CALCULUS INSTRUCTION

Establishing a new standard of care for calculus using trials with randomized student allocation

Laird Kramer[1,2]*, Edgar Fuller[1,3,4]*, Charity Watson[1,3,4], Adam Castillo[1,3,4]†, Pablo Duran Oliva[1,3,5], Geoff Potvin[1,2]

Calculus, the study of change in processes and systems, serves as the foundation for many STEM disciplines. Traditional, lecture-based calculus instruction may present a barrier for students seeking STEM degrees, limit their access to STEM professions, and block their potential to address society's challenges. A large-scale pragmatic trial with randomized student allocation was conducted to compare two calculus instruction styles: active student engagement (treatment condition) versus traditional, lecture-based instruction (control condition). A sample of 811 university students were studied across 32 sections taught by 19 instructors over three semesters at a large, US-based Hispanic-serving institution. Large effect sizes were consistently measured for student learning outcomes in the treatment condition, which demonstrates a new standard for calculus instruction and increased opportunities for completion of STEM degrees.

Calculus instruction needs substantial transformation because it is often a barrier to STEM degree attainment, especially for traditionally underrepresented groups (*1*–*3*), depriving both individuals and society of the potential benefits of their inclusion. National calls for calculus transformation are numerous (*4*, *5*), because failing calculus can contribute to a student's departure from STEM degree programs. Only ~40% of students entering universities with STEM degree intentions actually graduate with a STEM degree (*6*). More concerning is that the odds of female students switching out of a STEM program after a calculus course is ~1.5 times higher than that of comparable male students (*3*). Furthermore, Hispanic and Black/African American students had >50% higher failure rates than white students in calculus (*7*, *8*).

Evidence-based instruction, which is implemented in many STEM disciplines, has reliably led to profound improvement in student success (*9*–*11*). However, common approaches to calculus instruction continue to rely on traditional, lecture-based practices in which students are passive learners in the classroom and are expected to construct their knowledge mostly outside of the classroom by doing homework or in recitation sessions (*12*). Mathematics as a discipline thus needs to embrace its role in enabling STEM careers that will lead to prosperity for both individuals and society at large. "Calculus … must become a pump and not a filter" for the STEM pipeline, as noted in 1988 by Robert White, president of the National Academy of Engineering (*13*). Handlesman *et al.* (*14*) recently argued that "we must fix the classrooms where many students from historically excluded communities are discouraged from pursuing STEM" and that "… the continued exclusive use of lectures is malpractice at best, or an act of discrimination at worst." Thus, it is imperative that substantial transformation in calculus instruction takes place to promote more equitable learning environments for all students.

We present a large-scale trial featuring randomized student allocation into treatment or control conditions to rigorously compare an evidence-based, active student engagement calculus course against traditional, lecture-based calculus instruction. The work extends prior calculus research investigations (*15*–*17*) by including random assignment of students to treatment and control sections, as well as anonymized analysis of the identical end-of-semester learning outcomes. The study uses a pragmatic (*18*) design with random allocation of students to inform on the effectiveness of similar interventions at higher education institutions, reflecting real-world classroom constraints. In these contexts, blinding of the treatment and control conditions to both students and faculty is not possible because blinding is only feasible when the treatment and control conditions remain unknown to the participants during the period of study (such as in a clinical trial drug study). As with other public health or sociological interventions, enrollment of participants in this study revealed some aspects of a cohort structure, but it was still possible to maintain the essential aspects of random assignment by following a modified protocol, as was done in Zwarenstein *et al.* (*18*). The treatment condition integrated a suite of coherent strategies that have been independently found (*19*, *20*) to improve student learning; thus, the treatment was a distinct departure from traditional instruction, and it was not logically possible for the treatment condition to remain hidden from students or faculty after the treatment began. Random assignment of faculty to control or treatment conditions would not be possible because an individual faculty member's knowledge, philosophy, and experience with a variety of classroom strategies and instructional practices may intersect with the features of the treatment or control conditions. The experimental protocol thus included a group of instructors willing to adopt the instructional methods in the treatment condition. This comprehensive experimental approach was intended to secure the strongest possible evidence for critical stakeholders to sustain the treatment beyond the trial.

The treatment condition used the modeling practices in calculus (MPC) curriculum and pedagogy, and the control condition represented the preexisting, traditional instructional practices at the study institution. MPC integrates the practices of mathematicians as a central design tenet throughout the course. Instructors facilitate students' application of mathematical "habits of mind" (*21*) that foster deeper understanding of calculus concepts, including the identifying of patterns, hypothesis development and testing, making connections, and communicating ideas precisely to learn calculus throughout the course. Class time is devoted to students working collectively in small groups on predesigned notes and learning activities developing their calculus understanding with minimal lecturing. Treatment included learning assistants (*22*) who were undergraduate peers integrated within the instructional team to facilitate student learning and promote culturally responsive instruction. The curriculum promotes mathematical practices (sense-making, problem solving, argumentation, etc.) and established strategies to optimize student engagement, including cooperative learning, argumentation and metacognition, mathematical fluency, and a culturally responsive environment (*23*) (described in the supplementary materials section 2). The MPC design builds on the SCALE-UP calculus model (*24*) and intentionally embodies well-established recommendations for calculus instruction, including ambitious teaching practices and strategies promoted by national mathematics societies and national reports (*12*, *20*, *25*–*28*).

The study was performed at Florida International University (FIU) in Miami, Florida, the fourth-largest public research university in the United States, with 58,787 students, of which 41,795 are undergraduates as of fall 2019 (*29*). FIU is a Hispanic-serving institution, with 64% of students identifying as Hispanic/Latino/a/. Moreover, 79% of the students identify as members of historically underrepresented racial/ethnic minority groups, and 57% are women. The institution's size provided a unique opportunity to perform this study because there are 18 to 34 40-student sections of calculus 1 being

[1]STEM Transformation Institute, Florida International University, Miami, FL 33199, USA. [2]Department of Physics, Florida International University, Miami, FL 33199, USA. [3]Center for Transforming Teaching in Mathematics, Florida International University, Miami, FL 33199, USA. [4]Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, USA. [5]Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA 23284, USA.
*Corresponding author. Email: Laird.Kramer@fiu.edu (L.K.); Edgar.Fuller@fiu.edu (E.F.)
†Present address: Department of Mathematics, The University of Texas at Arlington, Arlington, TX 76019, USA.

taught each semester and primarily serving STEM majors. Furthermore, institutional conditions created urgency to transform calculus because historic pass rates in introductory calculus averaged 55% (range 13 to 88%) over the six semesters before the project's pilot.

Research design

A pragmatic trial (*30*–*32*) of the MPC approach was performed during the fall 2018, spring 2019, and fall 2019 semesters to rigorously test student outcomes. Students were randomly assigned individually to treatment and control conditions at the beginning of the semester after enrolling in sections on the basis of their scheduling preferences using the institution's enrollment system. To accommodate the randomized assignments, each of the experimental sections doubled in size from the usual 40 seats to 80 seats before enrollment opening. Instructor names and section sizes were invisible to students throughout the enrollment phase. Just before each term, the 80-seat sections were split into two 40-seat sections by assigning each student at random to either a treatment or control section.

Once assigned, the treatment sections implemented the MPC approach, whereas the control sections were unchanged. After assignment, students were free to change/drop/add course sections up until the regular institutional drop/add deadline (7 days after classes began). To account for such changes, enrollments were monitored, and only students who were randomly assigned to either a treatment or control section and remained in that section through the regular, nonpenalty drop/add deadline were included in the data for the experimental study reported below. In total, 1019 students were randomly assigned to either the treatment or control groups. Of these, 516 students were assigned to the treatment group, and 417 remained in the section at the drop/add deadline. At the same time, 503 students were assigned to the control group and 394 students remained in the section at the drop/add deadline. Our study followed the Consolidated Standards for Reporting Trials (CONSORT) guidelines (*18*, *33*, *34*). The specifics of recruiting, enrollment, assignment, and completion for the trial are discussed in sections 1.3 and 3.1 of the supplementary materials. The randomization process produced comparable groups by mathematical background and demographics; class sizes were typical for the course (see section 3 of the supplementary materials).

Faculty participating in the study included seven individuals teaching 16 treatment sections, along with 12 individuals teaching 16 control sections. Faculty recruited to teach the treatment sections indicated a willingness to adopt and implement the MPC approach, replicating the authentic condition of faculty reforming their classroom practice under the study design. To prepare for the new instructional approach, faculty participated in a 2-day, pre-semester professional development workshop and were provided with the MPC curricular materials. Consistency of the MPC treatment was monitored through weekly preparation meetings in which the course objectives and pacing were discussed. In-class monitoring by the project team was deemed overly intrusive and disruptive to classroom engagement. Control section faculty were not guided to use any particular practices and chose their normal instructional approach, best described as traditional lecture format with at most limited student engagement. Potential effects of instructor differences on learning outcomes were investigated and are presented in section 3 of the supplementary materials and summarized below.

The student outcome measures reported include identical end-of-semester learning measures, as well as course success data (i.e., course grades). The end-of-semester learning measures focused on evaluating learning using a set of identical assessment items (problems) developed by instructors spanning all calculus sections and spanning the major learning objectives of a calculus 1 course. The aim was to determine how well students understood essential elements of, and exhibited fluency and technical competency in, calculus at course end. Assessment items aligned to both local and national standards (*35*) were embedded in a cumulative final examination and were administered to all students in each treatment and control section. To ensure fidelity and fairness to both treatment and control sections, control and treatment faculty collaboratively developed a set of items to be administered to both conditions in identical format and wording. This set of identical items formed roughly two-thirds of the total final examination content, with the remaining items added by individual faculty in a separate section of the examination, allowing them to address their specific instructional goals. Furthermore, the examinations and problems were formatted identically and without course section identifiers to allow completely anonymized evaluation during the subsequent comparative analysis. The identical items covered core calculus topics including evaluating limits, identifying extrema, curve sketching, related rates, and evaluating indefinite integrals. For the second and third semesters, additional items focusing on implicit differentiation and optimization were added to the identical set of items (for details, see section 3.3 of the supplementary materials). Course success data (grades) reflect the overall assessment of students as assigned by each section's instructor. Course grade policies were established by individual instructors following departmental syllabus guidelines and were broadly consistent across sections and semesters.

Analysis of the end-of-course learning measures used a rubric for each problem, with five researchers testing the initial rubric on a subset of examinations to establish interrater reliability. The final rubric represented consensus on all elements and accounted for initial ambiguity or disagreement. The analysis was performed by a team of 10 trained evaluators, each of whom evaluated a completely anonymous set of student solutions. An average of two evaluators reviewed each solution for correctness on a scale from 0 to 100%. The evaluators were very consistent and interrater reliability was high (Cohen's kappa 0.827 in fall 2018 and 0.797 in spring 2019) (*36*, *37*). The same rubric was applied to the fall 2019 data given its high degree of agreement. Once all problems were evaluated, the research team deanonymized and sorted the results by treatment and control sections for the comparative analysis.

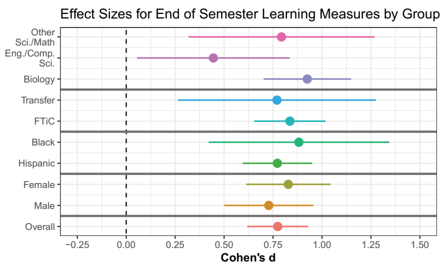

Fig. 1. Overall end-of-semester learning measures effect sizes (Cohen's *d*) broken out by major, race/ethnicity, and gender. Error bars indicate the 95% CI for effect size for each group.

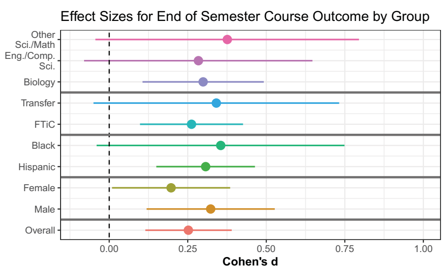

Fig. 2. Overall course success (earned grades of A, B, or C) effect sizes (Cohen's *d*) broken out by major, race/ethnicity, and gender. Error bars indicate the 95% CI for effect size for each group.

Results

The results indicate significant improvements in student learning for the MPC group across all three semesters. Students in the treatment group showed substantially higher scores on the identical end-of-semester learning outcomes: (fall 2018: d = 0.505, P < 0.01; spring 2019: d = 0.748, P < 0.001; fall 2019: d = 0.925, P < 0.001) compared with the control group. Combining results from all three semesters of trials (i.e., 32 sections and 811 total students), the overall standardized mean difference between treatment and control was d = 0.774 [95% confidence interval (CI) = 0.618 to 0.930] at the individual level, a medium/large effect size (*36*, *37*). An adjusted effect size (*38*) was computed using section-level cluster properties and the mixed-effects model structure described below and was found to be $d_T = 0.771$ (95% CI = 0.468 to 1.073; see section 3.4.2.1 of the supplementary materials).

The success of the MPC intervention could be seen across racial and ethnic groups, majors and academic pathways, and genders (Fig. 1). Similar medium/large overall effect sizes were observed for students in the treatment condition who identified as Black/African American ($d = 0.882$, $P < 0.001$) or Hispanic/Latino/a/ ($d = 0.772$, $P < 0.001$) when directly comparing the identical learning measures with their counterparts in the control condition. Although all STEM majors showed significantly improved learning, there were larger effect sizes for biology majors in the treatment group ($d = 0.925$, $P < 0.001$). Students matriculating onto campus as both first time in college (FTiC) and transfer students showed medium/large effect sizes and most were FTiC. Overall, treatment group students show more consistency in applying the tools of calculus to optimization problems, using derivatives to sketch graphs of functions, and in the evaluation of limits and integrals.

Potential biases arising in the random student assignment and faculty selections were investigated for hidden-level effects or confounders to establish limitations of the study (see section 3 of the supplementary materials). Random allocation of students provided equivariance in the demographics of student populations. Analyses showed that allowing students to drop/add sections during the open registration period after the initial assignment did not affect the measured outcomes. Faculty characteristics were compared and found to be similar in both background and prior course student grade distributions. A mixed-effects model with student fixed effects and random cluster effects due to section and instructor levels was fit with tests of fixed effects computed using Satterthwaite approximations (see section 3.2.4.1 of the supplementary materials). The explanatory power of the model was found to be high (conditional R^2 = 0.39, marginal R^2 = 0.31). The effect of treatment was statistically significant ($t_{(660)} = 5.68$, $P < 0.001$, semipartial $R^2 = 0.119$) (*39*), implying an estimated effect size of Cohen's $f = 0.368$ with covariates and cluster-level effects present. Random effects correlated with 0.085 of outcome variance (intraclass correlation coefficient = 0.13 with demographic covariates). A sensitivity analysis (see section 3.2.4.3 of the supplementary materials) showed that unmeasured confounders would need to be four times more powerful than any measured covariate, including student mathematics background, to be responsible for the observed effect.

Furthermore, students in the MPC treatment condition had improved course grades. Average grades were significantly higher by ~0.4 points (4.0 grade point scale) in MPC sections across all semesters of the study ($P < 0.001$, $d = 0.295$). This translated to success rates (A, B, or C grades) averaging 11% higher in MPC sections compared with traditional sections ($P < 0.001$, $d = 0.251$; Fig. 2). Outcomes were consistent across the three semesters of the experiment (Fig. 3). Moreover, the MPC sections also had lower course late drop rates (departure after the regular drop/add period ends) across all three semesters ($P < 0.05$, $d = 0.141$), suggesting that students more clearly perceived that they were likely to succeed in the course.

The trend of improved outcomes in course success was also observed for demographic subgroups, as can be seen in Fig. 2. A logistic regression model of success using gender identification, FTiC status, and Hispanic identification as independent variables showed the odds of a female-identified student in the treatment group passing the course to be 58% higher than the odds of a female-identified student in the control group ($b_1 = 0.46$, $P < 0.05$). Hispanic students' odds of passing the course were almost double that of their counterparts in the control group ($b_1 = 0.70$, $P < 0.001$). The likelihood of FTiC students in the treatment group passing the course increased by ~85% compared with FTiC students in the control group ($b_1 = 0.61$, $P < 0.01$) (for details, see section 3.4 of the supplementary materials).

Discussion and conclusions

This pragmatic trial demonstrates that student learning outcomes were significantly improved in the treatment condition. Contrary to previous research (*40*), this study shows that when students are expected to engage with calculus concepts collaboratively using intentional, evidence-based teaching strategies, they develop a better understanding of calculus concepts and techniques. The benefits of the MPC curriculum and pedagogy were realized regardless of racial/ethnic group, gender, or major/academic pathway. These trends suggest that the treatment includes culturally responsive and equitable strategies. Specifically, the MPC learning environment is designed to promote learning communities that provide ongoing support for learning mathematics through collaborative engagement and ongoing formative feedback. This aims to promote inclusion and increases access for students with different mathematical backgrounds, cultural identities, and life experiences by allowing them to use their mathematics skills in a supportive, nonthreatening environment.

The improved learning and course success for modeling practices in calculus reported in this study have profound implications for calculus instruction. This study demonstrates the substantial benefit to students of the MPC approach designed around established, evidence-based principles, and should motivate educators in mathematics and other STEM disciplines to adopt the same or similar approaches and conduct similar studies to replicate these findings. Improved student success also leads to more efficient student progress to graduation and boosts institutional effectiveness. Applying this study's 11% average improvement in pass rate to all 2000 first-time calculus students at FIU would translate to 220 additional students succeeding in calculus annually and reducing the instructional load by five sections annually. Extending this strategy to the ~300,000 students across the nation taking calculus 1 each

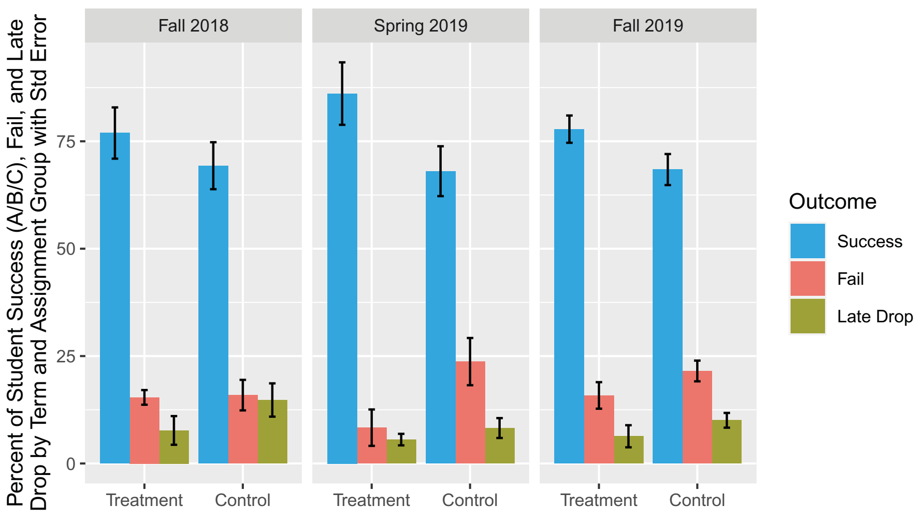

Fig. 3. Final course grade outcomes broken out by term and curriculum, including success (earned grades of A, B, or C), fail (earned grades of D or F), and late drops (withdrawals or drops after the institution's drop/add deadline). Vertical scale is percentage by outcome. Error bars indicate the 95% CI for the mean percentage of students in each outcome group over all sections in a term.

year, these results translate to the potential for an additional 33,000 students passing calculus each year, saving students an estimated $23.9 million on tuition [based on a three-credit course at the average public college/university tuition rate of $242/credit (*41*, *42*)]. Pragmatic trials provide guidance on what can be achieved by engaging faculty willing to change their instruction. These results potentially represent a lower bound on the long-term effects because faculty likely develop additional expertise through continued instruction and realize improved outcomes. The measured effect size provides the rationale to stop the control due to treatment benefit if one follows medical research protocols (*43*, *44*).

The experimental methodology described here establishes a new standard of care for calculus instruction and a high standard of evidence to bear on understanding the impacts on student learning. Improved learning of calculus aims to foster higher success in future STEM courses and develop the STEM "habits of mind" that students take with them into their future careers. Further, MPC shows the potential to address the disparities that differentially affect historically underrepresented groups, thus offering a mechanism to address Handelsman *et al.*'s (*14*) call to promote the success of historically excluded communities. We envision a mathematics experience for all students built on this approach and recommend that active student engagement must be deployed across all STEM disciplines to improve our development of future STEM professionals from all backgrounds.

REFERENCES AND NOTES

1. U. Treisman, *Coll. Math. J.* **23**, 362–372 (1992).
2. B. B. Alexander, A. C. Burda, S. B. Millar, *J. Women Minor. Sci. Eng.* **3**, 145–159 (1997).
3. J. Ellis, B. K. Fosdick, C. Rasmussen, *PLOS ONE* **11**, e0157447 (2016).
4. S. L. Ganter, D. J. Lewis, D. Hughes-Hallett, *Science* **279**, 2019–2025 (1998).
5. S. Olson, D. G. Riordan, *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President* (Executive Office of the President, 2012).
6. E. Seymour, N. M. Hewitt, *Talking About Leaving* (Westview Press, 1997).
7. E. Seymour, A.-B. Hunter, *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education* (Springer, 2019); http://link.springer.com/10.1007/978-3-030-25304-2.
8. A. K. Koch, B. Drake, *Digging into the Disciplines: The Impact of Gateway Courses in Accounting, Calculus, and Chemistry on Student Success* (John N. Gardner Institute for Excellence in Undergraduate Education, 2018).
9. M. J. Graham, J. Frederick, A. Byars-Winston, A. B. Hunter, J. Handelsman, *Science* **341**, 1455–1456 (2013).
10. M. Stains *et al*., *Science* **359**, 1468–1470 (2018).
11. C. Henderson *et al*., *Int. J. STEM Educ.* **4**, 14 (2017).
12. C. Rasmussen *et al*., *J. Res. Math. Educ.* **50**, 98–111 (2019).
13. L. A. Steen, in *Papers Presented at a Colloquium, Washington, DC, October 28-29, 1987* (ERIC, 1988); *MAA Notes Number 8*.
14. J. Handelsman *et al*., *Science* **376**, 1057–1059 (2022).
15. J. Bookman, C. P. Friedman, in *Assessment Practices in Undergraduate Mathematics*, B. Gold, S. Keith, W. A. Marion, Eds. (Mathematical Association of America, 1999), *MAA Notes*, p. 253.
16. M. Brown, in *Calculus: The Dynamics of Change*, W. Roberts, Ed. (Mathematical Association of America, 1996), vol. 39 of *MAA Notes*, pp. 52–58.
17. E. Tidmore, *Focus on Calculus* **7**, 5–6 (1994).
18. M. Zwarenstein *et al*., *BMJ* **337**, a2390–a2390 (2008).
19. S. Freeman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410–8415 (2014).
20. M. Abell, L. Braddy, D. Ensley, L. Ludwig, H. Soto-Johnson, *MAA Instructional Practices Guide* (Mathematical Association of America, 2018).
21. K. Saxe *et al*., *A Common Vision for Undergraduate Mathematical Sciences Programs in 2025* (Mathematical Association of America, 2015).
22. V. Otero, S. Pollock, N. Finkelstein, *Am. J. Phys.* **78**, 1218–1224 (2010).
23. G. Ladson-Billings, *Theory Pract.* **34**, 159–165 (1995).
24. L. Benson *et al*., in *Proceedings of the 2009 Annual Conference & Exposition, American Society of Engineering Education* (ASEE, 2009); pp. 14.159.1–14.159.7.
25. A. Kezar, A. C. Chambers, J. C. Burkhardt, *Higher Education for the Public Good: Emerging Voices from a National Movement* (Wiley, 2015).
26. D. Bressoud, *Math. Teach.* **109**, 179–185 (2015).
27. G. Sonnert, P. M. Sadler, S. M. Sadler, D. M. Bressoud, *Int. J. Math. Educ. Sci. Technol.* **46**, 370–387 (2015).
28. P. Zorn, E. Bressoud, W. David, M. Pearson, "Response to the PCAST Report to the President, Engage to Excel" (MAA, 2012); https://www.maa.org/sites/default/files/pdf/sciencepolicy/MAA-ResponsePCAST-E2E.pdf.
29. D. Burrows, "Common data set 2019-2020" (Florida International University CDS Archive, 2020); https://opir.fiu.edu/CDS/CDS2019.pdf.
30. C. J. Torgerson, *Education* **37**, 313–321 (2009).
31. M. Roland, D. J. Torgerson, *BMJ* **316**, 285 (1998).
32. B. Styles, C. Torgerson, *Educ. Res.* **60**, 255–264 (2018).
33. K. F. Schulz, D. G. Altman, D. Moher; CONSORT Group, *Trials* **11**, 32 (2010).
34. D. Moher *et al*., *Int. J. Surg.* **10**, 28–55 (2012).
35. M. A. Tallman, M. P. Carlson, D. M. Bressoud, M. Pearson, *International Journal of Research in Undergraduate Mathematics Education* **2**, 105–133 (2016).
36. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum, ed. 2., 1988).
37. J. C. Valentine, H. Cooper, "Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes" (US Department of Education, 2003); https://wmich.edu/sites/default/files/attachments/u58/2015/Effect_Size_Substantive_Interpretation_Guidelines.pdf.
38. L. V. Hedges, in *The Handbook Of Research Synthesis and Meta-Analysis*, H. Cooper, L. V. Hedges, J. C. Valentine, Eds. (Russell sage Foundation, 2009); 399–416.
39. B. C. Jaeger, L. J. Edwards, K. Das, P. K. Sen, *J. Appl. Stat.* **44**, 1086–1105 (2017).
40. M. T. Hora, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3024 (2014).
41. D. M. Bressoud, M. P. Carlson, V. Mesa, C. Rasmussen, *Int. J. Math. Educ. Sci. Technol.* **44**, 685–698 (2013).
42. T. Snyder, C. de Brey, S. Dillow, "Digest of education statistics, 2018" (NCES, 2019); https://nces.ed.gov/programs/digest/d19/tables/dt19_330.10.asp.
43. V. M. Montori *et al*., *JAMA* **294**, 2203–2209 (2005).
44. S. J. Pocock, *JAMA* **294**, 2228–2230 (2005).
45. Data for: L. Kramer *et al*., Establishing a new standard of care for calculus using trials with randomized student allocation, Dryad (2023); https://doi.org/10.5061/dryad.kkwh70s95.

ACKNOWLEDGMENTS

We thank the FIU administration and the State of Florida for initial support that made this project possible. The work would not be possible without the faculty, undergraduate learning assistants, and students participating in the study, who have earned our deep gratitude. We thank K. Rambo-Hernandez for insightful statistical analysis discussions, and we are indebted to the reviewers and editors for their insight and guidance. The research for this study was conducted under an institutional review board protocol approved by the Florida International University IRB (approval no. IRB-18-0211-a.m.02). Students participating in the study consented to participation at the beginning of each semester. Only students aged 18 or over were included. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. **Funding:** This work was supported by the National Science Foundation (grant DUE 1832450 to L.K., E.F., G.P., A.C., and C.W.). **Author contributions:** Conceptualization: L.K., E.F., G.P., A.C., C.W.; Funding acquisition: L.K., E.F., G.P.; Investigation: L.K., E.F., G.P., A.C., C.W., P.D.O.; Methodology: L.K., E.F., G.P., A.C., C.W., P.D.O.; Project administration: L.K., E.F.; Writing – original draft: L.K., E.F., G.P., A.C., C.W., P.D.O.; Writing – review and editing: L.K., E.F., G.P., A.C., C.W., P.D.O. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All experimental data are available at Dryad (*45*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade9803
Materials and Methods
Figs. S1 to S10
Tables S1 to S17
References (*46*–*109*)
MDAR Reproducibility Checklist

Submitted 20 September 2022; accepted 28 July 2023
10.1126/science.ade9803