rarely the case for Math 3A at the time of the study. As a result, the pace, and to some extent even the topics covered, varied much more for Math 3A than for LS30A.

Both LS30A and Math 3A employ TAs. One notable difference was that LS30A TAs have 50% time appointments (20 hours/week) and Math 3A TAs have 25% time appointments (10 hours/week). As a result, LS30A TAs attend lectures and have more time for student interactions such as drop-in office hours compared with Math 3A TAs, who are not required to attend lectures. All LS30A and Math 3A TAs new to teaching are required to complete a departmental pedagogy training course. By the second year of offering the new contextualized math curriculum, the LS30A instructional team also included learning assistants (LAs; see Talbot *et al.*, 2015), undergraduates who previously completed LS30A and who are trained to help TAs

facilitate collaborative learning during computational lab sessions. Finally, Math 3A utilizes graduate student Readers, who are hired on an hourly basis to grade weekly homework assignments. Starting in 2015, LS30A also began hiring Readers for the same purpose, leaving the TAs to focus on grading exams and giving feedback on weekly computational lab assignments.

Primary Sections. During the first 3 years of the study, enrollments in LS30A were lower than in Math 3A; however, enrollments increased over time, with LS30A enrollments exceeding those of Math 3A during the fourth and final year of the study. This trend has remained consistent through the current academic year with fewer than 300 students now enrolling in Math 3A annually. On average, during most of the study period, the section sizes were in the range of 100 to 300 students for LS30A. In year 1, individual sections of LS30A enrolled fewer than 100 students. By year 2, LS30A enrollments ranged from 100 to 240 students per section. In year 3, at least one section of LS30A enrolled up to 350 students. Individual sections of Math 3A, on the other hand, have always been capped at 210 students. During the study period, they were usually filled close to capacity.

Instructional Practices in Primary Sections. During the time frame of the study, both LS30A and Math 3A instructors relied heavily on lecture and real-time writing (“chalk talk”) as the dominant form of instruction. Starting in Fall 2016, in an attempt to make the courses more interactive, clicker questions were introduced into all sections of LS30A; clickers were not used in Math 3A. The grading strategies for LS30A and Math 3A were also distinct during the study period, with implications from the research that these strategies can differentially impact classroom climate (Hughes *et al.*, 2014; Schinske and Tanner, 2014). LS30A grades have always been assigned across all

sections of the course using criterion-based assessment practices, which help to cultivate a collaborative, rather than competitive, learning environment. For Math 3A, the grading structure varied by instructor, but most instructors during the time frame of the study would have used norm-referenced grading (e.g., grading on a curve), practices that promote a competitive learning environment that can adversely affect minoritized students (Covington, 1992). Plus/minus grades were given in both math courses.

Secondary Sections. Both math courses have secondary sections, but the format is vastly different. LS30A has a 2-hour weekly computational lab with graduate student TAs and undergraduate LAs. LS30A TAs are encouraged to use about half of the lab section time as a discussion section, allowing students to ask questions and clarify misunderstandings from lecture. The lab assignments themselves are largely self-guided, with students encouraged to collaborate in practicing computer-based coding applications in Python, and the TAs and LAs serve as facilitators of these collaborative discussions. Math 3A has a 1-hour weekly discussion section with graduate student TAs, who primarily lead problem-solving sessions with students.

The secondary section sizes for LS30A and Math 3A were always different. Enrollment in LS30A secondary sections is constrained by the capacity of the computer labs (max. 18–24 students), with an average enrollment over the study period of 18.8 students per section. Each LS30A TA is assigned two sections, and thus is responsible for oversight of approximately 40 students per term. Math 3A, on the other hand, maintained an average enrollment of 30.6 students per secondary section during the time frame of the study. Each TA was assigned two sections, and thus was responsible for oversight of around 60 students per term. Thus, the student to TA ratio in LS30A was lower (40:1) than in Math 3A (60:1).

Data and Sample

Several data sources are used in this study, including information provided by the university registrar and self-report data collected from end-of-term SRIs. All data analyses were conducted with human subjects' ethics board approval (IRB no. 13-001490).

To understand the grade outcomes for students in their science courses as a function of which math course they completed, the study sample included students who had taken as their *first* math course at the institution either the contextualized math course (LS30A) or the corresponding traditional calculus course for biology (Math 3A). The registrar provided students' demographic and admissions information (e.g., standardized test scores, high school grade point average [GPA]) and transcript data with course histories, including grades, for all students who completed Math 3A and/or LS30A from Fall 2013 through Winter 2017. Final grades for the two math courses were obtained for examination of trends in performance gaps, defined by differences in the mean and/or median grade in LS30A or Math 3A, for several groups based on demographic characteristics. Letter grades also were acquired for three lower-division science courses required of most life sciences majors: Chemistry 14A, which is the first chemistry course in a sequential four-course series, typically completed at

the time of the study during the first year; Life Sciences 2, which was commonly taken as the first of four biology courses completed before entry into upper-division major courses and for which Chemistry 14A was an enforced requisite; and Physics 6A, which was the first physics course in a sequential three-course series and often taken during the junior or senior year. Science grades were limited to students' most recent grades for their respective courses through Spring 2017, because that was the final term during which Physics 6A was offered at the institution. The Physics 6 curriculum was replaced by a new physics series for life sciences majors in Fall 2017. Notably, 2017–2018 was the last year that Life Sciences 2 was offered; the introductory biology curriculum has since been replaced by a sequence of three introductory biology courses with no chemistry requisites and specifically designed for first-year students.

In addition to limiting our sample to students' science course grades from Fall 2013 through Spring 2017, the sample was further narrowed to direct-admit students, because transfer students are not required to take math after matriculating to the university, as they would have completed their math requirements at their originating institutions before transfer. Consequently, those transfer students who elected to enroll in LS30A or Math 3A would have previously completed their math requisites for college, and thus neither course would qualify as their *first* math course taken. Of direct-admit students, only those students who completed either Math 3A or LS30A as their *first* math course at the institution and who took their respective courses only once were included in the sample to minimize multiple math treatment effects. Thus, anyone who repeated either LS30A ($N = 21$) or Math 3A ($N = 61$) for a letter grade was excluded from the sample. The few students who took both Math 3A and LS30A ($N = 23$) were also excluded from the sample. We further narrowed the math sample to include only those 1447 students who also completed Chemistry 14A, Life Sciences 2, and Physics 6A, our key outcomes of interest.

For the purposes of our research study, we focused on the most common science course-taking patterns as our sampling strategy, which ensured that we had a large enough sample population for meaningful grade comparisons and minimized confounding effects of the conditions or constraints that would have led to students scheduling their science courses according to a less common sequencing pattern. Such differences among students could not be explained, because we had no control over student choice and did not evaluate student reasoning for any of the course-taking patterns observed. Thus, the final sample included only those students who enrolled in Chemistry 14A during the same term as their first math course (63.9% of the math and science course sample, as compared with 18.5% who took Chemistry 14A *before* Math 3A or LS30A and 17.6% who took Chemistry 14A *after* Math 3A or LS30A), enrolled in Life Sciences 2 subsequent to Chemistry 14A (96.2%), and enrolled in Physics 6A subsequent to Life Sciences 2 (98.9% of the math and science course sample). These course enrollment and sequencing criteria yielded a final analytic sample of 909 students.

As part of the institution's course and instructor evaluation process, anonymous end-of-term surveys, SRIs, are administered to students to document their experiences in a course as well as their impressions of instructors. For this study, with instructor permission, SRI data were obtained for LS30A and

LS30B from Fall 2013 through Winter 2017 terms. The average response rate for all LS30A offerings was 79.2% ($N = 1120$) and for LS30B the average response rate was 73.9% ($N = 908$), both well within established thresholds for statistically valid survey response rates (Nulty, 2008).

Data Analysis

Logistic Regression and Propensity Score Analysis. Propensity score weights were employed to address potential selection bias or course assignment differences among students who enrolled in the contextualized math curriculum as opposed to traditional calculus for biology (Guo and Fraser, 2010). In general, use of propensity score estimates is a quasi-experimental technique used to adjust for confounding factors (covariates) in a study sample by constructing a control group composed of individuals with characteristics similar to those of individuals in the treatment group (Austin, 2011). This method provides a counterfactual framework (Rosenbaum and Rubin, 1983, 1984, 1985), which recognizes that, although we cannot go back in time and randomly assign students to the two different conditions, we can use logistic regression to predict a student's likelihood of being in the treatment condition (in this case, LS30A) compared with the control condition (in this case, Math 3A) and then estimate the treatment effect on any one individual in the sample.

A difference between propensity score weighting and propensity score matching is the ability of the former to include all sample individuals in the analyses, rather than only those who were successfully matched. This more-inclusive approach minimizes bias in the analysis due to incomplete matching between treatment and control groups and strengthens the generalizability of the treatment effect (Austin, 2011). Propensity score weighting generates output (ω) based on the inverse probability of treatment given a set of observed covariates [$e^{\hat{\chi}}$] (Olmos and Govindasamy, 2015). For individuals in the treatment group (e.g., LS30A), we use this equation: $\omega = 1 / e^{\hat{\chi}}$, and for individuals in the control group (e.g., Math 3A), we use this equation: $\omega = 1 / [1 - e^{\hat{\chi}}]$. As a result of transforming the variables into weight estimates via these two equations, the sample size in the weighted pseudo data set (calculated sample size) will be inflated compared with the unweighted data set (actual sample size).

Table 2 presents descriptive statistics for the sample before and after applying the weights that were derived using propensity score analysis. Weighting yields a more balanced sample with respect to representation of sex, race/ethnicity, Pell Grant recipient status (used in the study as a proxy for SES), and parent/guardian education status (the criterion used to establish first-generation status), as demonstrated by the relative lack of statistical significance (within our threshold of $\alpha < 0.01$) in group differences (based on two-proportion z -tests and independent-samples t tests). Moreover, the differences in means for scale variables, including high school GPA, Scholastic Aptitude Test [SAT] math scores, and Advanced Placement [AP] Biology and Calculus exam scores, are smaller after weighting, albeit group differences are still statistically significant ($p < 0.01$).

It is worth noting that the estimation of propensity scores is only as good as the variables in the model used to estimate them. Our regression models are limited to institutional data

and thus lack explicit measures of motivation that students may have had when enrolling in their courses. Despite this limitation, our estimations include a robust set of demographic and academic preparation measures. Supplemental Table S1 includes the covariates and their parameter estimates associated with the logistic regression model used to estimate propensity scores for sample individuals' probability to have enrolled in LS30A versus Math 3A.

We thought it was important to account for predictors in our model that we know from existing research to be factors affecting college math readiness (e.g., high school preparation). Others, such as race/ethnicity, were included, because they are supported by research as predictive of educational outcomes. In addition, we opted to avoid purely statistically motivated variable selection in our model (Heinze and Dunkler, 2017), which enabled us to account for possible shared predictive value among significant and insignificant variables, thereby preventing us from falsely attributing an effect to only the significant predictors.

Aligned with our goal to examine broad, course-level differences, our model did not include primary section within each course as a random effect for a nested regression analysis and instead assumes students' share experiences by course (LS30A vs. Math 3A). Random effects or a nested model presumes there is a substantive reason to group students together (e.g., by primary section) that makes them different from one another (Hox, 2010). We concluded that we simply do not have data to treat primary sections of the same course as amply distinct from one another. In our study, we were limited to course-level information about the structural details and instructional features of each math course (see Table 1).

Math and Science Course Grade Analysis. We compared course grades between LS30A and Math 3A students and, as a means to explore possible grade disparities, among subgroups of students within LS30A and Math 3A. Parametric t tests were used to compare mean grades, with the standardized magnitude of the difference, or effect size, calculated as Cohen's d coefficients (Lenhard and Lenhard, 2016). A value of 0.20 is considered a small effect, 0.50 is regarded as a medium effect, and 0.80 is a large effect (Cohen, 1988, 1992). Both sample size (Lumley *et al.*, 2002) and visual inspection of histogram plots (Skovlund and Fenstad, 2001; Fagerland, 2012) suggest the appropriateness of parametric t tests for comparing group outcomes. Nevertheless, applying the Kolmogorov-Smirnova and Shapiro-Wilk tests indicates that the samples do not meet the assumption of normality. Thus, nonparametric Mann-Whitney U -tests (Corder and Foreman, 2009) were used additionally to test for significant differences in the distributions of math and science course grades. Effect size calculations are not common with nonnormal data (see Tomczak and Tomczak, 2014) but become necessary when using nonparametric techniques (Leech and Onwuegbuzie, 2002). Pearson correlation coefficients (r) were calculated as an estimate of the effect size, or the strength of linear association between two variables (e.g., grades in the two math courses), using output from nonparametric tests and the equation: $r = Z / \sqrt{n}$, where Z is the Z -score and n is the number of observations on which Z is based (Tomczak and Tomczak, 2014). A Pearson correlation coefficient (r) of -1 corresponds to a perfect negative correlation and

TABLE 2. Descriptive statistics for logistic regression model before and after applying propensity score weighting

Sample characteristics	Unweighted						Weighted						
	Math 3A (N = 615)			LS 30A (N = 294)			Math 3A (N = 900)			LS 30A (N = 990)			
	%	Mean	SD	%	Mean	SD	%	Mean	SD	%	Mean	SD	Significant difference
Sex: female	66.8			58.8			64.6			66.5			0.383
Race: AAPI ^a	46.7			55.8			49.4		*	47.9			0.512
Race: Black	2.9			3.4			3.5			4.5			0.266
Race: international ^b	3.4			3.4			3.4			3.1			0.703
Race: Hispanic	20.7			11.9			17.9		**	19.4			0.402
Race: white (reference category)	22.9			22.8			22.7			21.3			0.470
Race: other (including unknown)	3.4			2.7			3.1			3.8			0.453
Pell Grant recipient	40.7			29.6			37.7		**	42.6			0.032
First generation (4-year graduate)	31.1			23.1			28.7		*	33.4			0.029
AP Biology exam: scored 3 or higher	51.9			65.0			54.9		***	48.5			0.005
AP Calculus exam: scored 3 or higher	42.1			25.5			37.3		***	35.5			0.424
PEERS participant ^c	4.6			6.1			5.6			5.6			1.000
Life sciences major	87.3			85.0			86.3			83.9			0.142
High school GPA	4.39	0.22		4.47	0.24		4.41	0.22	***	4.35	0.32		0.000
SAT Math score	678.70	62.52		712.07	64.58		687.58	62.66	***	671.90	83.31		0.000
AP Biology exam ^d	3.45	0.99		3.94	0.86		3.52	0.97	***	3.69	0.95		0.003
AP Calculus exam ^d	3.04	1.12		3.60	1.39		2.97	1.16	***	3.55	1.32		0.000

^aAAPI refers to students identifying as Asian American and Pacific Islander.

^bThe university codes international students' race or ethnicity as "foreign." Here we will simply refer to this race/ethnicity group as international students.

^cThe Program for Excellence in Education and Research in the Sciences (PEERS) is a cohort-based undergraduate STEM student retention initiative that recruits students from race/ethnicity groups underrepresented in STEM, low-income students, and students who enter the institution with challenging life circumstances.

^dBinary variable (passed AP exam vs. all else) used for propensity score estimation.

**p* < 0.05.

***p* < 0.01.

****p* < 0.001.

an r of +1 to a perfect positive correlation between variables. In this case, the effect size is small if the value of r is between 0.1 and 0.3 and large if r is above 0.5 (Cohen, 1988, 1992). Group means, medians, and results from their corresponding parametric and nonparametric tests for significant differences are provided in the tables, wherein we report p values ranging from $p < 0.05$ to $p < 0.001$. Our threshold for statistical significance, however, is $\alpha = 0.01$. This more stringent threshold helped us focus on those group differences that were more likely to be of practical importance, unless applying a less stringent threshold ($p < 0.05$) appeared to correspond to changes in the underlying quantities that were of potential practical importance as assessed by effect size.

To examine possible math course effects on different subgroups of students in their science courses, we ran parallel sets of blocked ordinary least-squares (OLS) regressions for all three science grade outcomes. Independent variables were organized into blocks comprising precollege characteristics (e.g., demographics, high school academic performance), math course grade, and whether students took LS30A or Math 3A as their first college math course. Plots of residuals and collinearity statistics confirm that the data do not violate regression assumptions. Analysis of variance (ANOVA) was conducted to confirm statistical significance of the models ($\alpha = 0.01$).

Self-Report Data Analysis. Two SRI questions of interest for this study asked LS30A and LS30B students to gauge their “subject interest before course” and “subject interest after course” using four categorical response choices: N/A, low, medium, and high. We omitted N/A responses and assigned numeric values on a 3-point scale to the remaining three categorical data as follows: 1 (low), 2 (medium), and 3 (high). Descriptive analysis of students’ self-report data to these two survey items produced histograms of response frequencies in three categories (low level of interest, medium level of interest, and high level of interest) at two time points (before and after the course). We conducted z -tests to compare the distributions of response frequencies for LS30A and LS30B. Unfortunately, we were not able to access comparable data for Math 3A. A copy of the entire SRI instrument is provided (see Supplemental Figure S1).

RESULTS

Analysis of science course (chemistry, life science, physics) grades among students who completed either contextualized math (LS30A) or traditional calculus for biology (Math 3A) as their first math course provides evidence supporting the efficacy of the new math curriculum in improving learning. In addition to these cognitive gains, examination of students’ responses to survey items on end-of-term SRIs reveal marked gains in student interest in the subject matter, a noncognitive measure that could contribute to student motivation to persist in their course work as STEM majors (Graham *et al.*, 2013).

Comparing Grades in Science Courses by First Math Course Completed

To determine the possible effects of the math courses on academic performance in subsequent science courses, we first examined group differences between Math 3A and LS30A with respect to course grades in Chemistry 14A, Life Science 2, and Physics 6A. Table 3 includes descriptive statistics and signifi-

cance test results for 907 students who earned a passing grade (at least a “C–”) in their respective math classes, and thus satisfied the math co- or prerequisite for the three science courses examined in this study. Nonparametric tests for significant differences, for both unweighted and weighted samples, confirm that students who enroll in and pass LS30A earn higher grades in their subsequent science classes than do their counterparts in Math 3A. Recall, the apparent change in sample size between the unweighted data set (actual sample size) and weighted pseudo data set (calculated sample size) results from the transformation of variables used to calculate propensity scores in the logistic regression model. By more conservative (i.e., weighted) estimates, LS30A students earn grades ranging from 0.10 grade points higher in Chemistry 14A to 0.17 grade points higher in Life Science 2, on average, than Math 3A students. In other words, these results indicate that students who enrolled in LS30A earned significantly higher grades in chemistry ($p < 0.001$), life sciences ($p < 0.001$), and physics ($p < 0.01$) courses than their peers who had taken Math 3A as their first math course at the university.

To better understand the magnitude of mean grade differences between LS30A and Math 3A students and their practical significance, we estimated Cohen’s d coefficients based on parametric t test data. By more conservative (i.e., weighted) estimates, the effect size is considered negligible (<0.2) to rather small (≥ 0.2 but <0.5), ranging from 0.15 for Chemistry 14A to 0.23 for Life Science 2. The effect size is much larger for the unweighted samples, ranging from 0.54 in Physics 6A to 0.7 in Chemistry 14A (medium effect ≥ 0.5 but <0.8). That said, we acknowledge that these effect size estimates may be affected by departures from normality in the data set, which is what motivated us to employ nonparametric tests for significance in the first place.

The Pearson correlation coefficients (r) were calculated as a means to gauge effect size based on the nonparametric data. These results mirror those estimated using Cohen’s d coefficients. For the weighted samples, the effect size is considered negligible (<0.1) for Chemistry 14A and small (0.1–0.3) for Life Science 2 (0.13) and Physics 6A (0.11). These results are interesting, because they support our hypotheses, in which we predicted some grade improvements for LS30A students relative to their Math 3A peers in Life Science 2 and Physics 6A based on the potential relevance of content between the courses. Furthermore, we did not expect large differences between these two student populations in their Chemistry 14A grades, as we were unable to identify obvious content connections to either math course. Given that there was a significant difference in both mean science course grades and science course grade distributions, respectively, between LS30A and Math 3A student groups, it appears that LS30A students may benefit in their learning from acquisition of more generalizable and transferrable cognitive or noncognitive skills.

In addition to group differences by first math class taken, and in response to pervasive disparities among students in STEM academic performance and retention (PCAST, 2012), we examined possible math course effects alongside those of students’ sex, race/ethnicity, SES, first-generation status, and academic preparation. Table 4 shows all three sets of OLS regressions (both unweighted and weighted versions for three science course grade outcomes); omnibus tests confirm that the models

TABLE 3. Descriptive statistics of mean and median grades in chemistry, life science, and physics courses for the sample population^a disaggregated by math class taken and science course completed between 2013 Fall and 2017 Spring

	Math 3A			LS30A			Parametric ^b			Nonparametric ^c		
	N	Mean	SD	Median	SD	Mean	Median	Cohen’s <i>d</i>	Significant difference	<i>r</i>	Significant difference	
Unweighted sample												
Chemistry 14A	615	2.75	0.62	2.70	0.68	3.20	3.30	0.70	0.000	0.32	0.000	***
Life Sciences 2	614	2.89	0.73	3.00	0.69	3.32	3.30	0.60	0.000	0.29	0.000	***
Physics 6A	614	3.01	0.76	3.00	0.66	3.40	3.70	0.54	0.000	0.27	0.000	***
Weighted sample												
Math 3A												
	N	Mean	SD	Median	SD	Mean	Median	Cohen’s <i>d</i>	Significant difference	<i>r</i>	Significant difference	
Chemistry 14A	900	2.79	0.62	2.70	0.77	2.89	3.00	0.15	0.001	0.08	0.000	***
Life Sciences 2	898	2.92	0.73	3.00	0.76	3.09	3.30	0.23	0.000	0.13	0.000	***
Physics 6A	899	3.05	0.75	3.00	0.73	3.20	3.30	0.21	0.000	0.11	0.002	**
LS30A												
	N	Mean	SD	Median	SD	Mean	Median	Cohen’s <i>d</i>	Significant difference	<i>r</i>	Significant difference	

^aChemistry and physics grades for students with at least a “C-” (GPA 1.7) in math course. Sample excludes two cases from LS30A.
^bFor parametric statistics, *t* tests were used to compare mean grades, and Cohen’s *d* coefficients were calculated as measure of effect size (small effect 0.2, medium effect 0.5, and large effect 0.8).
^cFor nonparametric statistics, the Mann-Whitney *U*-test was used to compare grade distributions, and Pearson correlation coefficients (*r*) were calculated as a measure of effect size (small effect 0.1–0.3, medium effect 0.3–0.5, large effect > 0.5).
 **p* < 0.05.
 ***p* < 0.01.
 ****p* < 0.001.

are statistically significant ($p < 0.001$ for all ANOVAs), with adjusted R^2 values for full, final models ranging from 0.274 (predicting Physics 6A grades) to 0.498 (predicting Chemistry 14A grades). The coefficients for all six full models are provided in Table 4.

Using the standardized coefficient, beta, to compare effect sizes within models for weighted samples, we observe that the largest effect sizes are associated with how well students performed on their math SAT test and in their math classes. That said, the models also confirm that *which* math class students completed also matters in their science grade outcomes. The addition of math course (LS30A vs. Math 3A) contributes predictive value to each model, as evidenced by the statistical significance of the positive math course coefficient as well as the significance of change in R^2 value with the addition of the math course variable (ΔR^2 Chemistry 14A_{weighted} = 0.019, $p < 0.001$; ΔR^2 Life Sciences 2_{weighted} = 0.025, $p < 0.001$; ΔR^2 Physics 6A_{weighted} = 0.021, $p < 0.001$). Moreover, these findings are consistent regardless of outcome measure (science course grades) or weight (the more conservative, weighted estimates parallel the unweighted results). Thus, students who take LS30A as their first college math course tend to have significantly higher average chemistry, life science, and physics grades compared with those who take Math 3A, even after controlling for demographic characteristics, high school academic preparation, and math grade. This regression analysis lends further support to our hypotheses and suggests that LS30A students are benefiting in their learning in all three science courses, whether due to apparent content-based connections or to a more generalizable and positive impact on cognitive or noncognitive skill development.

Student Performance and Equity of Learning in Contextualized Math

For both traditional calculus and contextualized math courses, we calculated the mean and median course grades and examined the distribution of final letter grades ($N = 909$; see Table 5 and corresponding histograms in Figure 1) as a performance indicator that could possibly be attributed to differences in student learning or to other cognitive and/or noncognitive benefits associated with differences in course structure and instructional approach (see Table 1). Notably, in our previous analyses of science course grades, we used weighted samples in predicting grade outcomes in order to adjust for potential bias in covariates and produce findings that are more generalizable to the larger undergraduate life sciences population. For this part of our study, we used unweighted math grades to compare student performance within our actual (i.e., unweighted) sample.

Math 3A students in our sample, on average, earned 3.25 (SD = 0.69) grade points for their final course grade, or slightly less than a “B+” (3.3 on a 4.0 grade point scale). LS30A students, on the other hand, earned an average grade of 3.41 (SD = 0.72), or slightly higher than a “B+” average. Parametric and nonparametric tests confirm that the difference in mean grades and in the overall grade distribution, respectively, between the two classes is statistically significant (at $p < 0.001$). Both Cohen’s *d* and Pearson’s correlation coefficient (*r*) indicate the effect size is practically small yet still significant. On its own, this finding provides some support for the efficacy of LS30A as having a positive impact on student performance but does not

TABLE 4. Ordinary-least squares (OLS) regression models predicting chemistry, life science, and physics grades

	Unweighted (N = 908)				Weighted (N = 1886)				
	B	SE	Beta	Significance ^d	B	SE	Beta	Significance ^d	
	(Constant)	-1.085	0.401		0.007	**	-1.621	0.231	0.000
Sex: female	-0.071	0.035	-0.050	0.044	*	-0.043	0.025	0.083	
Pell Grant recipient	0.016	0.039	0.012	0.677		-0.018	0.028	0.516	
First-generation, 4-year graduate	-0.031	0.043	-0.021	0.466		-0.050	0.031	0.106	
Race: other (including NA/AN ^a and unknown)	0.086	0.099	0.022	0.385		-0.095	0.068	0.160	
Race: Black	-0.173	0.102	-0.044	0.088		-0.316	0.066	0.000	***
Race: international ^b	0.043	0.103	0.012	0.674		-0.004	0.073	0.958	
Race: Hispanic	-0.088	0.057	-0.050	0.126		-0.199	0.041	0.000	***
Race: AAPI ^c	-0.014	0.043	-0.010	0.746		-0.097	0.031	0.002	**
High school GPA	0.278	0.078	0.093	0.000	***	0.326	0.047	0.000	***
SAT Math score or converted ACT score if higher than or missing SAT	0.002	0.000	0.174	0.000	***	0.003	0.000	0.000	***
AP Biology exam: scored 3 or higher	0.091	0.036	0.067	0.011	*	0.100	0.026	0.000	***
AP Calculus exam: scored 3 or higher	-0.089	0.035	-0.064	0.012	*	-0.096	0.025	0.000	***
Math grade	0.448	0.026	0.466	0.000	***	0.393	0.018	0.000	***
LS30A (vs. Math 3A)	0.244	0.037	0.169	0.000	***	0.197	0.023	0.000	***
Adjusted R ²	0.457					0.498			
R ² change (add math course)	0.026			0.000	***	0.019		0.000	***
Life Sciences 2									
	Unweighted (N = 908)				Weighted (N = 1887)				
	B	SE	Beta	Significance ^d	B	SE	Beta	Significance ^d	
	(Constant)	0.084	0.489		0.864		-0.321	0.284	0.259
Sex: female	-0.052	0.043	-0.033	0.227		-0.039	0.031	0.209	
Pell Grant recipient	-0.073	0.048	-0.047	0.131		-0.086	0.035	0.014	*
First-generation, 4-year graduate	-0.082	0.053	-0.050	0.118		-0.111	0.038	0.003	**
Race: other (including NA/AN ^a and unknown)	0.137	0.121	0.032	0.260		0.020	0.084	0.815	
Race: Black	0.040	0.124	0.009	0.749		-0.382	0.081	0.000	***
Race: international ^b	0.022	0.125	0.005	0.862		0.027	0.090	0.763	
Race: Hispanic	-0.048	0.070	-0.025	0.495		-0.035	0.050	0.485	
Race: AAPI ^c	-0.062	0.052	-0.042	0.236		-0.094	0.038	0.013	*
High school GPA	0.238	0.095	0.073	0.012	*	0.322	0.058	0.000	***
SAT Math score or converted ACT score if higher than or missing SAT	0.000	0.000	0.038	0.260		0.001	0.000	0.001	**
AP Biology exam: scored 3 or higher	0.145	0.043	0.096	0.001	***	0.096	0.032	0.002	**
AP Calculus exam: scored 3 or higher	-0.082	0.043	-0.053	0.058		-0.056	0.030	0.064	
Math grade	0.476	0.032	0.450	0.000	***	0.419	0.022	0.000	***
LS30A (vs. Math 3A)	0.259	0.046	0.162	0.000	***	0.238	0.029	0.000	***
Adjusted R ²	0.338					0.322			
R ² change (add math course)	0.024			0.000	***	0.025		0.000	***

(Continues)

TABLE 4. Continued

	Unweighted (N = 907)			Weighted (N = 1885)					
	B	SE	Beta	Significance ^d	B	SE	Significance ^d	Beta	Significance ^d
Physics 6A									
(Constant)	-0.258	0.514		0.616	0.103	0.292		0.724	
Sex: female	-0.133	0.045	-0.084	0.003	-0.132	0.032	**	-0.084	0.000
Pell Grant recipient	-0.111	0.050	-0.071	0.028	-0.118	0.036	*	-0.077	0.001
First-generation, 4-year graduate	-0.001	0.055	0.000	0.989	0.021	0.039		0.013	0.598
Race: other (including NA/AN ^a and unknown)	0.191	0.127	0.044	0.133	0.253	0.086		0.062	0.003
Race: Black	-0.205	0.130	-0.047	0.116	-0.291	0.083		-0.076	0.000
Race: international ^b	0.007	0.132	0.002	0.958	0.096	0.093		0.023	0.298
Race: Hispanic	-0.240	0.074	-0.122	0.001	-0.247	0.052	**	-0.129	0.000
Race: AAPI ^c	-0.112	0.055	-0.074	0.042	-0.117	0.039	*	-0.079	0.003
High school GPA	0.296	0.100	0.089	0.003	0.242	0.060	**	0.090	0.000
SAT Math score or converted ACT score if higher than or missing SAT	0.002	0.000	0.133	0.000	0.002	0.000	***	0.163	0.000
AP Biology exam: scored 3 or higher	0.021	0.046	0.014	0.651	0.059	0.032		0.040	0.068
AP Calculus exam: scored 3 or higher	-0.012	0.045	-0.008	0.788	0.011	0.031		0.007	0.736
Math grade	0.353	0.033	0.329	0.000	0.288	0.023	***	0.284	0.000
LS30A (vs. Math 3A)	0.222	0.048	0.137	0.000	0.220	0.030	***	0.147	0.000
Adjusted R ²	0.292				0.274				
R ² change (add math course)	0.017			0.000	0.021		***	0.000	0.000

^aNA/AN subgroup includes students who identify as Native American or Alaskan Native.

^bThe university codes international students' race or ethnicity as "foreign." Here we will simply refer to this race/ethnicity group as international students.

^cAAPI refers to students identifying as Asian American and Pacific Islander.

^dA dash would have denoted true zero values in the table; however, there are no true zero values. More decimal places would be needed to show values beyond the three decimal places shown. Omnibus tests confirm models are statistically significant ($p < 0.001$ for all ANOVAs).

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

TABLE 5. Descriptive statistics of unweighted mean and median grades in math courses for the sample population ($N = 909$) disaggregated by the first math class completed between 2013 Fall and 2017 Spring

Course	N	Parametric ^a				Nonparametric ^b		
		Mean	SD	Cohen's d	Significant difference	Median	r	Significant difference
Math 3A	615	3.25	0.69	0.24	***	3.30	0.14	***
LS30A	294	3.41	0.72			3.85		

^aFor parametric statistics, t tests were used to compare mean grades, and Cohen's d coefficients were calculated as measure of effect size (small effect 0.2, medium effect 0.5, and large effect 0.8).

^bFor nonparametric statistics, the Mann-Whitney U -test was used to compare grade distributions, and Pearson correlation coefficients (r) were calculated as a measure of effect size (small effect 0.1–0.3, medium effect 0.3–0.5, large effect >0.5).

*** $p < 0.001$.

rule out confounding factors such as differences in instructors' grading practices (see Table 1). When combined with the previous findings for science grades, which we would argue are far better indicators of learning, the difference in average grades in the math courses provides more compelling evidence of the benefits conferred to students who complete the contextualized math curriculum.

In addition to teaching math in a biology context and better preparing students for science courses, LS30A seeks to offer the benefit of a more equitable and inclusive learning environment. As previously mentioned (see Table 1), LS30A instructors meet regularly to not only coordinate curricular content and pacing of the course, but also to ensure consistent messaging to students about growth mindset by both instructors and TAs (Cohen *et al.*, 1999; Dweck, 1999; Canning *et al.*, 2019). We hypothesized that this approach to learning college-level math might reduce performance gaps between students historically underserved and underrepresented in life sciences fields.

Table 6 shows unweighted mean and median grades for student subgroups in Math 3A and LS30A. In line with persistent disparities throughout STEM education, there are apparent differences in average grades among students in both math courses with respect to students' sex, race/ethnicity, SES, and first-generation status. For the latter three student characteristics, nonparametric tests confirm that the difference in the overall Math 3A grade distributions between the two subgroups is statistically significant ($p < 0.001$) and corresponds to medium effect sizes as measured by Pearson correlation coefficients (r), with a range of -0.19 to -0.26 . In other words, students who identified as a member of a racial/ethnic group underrepresented in

STEM, as lower SES, or as a first-generation college student were less likely to earn higher grades in Math 3A than their counterparts in the same course. We did not observe a statistically significant or practical difference by sex in Math 3A. The differences in LS30A grade distributions between subgroups with respect to all four social identity characteristics did not meet our threshold for statistical significance ($\alpha = 0.01$). That said, the practical differences between median grades by subgroup in LS30A (0.3 to 0.7) are comparable to what we observed in Math 3A (0.3 to 0.7), except for sex, yet the corresponding effect sizes are negligible to small for LS30A (effect size -0.04 to -0.12). We attribute the lack of detectable significant differences in median grades to the smaller sample size (N) for LS30A compared with Math 3A (Gelman and Stern, 2006). Nonetheless, these findings support our hypothesis in showing that the grade gaps between students who identified as a member of a racial/ethnic group underrepresented in STEM, as lower SES, or as a first-generation college student are practically smaller, as approximated by effect size, in LS30A as compared with Math 3A (see Table 6). There is some indication, however, that a performance gap exists for women and lower-income students in LS30A, suggesting additional improvements to the instructional approach are merited.

Gains in Student Interest in Contextualized Math Courses

Our analysis of grade data for science courses provides direct evidence of the cognitive benefits of the contextualized math curriculum (see Tables 3 and 4). Confirmation of a reduced grade gap in LS30A compared with Math 3A with respect to student social identity groupings (see Table 6) suggests there are

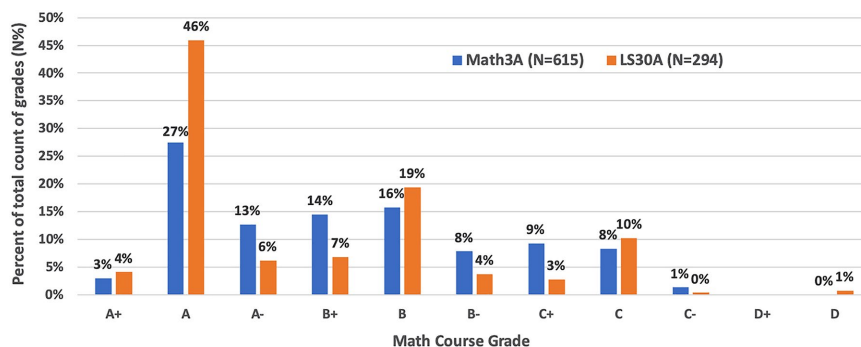


FIGURE 1. Grade distributions for math courses completed between 2013 Fall and 2017 Spring. Nonparametric Mann-Whitney U -tests confirmed a statistically significant difference in the distributions (medians) of math course grades.

likely noncognitive factors affecting classroom climate that positively impact student success in the contextualized math curriculum. We also were interested in examining how this transformational approach to teaching college-level math influences students' interest in the subject matter—that is, learning math in the context of biology. We predicted a positive shift in subject matter interest if we were successful in establishing the relevance of math to life sciences majors.

With access to SRI data across all 4 years of the study for both LS30A and LS30B, we analyzed the response frequencies for survey items that asked students to

TABLE 6. Descriptive statistics of unweighted mean and median grades in math courses for the sample population disaggregated by math class taken and student characteristics (unweighted sample N = 909)

	Math 3A (N = 615)						LS 30A (N = 294)					
	Parametric ^a			Nonparametric ^b			Parametric ^a			Nonparametric ^b		
	N	Mean	SD	Cohen's d	Significant difference	Median	r	Significant difference	Median	r	Significant difference	
Sex												
Male	204	3.29	0.70	0.09	0.293	3.30	-0.05	0.253	4.00	0.26	0.025	
Female	411	3.23	0.68			3.30			3.70			
Race/ethnicity												
Non-URG ^c	468	3.35	0.65	0.63	0.000	3.30	-0.26	0.000	4.00	0.11	0.506	
URG ^c	147	2.93	0.70			3.00			3.70			
SES ^d												
No Pell Grant	365	3.38	0.64	0.47	0.000	3.70	-0.22	0.000	4.00	0.28	0.030	
Pell Grant recipient ^d	250	3.06	0.72			3.00			3.30			
Parent/legal guardian education ^e												
Continuing generation ^e	424	3.34	0.66	0.43	0.000	3.30	-0.19	0.000	4.00	0.15	0.268	
First-generation (4-year graduate)	191	3.05	0.72			3.00			3.50			

^aFor parametric statistics, t tests were used to compare mean grades, and Cohen's d coefficients were calculated as measure of effect size (small effect 0.2, medium effect 0.5, and large effect 0.8).
^bFor nonparametric statistics, Mann-Whitney U-test was used to compare grade distributions, and Pearson correlation coefficients (r) were calculated as a measure of effect size (small effect 0.1-0.3, medium effect 0.3-0.5, large effect >0.5).

^cURG subgroup includes students who identify as Black/African American, Latinx/Hispanic, Native American, or Alaskan Native as their race or ethnicity.

^dFederal Pell Grant status serves as proxy for SES; those students eligible to receive a Pell Grant report an adjusted gross family income of less than \$60,000 per year.

^eFor this study, continuing-generation students refer to those students with at least one parent or legal guardian who previously completed college with a bachelor's degree.

*p < 0.05.

**p < 0.01.

***p < 0.001.

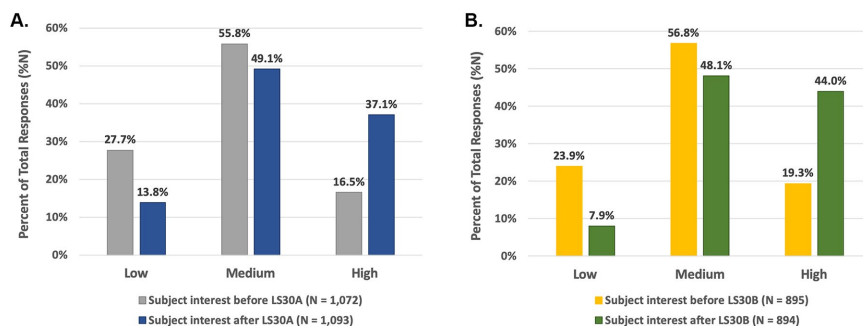


FIGURE 2. Changes in students' interests in LS30A (A) and LS30B (B) over the duration of each course. Response options were assigned numeric values on a 3-point scale: 1) low level of interest, 2) medium level of interest, and 3) high level of interest. Histograms reflect student responses comparing two relative time points: before and after each course. The z-tests confirmed a statistically significant positive shift in students' level of interest for each course.

retrospectively report their levels of interest in the subject matter at the beginning of the term and at the end of the term. We then looked for patterns corresponding to changes in student interest over time (see histograms in Figure 2). We used these particular survey items as a proxy for student attitudes about math in a biology context and motivation to persist in course work critical to the quantitative preparation of life sciences majors.

Overall, z-tests of the SRI data reveal that students had a statistically significant positive shift in their level of interest over the duration of each course as shown in Figure 2 for both LS30A ($p < 0.01$) and LS30B ($p < 0.001$). Before the course, fewer than 20% of students in LS30A (16.5%) and LS30B (19.3%) reported high levels of interest in the subject matter. However, after the course, nearly 40% of students in LS30A (37.1%) and LS30B (44.0%) reported high levels of interest. Moreover, fewer students expressed low levels of interest in the subject matter by the end of each course (13.8% for LS30A, 7.9% for LS30B) as compared with the start of each course (27.7% for LS30A, 23.9% for LS30B). Altogether, the SRI data suggest that this innovative instructional approach to teaching college-level math is improving student attitudes and motivating their sustained engagement in introductory math courses, which historically have been a barrier to equitable attainment of STEM degrees in life sciences. As noted earlier, there were no comparable SRI data available to us for Math 3A. However, these results do contrast prior survey results in which life sciences majors reported low levels of satisfaction with the Math 3 series (B.V.V., 2013, unpublished data).

DISCUSSION

The overarching goal of this study was to conduct a broad, course-level comparison of outcomes for students who completed either LS30A or Math 3A as their first math course at a large, public research institution. The study was propelled by interest in the cognitive and noncognitive benefits that a transformative math curriculum might afford first-year life sciences students, who frequently see calculus courses as unwelcoming and irrelevant obstacles en route to their undergraduate degrees in biological sciences (Bialek and Botstein, 2004; Steen, 2005;

Laursen *et al.*, 2011; PCAST, 2012). The new contextualized math curriculum was designed to provide students with opportunities to learn and practice quantitative and computational skills relevant to contemporary STEM careers and research in life sciences (Marshall and Durán, 2018). Notably, our study was further motivated as a means to promote buy-in from colleagues in chemistry and physics departments, assuaging their concerns about the new contextualized math curriculum and its efficacy in preparing students for their respective service courses. These apprehensions prompted us to track student grade outcomes not only in a life sciences course, but also in chemistry and physics courses required of life sciences majors.

The findings from our research study demonstrate that students who completed the contextualized math curriculum earned significantly higher grades in their science courses Chemistry 14A, Life Science 2, and Physics 6A (see Tables 3 and 4) compared with their peers who took traditional calculus for biology (Math 3A) as their first math course. Importantly, we are able to minimize course performance differences in their science courses attributable to dissimilarities in student characteristics or academic preparation by applying propensity score weighting to our sample population (see Table 2). This strategy is designed to address potential selection bias among enrolled students by accounting for differences in academic background characteristics (SAT scores, high school GPA, AP scores) and demography (sex, race/ethnicity, SES, first-generation status). The improvement in student performance as ascertained by grade data is consistent with our initial hypothesis for Life Science 2 and Physics 6A, where overlap with the new math courses in conceptual applications and skills was anticipated. That we also saw grade improvements in Chemistry 14A, where content connections were less obvious, lends support to the idea that the contextualized math courses might confer benefits beyond content knowledge. Aspects of the course structure and pedagogy may instill strategies in students that make them better learners. For example, with growth mindset being emphasized consistently by instructors (Dweck, 1999), we might speculate that LS30A students become more resilient in their reaction to setbacks (Master, 2015) and thus are better positioned to persevere through difficult material or science courses with a chilly or hostile classroom climate (Cabrera *et al.*, 1999; Yosso *et al.*, 2009; Jensen and Deemer, 2019).

Study findings also show that students earned higher grades in LS30A than their counterparts in Math 3A (see Table 5 and Figure 1). The dissimilarity in instructors' grading practices (see Table 1), however, is a confounding factor affecting our math grade comparison.

Instead, we turned our attention to ascertaining potential impacts of the contextualized math curriculum on historically persistent performance gaps in math grade outcomes, specifically with respect to sex, race/ethnicity, SES, and first-generation status. As shown in Table 6, the grade gaps for LS30A students, as assessed by effect size, in three of four subgroup comparisons are reduced relative to their Math 3A counterparts,

suggesting LS30A may indeed be creating a more inclusive learning environment for minoritized and first-generation college students (Dewsbury and Brame, 2019).

During the time frame of the study, LS30A grades were determined using criterion-based assessment practices, whereas Math 3A grades mostly were assigned using norm-referenced grading strategies such as grading on a curve (Schinske and Tanner, 2014). Research shows that grading practices can have an impact on classroom climate, in that criterion-based assessment tends to foster a collaborative learning environment and norm-referenced strategies typically promote a competitive learning environment (Hughes *et al.*, 2014), the latter of which can differentially and negatively impact minoritized students (Covington, 1992). By using inclusive grading practices (i.e., criterion-based assessment), in combination with actively endorsing a growth mindset among students, we posit that LS30A instructors are creating a supportive, positive classroom climate that promotes students' sense of belonging (Walton and Cohen, 2007), leading to improved academic achievement (Dewsbury and Brame, 2019). In addition, starting in 2015, a fraction of LS30A students (6.1%) began participating in a cohort-based undergraduate STEM student retention program that recruits science students who identify as members of racial/ethnic groups underrepresented in STEM, low-income students, and students who enter the institution with challenging life circumstances (Toven-Lindsey *et al.*, 2015). This cocurricular program was discontinued for Math 3A students during the study time frame. Being part of this learning community undoubtedly enriched the learning environment experienced by many minoritized life sciences majors in LS30A but was not sufficient to eliminate the performance gap for all subgroups (see Table 6). Future studies of the performance gap in the contextualized math curriculum could help to unpack the differential impacts of the various contributing factors on classroom climate.

Altogether, our new approach to teaching college-level math strengthens the academic preparation of all life sciences majors, including those students historically underserved and underrepresented in STEM. The cognitive gains made by LS30A students are complemented by gains in at least one noncognitive measure, students' interest in the subject matter (see Figure 2). Based on responses to two items on the end-of-term SRIs for both LS30A and LS30B, our study showed a statistically significant positive shift in students' subject matter interest upon completion of each course. These results are consistent with previous studies that showed improvements in students' attitudes toward math when it was taught within a biology context (Dufus and Olifer, 2010; Usher *et al.*, 2010; Eaton and Highlander, 2017; Aikens *et al.*, 2021). Furthermore, given prior evidence that increased subject matter interest can be a predictor of academic achievement (Schiefele *et al.*, 1992), these noncognitive outcomes support and reinforce the cognitive gains, particularly with respect to science course grades, revealed in this study. Finally, these results contrast with previous surveys of life sciences students who reported low levels of satisfaction with the traditional calculus for biology curriculum (B.V.V., 2013, unpublished data). This finding suggests teaching calculus using biology examples may not be sufficient for inspiring life sciences students' interest in math, its applications to quantitative science disciplines, and its relevance to the biology major. We con-

clude that increasing student interest in the contextualized math curriculum is likely a combination of the manner by which the mathematical concepts and skills are taught (i.e., focusing on mathematical modeling, integrating key calculus concepts as needed to support biological applications, and emphasizing conceptual understanding and computational applications relevant to understanding living systems), the structure (e.g., lower student to TA ratio, LA-supported secondary sections, computational lab), and instructional approaches that support an inclusive classroom climate (e.g., criterion-based assessment practices, emphasis on a growth mindset).

Implications of the Research Findings

Demonstrating that students in the contextualized math curriculum fared better in subsequent chemistry and physics courses led to a positive sea change in the attitude about the contextualized math curriculum and increased the confidence of our colleagues in the chemistry and physics departments, as evidenced by their support of a college senate curriculum committee proposal to credit the units earned from LS30A and LS30B completion toward the quantitative requirements of life sciences majors. Our research-driven approach to curricular change not only helped to solidify a formal agreement to modify the major requirements, but it also motivated the subsequent investment of chemistry and physics faculty in curricular change in their service courses for life sciences students. For example, the physics department initiated a major reform of the Physics 6 curriculum taken by life sciences students, replacing it in 2017 with a new curriculum that approaches the content and pedagogy in ways that better align with the interests and learning goals of the life sciences students it serves. Similarly, in 2019 the chemistry department began reviewing and revising the content, structure, and pedagogy of chemistry courses for life sciences students to enhance their success. Thus, in addition to documenting the positive student outcomes resulting from the implementation of the novel math curriculum, our research drove changes in teaching culture across the sciences.

An important goal in our continuing efforts to improve the quantitative skills of life sciences students is to ensure that the mathematical concepts learned in these contextualized math courses become an integrated component of the entire curriculum of a biology major. In other words, a companion approach to teaching math in the context of biology is to then teach biology in ways that apply the quantitative concepts and skills that students learn in their math courses (Usher *et al.*, 2010; Feser *et al.*, 2013). Such an approach might include using the same biological examples in both the math and life sciences courses to give students a more cohesive and consistent framework for applying and deepening their knowledge. This approach has been implemented successfully in biology courses with physics topics (Geller *et al.*, 2018) in which students' interest in physics was enhanced when covering a topic relevant to what they were learning in biology.

Other approaches might include using supplementary interactive Web modules to introduce quantitative exercises into biology courses (Thompson *et al.*, 2010) or using textbooks interwoven with examples that integrate math into biology concepts (Campbell *et al.*, 2020). A committee of life sciences instructors was recently established to explore this latter approach, integrating math examples from the LS30A/B series

into the new introductory biology courses. Yet another strategy might involve creating course-based research opportunities for students to expand their skills in modeling and coding by investigating complex, real-world problems relevant to the challenges faced by 21st-century biologists (NRC, 2009). We hypothesize that such a strategy might be an effective way to motivate or sustain biology students' interest in computer programming following its introduction in first-year math courses (Matthews *et al.*, 2010).

Research Study Limitations and Future Research

It is important to consider the findings from our research in the context of several limitations. First, in generating the sample population from registrar data, we selected students who conformed to the most common course-taking patterns. We did not examine the basis of the decision-making process that led others to complete their course work according to a less common sequencing pattern. Likely reasons might include scheduling conflicts or counseling during their first-year orientation by a student advisor with less experience or less knowledge of the courses or typical patterns of enrollment. Our sampling strategy aimed to minimize confounding factors that manifest in or affect course-taking patterns; however, we recognize that we may have inadvertently missed issues influencing the success or failure of those students who complete their math and science course work according to less common sequencing patterns. Future research examining student outcomes in the contextualized math courses should consider these smaller student populations separate from the majority group to identify potential advantages or disadvantages associated with less common course-taking patterns. This information could prove useful, for instance, in devising a more personalized, data-informed advising strategy for all students.

Observational data using the Classroom Observation Protocol for Undergraduate STEM (Smith *et al.*, 2013) indicate that LS30A instructors relied on lecture and real-time writing (“chalk talk”) as the primary instructional mode during the study period (M.K.E., unpublished data). Thus, there is still room to improve the pedagogy and classroom climate in the contextualized math courses, such as by integrating more active learning and inclusive teaching during lectures (Dasgupta and Asgari, 2004; Walton *et al.*, 2015; Theobald *et al.*, 2020). Such efforts could help to actualize closure of enduring performance gaps for women and lower-income students in the contextualized math courses (see Table 6). Future research should continue to include classroom observation data to inform pedagogical improvements in real time. This, in addition to instructor interview data, course syllabi, and student feedback, could be combined to support instructor groupings by various pedagogical criteria (e.g., Stains *et al.*, 2018) to parse out math section-level differences and other nuances specific to the student experience, none of which was possible in a study designed to measure broader course-level differences. By extension, our statistical analyses could be expanded to include multilevel analyses and to explore possible interaction effects of variables in our main effects models, which could lend insight into math section-level or instructor-level differences. Both are important to consider in future studies of LS30A and LS30B.

In our study, we did not investigate attrition data or DF grade frequencies for the math courses, because our outcome

variables were the subsequent science course grades, which required students to complete their first math course with a passing grade. No-pass rates, in particular, are a substantive area of inquiry in ongoing LS30A/B curricular development efforts, in part, due to their negative impact on STEM persistence and/or time to degree (Stewart *et al.*, 2015; Yue and Fu, 2017). Toward the goal of providing a more individualized learning path for less-prepared students, we are integrating supplemental online learning modules designed to eliminate blind spots that impede a novice student's understanding of math concepts (Lee *et al.*, 2018). This approach has been shown to increase student engagement, boost exam scores without compromising rigor, and reduce course attrition rates, particularly for women, in a bioinformatics course. A resulting follow-up study should reveal whether or not such an intervention is sufficient to close the achievement gap that persists for women in the contextualized math curriculum, reduce the frequency of DF grades in these courses, and improve the learning of all students.

One final limitation involves consideration of the SRI data and other sources of self-report data. Ideally, we would have compared the shift in subject matter interest in the contextualized math courses, LS30A and LS30B, to that observed for our traditional calculus for biology courses, Math 3A, 3B, and 3C. However, per institutional policy, access to SRI data requires instructor permission, and we were only given permission by the LS30A and LS30B instructors. Nevertheless, the SRI data we did obtain covered the same time frame as the institutional data used to measure cognitive gains and thus told a parallel story about the noncognitive gains made by the same student population. However, because the SRI data are anonymous, we were unable to link student responses to registrar data, thereby hampering further studies. For example, we could not explore whether increased interest in the subject matter among LS30A students was directly correlated with improvements in individual student achievement as predicted by previous studies (Schiefele *et al.*, 1992). In addition, these SRI data come from a homegrown, unpublished instrument that has not undergone validity testing (see Supplemental Figure S1), so future studies that assess student interest and other noncognitive measures would be strengthened by using a validated survey instrument.

As part of a formative evaluation of the new math courses, postcourse surveys were administered to students at the end of LS30A and again at the end of LS30B with items monitoring additional noncognitive measures, such as students' confidence in their science and math ability, relevance of course material to students' respective majors and career goals and its real-life applications, and factors affecting classroom climate and motivation to persist in the life sciences curriculum (see Supplemental Table S2). The overall response rate for these surveys was low (11.5%); they were administered to students only through 2016, and, consequently, the survey instrument never underwent validity testing. Despite these drawbacks, our provisional survey results align with and provide additional support for the findings from the SRI data analysis in suggesting that this transformative approach to teaching college-level math is positively changing student attitudes as well as increasing student motivation to persist in their math and science course work. Future research could entail relaunching survey efforts with incentives to improve response rates as well as administering them to a

comparison group such as Math 3A students, which would allow for both validity and reliability testing. Since the time of the original survey administration, at least two validated assessment instruments relevant to our study have been published that could provide additional insight into student outcomes in the contextualized math courses. Stanhope and colleagues designed an exam called BioSQuaRE that measures undergraduate students' quantitative reasoning skills within a biological context (Stanhope *et al.*, 2017), and Andrews and colleagues developed the Math-Biology Values Instrument, which measures life sciences majors' interest in using math to understand biology as well as the perceived utility of math in a life sciences career (Andrews *et al.*, 2017).

Concluding Remarks

The transformation of the introductory math curriculum for life sciences students was implemented in the face of significant pushback from STEM departments responsible for teaching service courses taken by life sciences majors in chemistry, physics, and math. A comprehensive research study of student outcomes was critical to obtaining unanimous vote approval of the contextualized math courses, LS30A and LS30B, by the institution's curriculum oversight committee in the senate. Thus, this endeavor is not only a research study for maximizing student learning in gateway courses critical to the persistence of life sciences students, but it also serves as a case study for overcoming cultural barriers in large, public, research-intensive institutions. Rigorous student assessment played an indispensable role in leveraging the success of an educational intervention to foster positive changes in how the vast majority of life sciences students at our institution are now engaged in learning math. Our findings clearly demonstrate that the innovative approach taken to teaching college-level math has had a positive impact on student learning in subsequent science courses and on narrowing performance gaps in math. It also has helped to inspire the interests of life sciences students in quantitative biology.

ACKNOWLEDGMENTS

This research study was supported, in part, by a grant to the University of California Los Angeles (UCLA), from the National Science Foundation's Improving Undergraduate STEM Education (IUSE) program (DUE award no. 1432804). Additional support was provided by UCLA's Center for the Advancement of Teaching Instructional Improvement Program (IIP grant no. 13-29) and UCLA's Center for Education Innovation and Learning in the Sciences (CEILS). We thank the Life Sciences (LS) Division for providing funding for the implementation of the new math curriculum. Special thanks to LS Dean V. Sork, whose leadership and vision for transforming the first-year academic experiences of all life sciences majors created an opportunity for ambitious course reform that otherwise would not have been possible. Considerable thanks to LS department chair F. Laski for providing the administrative home for the courses. We thank J. Shevtsov, E. Deeds, J. Keranen, S. Venugopal, K. McCully, and all other instructors, for their contributions to the development and teaching of the new math curriculum. We also thank our collaborator T. Hasson for her advice and leadership in modifying the STEM student retention program, PEERS, to align with and support students enrolling in the new math courses. Finally, we thank S. Shutta-Morgan for assisting with

data entry; J. Fregoso for helping with early versions of figures and tables; and science course instructors D. Pires, J. Casey, J. Samani, and S. Shaked for their critical interrogation of topics and concepts in the math courses.

REFERENCES

- Aikens, M. L., Eaton, C. D., & Highlander, H. C. (2021). The case for biocalculus: Improving student understanding of the utility value of mathematics to biology and affect toward mathematics. *CBE—Life Sciences Education*, 20(1), ar5. <https://doi.org/10.1187/cbe.20-06-0124>
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education—A call to action*. Retrieved November 3, 2020, from www.visionandchange.org
- Andrews, S. E., Runyon, C., & Aikens, M. L. (2017). The Math-Biology Values Instrument: Development of a tool to measure life science majors' task values of using math in the context of biology. *CBE—Life Sciences Education*, 16(3), ar45. <https://doi.org/10.1187/cbe.17-03-0043>
- Association of American Medical Colleges-Howard Hughes Medical Institute. (2009). *Scientific foundations for future physicians* (Committee report). Retrieved November 3, 2020, from www.hhmi.org/grants/sfp.html
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Benken, B. M., Ramirez, J., Li, X., & Wetendorf, S. (2015). Developmental mathematics success: Impact of students' knowledge and attitudes. *Journal of Developmental Education*, 38(2), 14–22.
- Bialek, W., & Botstein, D. (2004). Introductory science and mathematics education for 21st-century biologists. *Science*, 303(5659), 788–790. doi: 10.1126/science.1095480
- Cabrera, A. F., Nora, A., Terenzini, P. T., Pascarella, E., & Hagedorn, L. S. (1999). Campus racial climate and the adjustment of students to college—a comparison between White students and African-American students. *Journal of Higher Education*, 70(2), 134–160. <https://doi.org/10.2307/2649125>
- Callier, V., Singiser, R. H., & Vanderford, N. L. (2014). Connecting undergraduate science education with the needs of today's graduates. *F1000Res*, 3, 279. <https://doi.org/10.12688/f1000research.5710.1>
- Campbell, A. M., Heyer, L. J., & Paradise, C. J. (2020). *Integrating concepts in biology*. Retrieved June 10, 2021, from www.trinity.com/trubook-integrating-concepts-in-biology-by-campbell-heyer-paradise.html
- Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, 5(2). <https://doi.org/10.1126/sciadv.aau4734>
- Caudill, L., Hill, A., Hoke, K., & Lipan, O. (2010). Impact of interdisciplinary undergraduate research in mathematics and biology on the development of a new course integrating five STEM disciplines. *CBE—Life Sciences Education*, 9(3), 212–216. <https://doi.org/https://doi.org/10.1187/cbe.10-03-0020>
- Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, 25(10), 1302–1318. <https://doi.org/10.1177/0146167299258011>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum. Retrieved November 3, 2020, from <http://www.loc.gov/catdir/enhancements/fy0731/88012110-d.html>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Comar, T. D. (2008). The integration of biology into calculus courses. *Primus*, 18(1), 49–70.
- Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: A step-by-step approach*. Hoboken, NJ: Wiley.
- Covington, M. V. (1992). *Making the grade: A Self-worth perspective on motivation and school reform*. Cambridge: Cambridge University Press.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, 40(5), 642–658.

- Depelteau, A. M., Joplin, K. H., Govett, A., Miller, H. A., 3rd, & Seier, E. (2010). SYMBIOSIS: Development, implementation, and assessment of a model curriculum across biology and mathematics at the introductory level. *CBE—Life Sciences Education*, 9(3), 342–347. <https://doi.org/10.1187/cbe.10-05-0071>
- Dewsbury, B., & Brame, C. J. (2019). Inclusive teaching. *CBE—Life Sciences Education*, 18(2), fe2. <https://doi.org/10.1187/cbe.19-01-0021>
- Duffus, D., & Olifer, A. (2010). Introductory life science mathematics and quantitative neuroscience courses. *CBE—Life Sciences Education*, 9(3), 370–377. <https://doi.org/10.1187/cbe.10-03-0026>
- Dweck, C. S. (1999). *Essays in social psychology. Self-theories: Their role in motivation, personality, and development*. Hove, East Sussex, UK: Psychology Press.
- Eaton, C. D., & Highlander, H. C. (2017). The case for biocalculus: Design, retention, and student performance. *CBE—Life Sciences Education*, 16(2), ar25. <https://doi.org/10.1187/cbe.15-04-0096>
- Edelstein-Keshet, L. (2005). Adapting Mathematics to the New Biology. In Steen, L. A. (Ed.), *Math and Bio 2010: Linking Undergraduate Disciplines* (pp. 63–73). Washington DC: The Mathematical Association of America.
- Epstein, J. (2013). The Calculus Concept Inventory—Measurement of the effect of teaching methodology in mathematics. *Notices of the American Mathematical Society*, 60(8), 1018–1026. <https://doi.org/10.1090/noti1033>
- Fagerland, M. W. (2012). t-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Medical Research Methodology*, 12, 78. <https://doi.org/10.1186/1471-2288-12-78>
- Feser, J., Vasaly, H., & Herrera, J. (2013). On the edge of mathematics and biology integration: Improving quantitative skills in undergraduate biology education. *CBE—Life Sciences Education*, 12(2), 124–128. <https://doi.org/10.1187/cbe.13-03-0057>
- Garfinkel, A., Shevtsov, J., & Guo, Y. (2017). *Modeling life: The mathematics of biological systems*. New York: Springer International. <https://doi.org/10.1007/978-3-319-59731-7>
- Geller, B. D., Turpen, C., & Crouch, C. H. (2018). Sources of student engagement in Introductory Physics for Life Sciences. *Physical Review Physics Education Research*, 14(1), 010118.
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A. B., & Handelsman, J. (2013). Science education. Increasing persistence of college students in STEM. *Science*, 341(6153), 1455–1456. <https://doi.org/10.1126/science.1240487>
- Gross, L. J. (2000). Education for a biocomplex future. *Science*, 288(5467), 807. <https://doi.org/10.1126/science.288.5467.807>
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis*. Thousand Oaks, CA: Sage.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449–455. <https://doi.org/10.1037/0003-066x.53.4.449>
- Heinze, G., & Dunkler, D. (2017). Five myths about variable selection. *Transplant International*, 30(1), 6–10. <https://doi.org/10.1111/tri.12895>
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Abingdon, UK: Routledge.
- Hughes, B. E., Hurtado, S., & Eagan, M. K. (2014). *Driving up or dialing down competition in introductory STEM Courses: Individual and classroom level factors*. Washington, DC: Association of the Study of Higher Education. Retrieved September 26, 2020, from www.heri.ucla.edu/nih/downloads/ASHE2014-Competition-in-Introductory-STEM-Courses.pdf
- Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, 102(4), 880–895.
- Jensen, L. E., & Deemer, E. D. (2019). Identity, campus climate, and burnout among undergraduate women in STEM fields. *Career Development Quarterly*, 67, 96–109. <https://doi.org/10.1002/cdq.12174>
- Laursen, S., Hassi, M.-L., Kogan, M., Hunter, A.-B., & Weston, T. (2011). *Evaluation of the IBL Mathematics Project: Student and Instructor Outcomes of Inquiry-Based Learning in College Mathematics, Assessment & Evaluation Center for Inquiry-Based Learning in Mathematics* (Report to the Educational Advancement Foundation and the IBL Mathematics Centers, Issue). Retrieved November 3, 2020, from https://www.colorado.edu/eeer/sites/default/files/attached-files/iblmathreportall_050211.pdf
- Lee, C. J., Toven-Lindsey, B., Shapiro, C., Soh, M., Mazrouee, S., Levis-Fitzgerald, M., & Sanders, E. R. (2018). Error-discovery learning boosts student engagement and performance, while reducing student attrition in a bioinformatics course. *CBE—Life Sciences Education*, 17(3), ar40. <https://doi.org/10.1187/cbe.17-04-0061>
- Leech, N., & Onwuegbuzie, A. J. (2002). A call for greater use of nonparametric statistics. Paper presented at: Annual meeting of the Mid-South Educational Research Association (Chattanooga, TN). Retrieved August 27, 2021, from <https://files.eric.ed.gov/fulltext/ED471346.pdf>
- Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. *Psychometrica*. Retrieved October, 2019 from https://www.psychometrica.de/effect_size.html
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23, 151–169. <https://doi.org/10.1146/annurev.publhealth.23.100901.140546>
- Marshall, J. A., & Durán, P. (2018). Are biologists getting the mathematical training they need in college? *Biochemistry and Molecular Biology Education*, 46(6), 612–618.
- Master, A. (2015). *Praise that makes learners more resilient*. Retrieved January 25, 2021, from <http://mindsetscholarsnetwork.org/wp-content/uploads/2015/09/Praise-That-Makes-Learners-More-Resilient.pdf>
- Matthews, K. E., Adams, P., & Goos, M. (2010). Using the principles of BIO2010 to develop an introductory, interdisciplinary course for biology students. *CBE—Life Sciences Education*, 9(3), 290–297. <https://doi.org/10.1187/cbe.10-03-0034>
- Moore, G. W., Slate, J. R., Edmonson, S. L., Combs, J. P., Bustamante, R., & Onwuegbuzie, A. J. (2010). High school students and their lack of preparedness for college: A statewide study. *Education and Urban Society*, 42(7), 817–838. <https://doi.org/10.1177/0013124510379619>
- National Research Council (NRC). (2003). *Bio 2010: Transforming undergraduate biology education for future research biologists*. Washington, DC: National Academies Press. <https://doi.org/https://doi.org/10.17226/10497>
- NRC. (2009). *A new biology for the 21st century*. Washington, DC: National Academies Press. <https://doi.org/https://doi.org/10.17226/12764>
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, 33(3), 301–314. <https://doi.org/10.1080/02602930701293231>
- Olmos, A., & Govindasamy, P. (2015). A practical guide for using propensity score weighting in R. *Practical Assessment, Research, and Evaluation*, 20, ar13. <https://doi.org/10.7275/jjtm-r398>
- President’s Council of Advisors on Science and Technology. (2012). *Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology. Retrieved July 6, 2019, from <https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports>
- Rheinlander, K., & Wallace, D. (2011). Calculus, biology and medicine: A case study in quantitative literacy for science students. *Numeracy*, 4(1), 3.
- Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, 13, 137–172.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387), 516–524.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, 39(1), 33–38.
- Rubinstein, A., & Chor, B. (2014). Computational thinking in life science education. *PLoS Computational Biology*, 10(11), e1003897. <https://doi.org/10.1371/journal.pcbi.1003897>
- Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In Renninger, K. A., Hidi, S., & Krapp, A. (Eds.), *The role of interest in learning and development* (pp. 183–212). Mahwah, NJ: Erlbaum.

- Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education*, 13(2), 159–166. <https://doi.org/10.1187/cbe.CBE-14-03-0054>
- Skovlund, E., & Fenstad, G. U. (2001). Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? *Journal of Clinical Epidemiology*, 54(1), 86–92.
- Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12(4), 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... & Young, A. M. (2018). Anatomy of STEM teaching in North American universities: Lecture is prominent, but practices vary. *Science*, 359(6383), 1468–1470. <https://doi.org/10.1126/science.aap8892>
- Stanhope, L., Ziegler, L., Haque, T., Le, L., Vinces, M., Davis, G. K., ... & Overvoorde, P. J. (2017). Development of a Biological Science Quantitative Reasoning Exam (BioSQuARE). *CBE—Life Sciences Education*, 16(4), ar66. <https://doi.org/10.1187/cbe.16-10-0301>
- Stewart, S., Lim, D. H., & Kim, J. (2015). Factors influencing college persistence for first-time students. *Journal of Developmental Education*, 38(3), 12–20.
- Talbot, R. M., Hartley, L. M., Marzetta, K., & Wee, B. S. (2015). Transforming undergraduate science education with learning assistants: Student satisfaction in large enrollment courses. *Journal of College Science Teaching*, 44(5), 28–34.
- Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences USA*, 117(12), 6476–6483. <https://doi.org/10.1073/pnas.1916903117>
- Thompson, K. V., Nelson, K. C., Marbach-Ad, G., Keller, M., & Fagan, W. F. (2010). Online interactive teaching modules enhance quantitative proficiency of introductory biology students. *CBE—Life Sciences Education*, 9(3), 277–283. <https://doi.org/10.1187/cbe.10-03-0028>
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 1(21), 19–25.
- Toven-Lindsey, B., Levis-Fitzgerald, M., Barber, P. H., & Hasson, T. (2015). Increasing persistence in undergraduate science majors: A model for institutional support of underrepresented students. *CBE—Life Sciences Education*, 14(2), ar12. <https://doi.org/10.1187/cbe.14-05-0082>
- Usher, D. C., Driscoll, T. A., Dhurjati, P., Pelesko, J. A., Rossi, L. F., Schleiniger, G., ... & White, H. B. (2010). A transformative model for undergraduate quantitative biology education. *CBE—Life Sciences Education*, 9(3), 181–188. <https://doi.org/10.1187/cbe.10-03-0029>
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96. <https://doi.org/10.1037/0022-3514.92.1.82>
- Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468–485. <https://doi.org/10.1037/a0037461>
- Yosso, T., Smith, W., Ceja, M., & Solórzano, D. (2009). Critical race theory, racial microaggressions, and campus racial climate for Latina/o undergraduates. *Harvard Educational Review*, 79(4), 659–691. Retrieved June 10, 2021, from www.jstor.org/stable/2696265
- Yue, H., & Fu, X. (2017). Rethinking graduation and time to degree: A fresh perspective. *Research in Higher Education*, 58, 184–213. <https://doi.org/10.1007/s11162-016-9420-4>