Abstract Title Page
Not included in page count.

Title: "Measuring Student Success from a Developmental Mathematics Course at an Elite Public Institution"

Authors and Affiliations:
Julian Hsu, University of Michigan – Economics
William J. Gehring, University of Michigan – Psychology

Abstract Body

Background / Context:

Description of prior research and its intellectual context.

Developmental or remedial college courses are designed to help underprepared students. Although most of the literature examines developmental courses in community college, around 18% of students at four-year institutions take developmental courses (Sparks & Malkus, 2013). We evaluate a developmental course in such a public institution. Using administrative transcript data from an elite state university and a regression discontinuity strategy, we estimate causal impacts of taking a pre-calculus course on future course decisions, major declarations, and college persistence. Our work pushes current literature in a new direction - using instructor characteristics to explore driving factors of these impacts. Preliminary results show that the recommendations to take pre-calculus have significant bite, and taking pre-calculus helps low performing students' grades in future calculus courses. Looking at course-specific outcomes elucidates students' course-taking decisions and informs course-level academic policies.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

Our paper asks whether placement recommendations for a developmental math course at an elite public institution impact students' future academic performance, course-taking, and college outcomes. We use these specific outcomes to measure whether developmental courses help students develop the skills necessary to succeed in college, inspire them to take different courses, and overall help them graduate or persistent in college.

In addition to examining student outcomes, we dig deeper than the current literature and examine the ways in which instructor characteristics can drive these outcomes. We ask whether instruction at this university in a program for low-achieving students and particularly underprepared low-income, first-generation, and underrepresented minority students achieves its goal of reducing achievement gaps. This informs specific course and instructor policies to help underprepared students in their first semesters in college.

Setting:

Description of the research location.

Our research setting is an elite public institution. This institution is highly selective and has a large number of students, with an undergraduate enrollment of over twenty thousand. This institution house separate liberal arts, engineering, and business colleges.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

We collected? information on all students and instructors who enrolled or taught in all courses at this institution from 2002 to 2015.

Significance / Novelty of study:

Description of what is missing in previous work and the contribution the study makes.

Our study advances the literature studying developmental courses in three ways. First, we study the context of an elite four-year public institution, whereas the previous literature has focused on community colleges (Bettinger & Long, 2005,2009; Boatman & Long, 2010; Clayton & Rodriguez, 2012). At an elite institution with higher quality resources, it is extremely important for underprepared students to master the skills necessary to take advantage of the higher equality education.

Second, we use student transcript data to track student course-taking as a result of these developmental courses. Previous literature largely focuses on college persistence, dropout, and accumulated credits. An additional measure to augment these is course-taking: The literature has not answered whether students who take developmental courses are more likely to take Science Technology Engineer Mathematics (STEM) courses, or whether dropping out of college is related to the courses students concurrently taken with developmental courses. Answering these questions better identifies how developmental courses impact students' academic decisions, and can encourage STEM participation.

Third, we leverage instructor data to explore factors for measuring impacts, informing policy to help students early in college. Instructors can have significant effects on student outcomes such as major decisions (Ost & Main, 2014) and persistence (Price, 2010). Previous literature has found instructor impacts in general (Hoffman & Oreopoulus, 2009; Figlio, Schapiro, & Soter, 2015), but our work is the first to look at instructor impacts in developmental courses. This is particularly important for developmental courses, which are intended to have lasting impacts. Our focus on a program specifically targeting underprepared students seeks to determine the factors that can increase the likelihood of success for these students.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.
A major estimation hurdle is that students who take developmental courses are different from students who do not. A simple comparison of students who take developmental courses against students who do not is not an "apple-to-apples" comparison. For example, students with high SAT math scores most likely do not take developmental courses and cannot be compared against students who do. We use a regression-discontinuity (RD) framework to overcome this hurdle and use a fair comparison of students.

With our RD strategy, we can measure the causal impact of receiving a recommendation for developmental courses. In our context, students are recommended to take a math developmental course based on explicit cutoffs in a calculated index. The RD framework argues students in a narrow interval around the cutoffs are indistinguishable. This allows an "apples-to-apples" comparison of students who just managed to receive a recommendation to students who did not. Using this RD framework, we can identify the causal impact of receiving the recommendation to take a math developmental course.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.
One key assumption in RD is that students are indistinguishable around the cutoff. We test this assumption by plotting the characteristics of the students as a function of the calculated index and the cutoff (McCall & Biebly, 2012; Imbens & Lemieux, 2007). This index is calculated based on students' pre-college test scores: high school grade point average, SAT or ACT math scores, and a math placement exam. Therefore we expect to see a smooth correlation between the

index and student characteristics, but no discrete changes around the cutoff. We show several plots in Figures (1) – (3). These figures plot how the percent of black students, ACT and SAT math scores vary with the calculated index. It also plots predicted linear regressions of the index on these characteristics. These figures show the proportion of being black, ACT math and SAT math scores do not discretely change at the cutoff. The figures for other student characteristics look similar, showing a smooth pattern through the cutoff.

Another key assumption in RD is students cannot easily manipulate their indexes and thus the calculated recommendations. We do not believe this a large issue in this study. Recommendations are calculated through the Office of the Registrar, using a formula unknown to students. In addition, these recommendations are not binding and students are free to ignore them, eliminating any incentive to manipulate the recommendations.

Research Design:

Description of the research design.

The first step to applying our RD to the data is to recreate students' recommendations. We gathered the administrative formulas the Office of the Registrar used to calculate student recommendations, as well as accessed the student administrative data. We show in Figure (4) we are able to reproduce student recommendations. This verifies we are using the correct student data and applying the correct administrative formulas to calculate the index.

Next we select our sample to contain only students with calculated indexes in a narrow interval around the cutoffs used to determine recommendation. There are four mutually exclusive recommendations:

1. Definitely Take Pre-Calculus;
2. Tentative Take Pre-Calculus;
3. Tentative Take Calculus I; and
4. Definitely Take Calculus I or Higher.

Not only are there recommendations for different courses, but recommendations come in varying strengths. For our analysis, we focus on the difference between Definitely and Tentatively Take Pre-Calculus. In future analysis, we will focus on the difference between Tentatively Take Pre-Calculus and Tentative Take Calculus I.

Finally, we use a linear regression on our selected sample to estimate the causal impact of the recommendation. We include the index distance above and below the cutoff as additional controls. For a student *i*¸ we estimate the causal impact of the recommendations on a host of outcomes Y_i:

$$Y_i = \alpha(Receive\ Recommendation)_i + \beta_1(Index\ Distance\ Below\ Cutoff)_i + \beta_2(Index\ Distance\ Above\ Cutoff)_i$$

Where α is our coefficient of interest. α measures the difference in outcome Y_i between students who just managed to receive the recommendation and students who did not. In an RD framework, these students are indistinguishable.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

We gathered administrative student and instructor data from the Office of the Registrar. These data contain records for each student and instructor who participated in class taught at this elite public institution from 2002 to 2015. Student data includes not only standard demographic and admission test data, but also detailed transcript data. This allows us to track students' entire course history while at this institution. Transfer credit records are also collected. This is an incredibly rich dataset that allows us to measure a large variety of student outcomes, such as how many Accounting, Biology, Economics, or Psychology courses the student has taken in any given semester.

The instructor data is equally as rich. It contains demographic characteristics of instructors, as well as their status in the institution (faculty, lecturer, graduate student, etc.). For graduate student instructors, GRE and TOEFL scores are available. We are particularly interested in instructor evaluation data. We use instructors' evaluation histories to see whether instructor teaching quality ratings are related to student outcomes through the developmental courses. We are also able to relate the quality ratings in the developmental course to performance in subsequent classes.

Findings / Results:

Description of the main findings with specific details.

Our preliminary findings can be found in Figure (4). Using the linear regression and RD framework, we find students who receive a Definite Recommendation to take Pre-Calculus are more likely to take Pre-Calculus compared to students who receive only a Tentative Recommendation. This shows these recommendations have a significant effect on students' course-taking decisions, and is our first step to looking at how these recommendations affect how students choose other courses.

In future work, we will measure effects on other courses, such as whether students who receive a Definite Recommendation are less likely to take STEM courses. This allows us to see students' course-taking trajectories when they receive different recommendations. It also elucidates how students may choose to eventually drop out or persist in college.

To explain possible factors for these impacts, we will look for instructor effects among those who teach the Pre-Calculus and Calculus I courses as well as the effects of the university's program for underprepared students. The quality of instruction may contribute to the impact on students, and the university assumes that the program has beneficial effects.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

To our knowledge, our work is the first to evaluate developmental courses at an elite public institution. Our findings will inform policies to help underprepared students at higher quality institutions, where the benefits of mastering the college skills are higher than in community colleges. Our results on student course-taking links how students start college to college persistence and dropout. These developmental courses can be used not only to help students through college, but also enable them to enter specific fields they might otherwise avoid. Finally, our results on individual instructor and developmental program effects will help to inform training and hiring decisions to further increase the chances underprepared students will succeed.

Appendices

Appendix A. References

Bettinger, E. & Long, B.T. (2005). Remediation at the Community College: Student Participation and Outcomes. *New Directions for Community College, 129, 17-26.*

Bettinger E. & Long, B.T. (2009). Addressing the Needs of Underprepared Students in Higher Education: Does College Remediation Work? *Journal of Human Resources, 44(3), 736-771.*

Boatman, A. & Long, B.T. (2010). Does Remediation Work for All Students? How the Effects of Postsecondary Remedial and Developmental Courses Vary by Level of Academic Preparation. *National Center for Postsecondary Research Working Paper.*

Clayton, J. & Rodriguez, O. (2012). Development, Discouragement, or Diversion? New Evidence on the Effects of College Remediation. *NBER Working Paper 18328*, *Cambridge, MA.*

Figlio, D., Schapiro, M., & Soter, K. (2015). Are Tenure Track Professor Better Teachers? *The Review of Economics and Statistics, forthcoming*.

Hoffman, F. & Oreopoulus, P. (2009). Professor Qualities and Student Achievement. *The Review of Economics and Statistics, 19(1), 83-92.*

Imbens, G. & Lemieux, T. (2007). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics, 142(2), 615-6350*

Main, J. & Ost, B. (2014). The Impact of Letter Grades on Student Course Selection and Major Choices: Evidence from a Regression-Discontinuity Design. *The Journal of Economic Education*, *45(1), 1-10.*

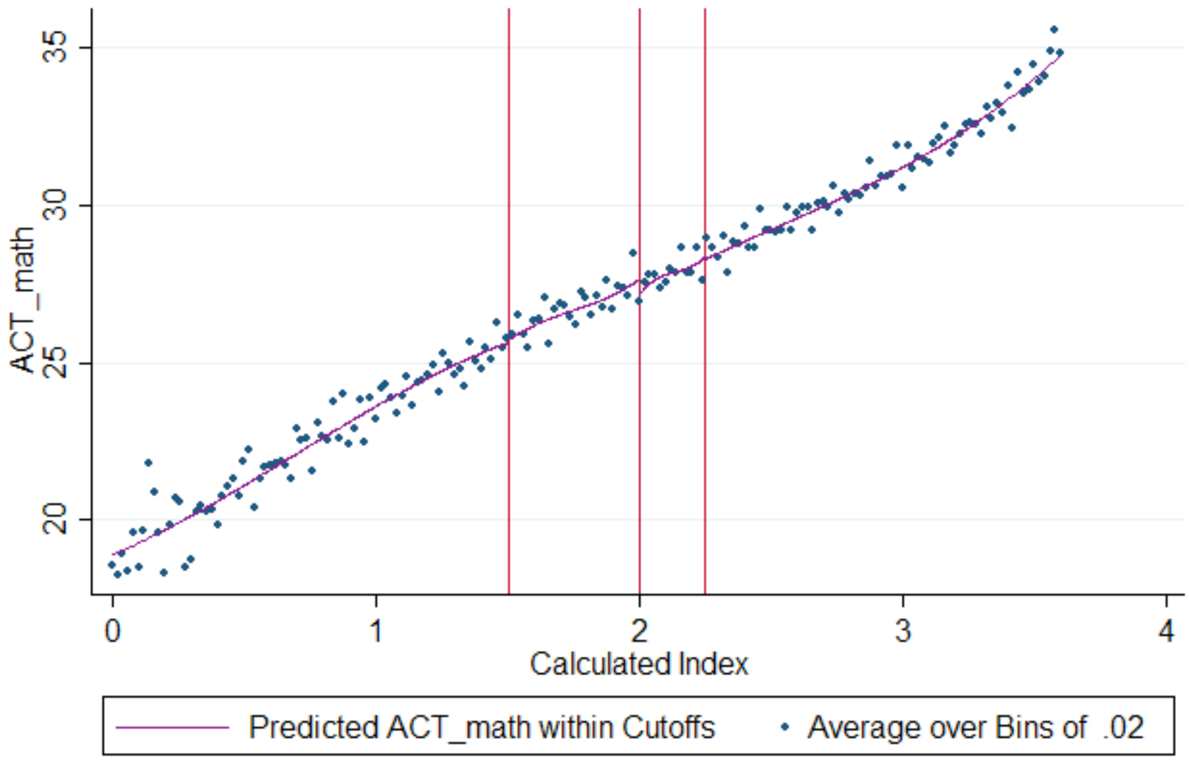
McCall, B. & Bielby, R. (2012). Regression Discontinuity Design: Recent Developments and a Guide to Practice for Researchers in Higher Education. In Smart, J. & M. Paulsen (Eds.). Higher Education: Handbook of Theory and Research XXVII. 249-290.

Price, J. (2010). The Effect of Instructor Race and Gender on Student Persistence in STEM Fields. *Education of Economics Review*, *29, 901-910.*

Sparks, D. & Malkus, N. (2013). First-Year Undergraduate Remedial Coursetaking: 1999-2000, 2003-04, 2007-08. *National Center for Education Statistics.*

Thistlethwaite , D. & Campbell, D. (1960) Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51(6), 309-317.*

Figure 1: Average ACT Math Score over the Calculated Index



Vertical lines at 1.5, 2, and 2.25 are the cutoffs for Placement Recommendations
Bins of .02 were used to create average of ACT_math, strictly for scatter and not line plot.
Calculated Indices below zero are not shown for graphical purposes

Figure 2: Proportion of Black Students over the Calculated Index

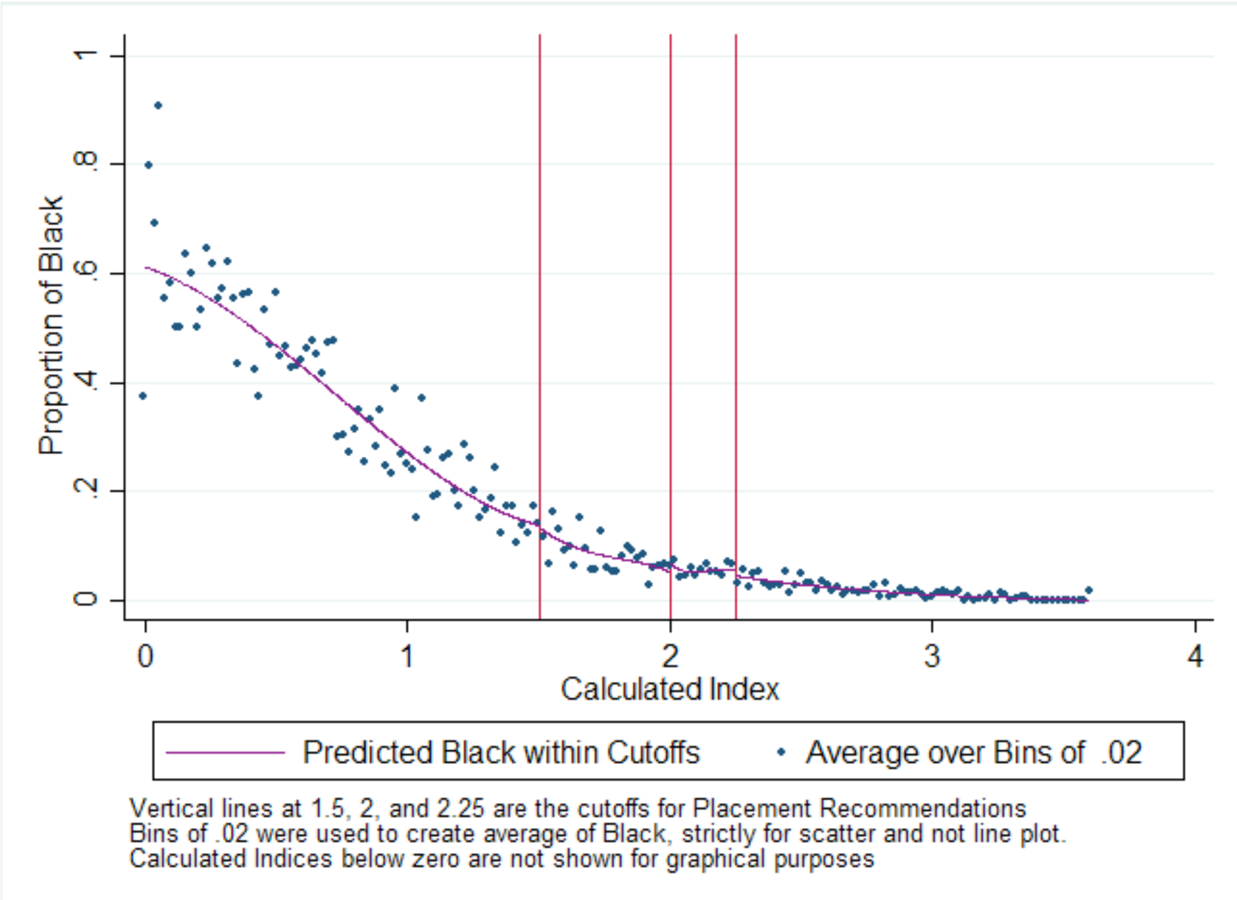


Figure 3: Average SAT Math Score over the Calculated Index

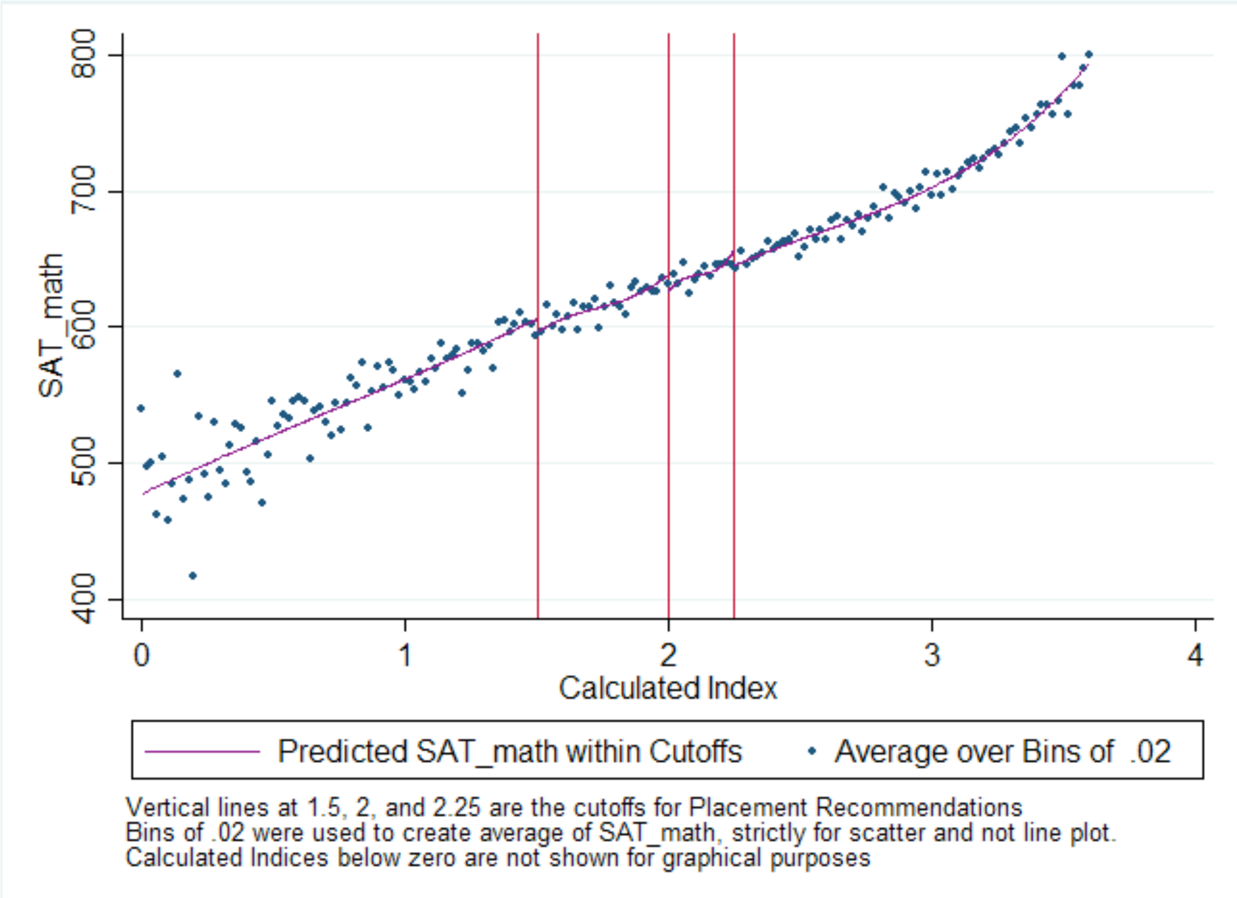


Figure 4: Preliminary Results of Ever Taking Pre-Calculus