Causal Quantitative Research and Implications for Developmental Mathematics Reform: A Literature Review

Zachary Beamer, EdD
Associate Professor of Mathematics
Piedmont Virginia Community College

Many students enter higher education with poor preparation for college mathematics courses. These students are often placed through high-stakes placement testing into developmental coursework that bear no credit towards graduation, but many students who begin in developmental coursework never succeed in a college-level mathematics course. Recent quantitative scholarship estimating the impacts of developmental placement and coursework on performance in gatekeeper courses and graduation has cast doubt on the effectiveness of common remediation practices. Such scholarship has prompted major ongoing reform efforts across the nation.This article highlights some of these findings, while exploring the assumptions of such research and the implications these assumptions have on how to interpret this scholarship.

Higher education institutions have long faced the challenge of providing adequate support to students ill-prepared for success in college-level mathematics. These students have often been assessed using high-stakes placement testing and required to take prerequisite developmental coursework before beginning a college-level, credit-bearing mathematics course. This problem is particularly acute across the nation's community colleges. According to the most recent nationwide longitudinal data from the National Center for Education Statistics (NCES), students entering two-year public colleges in 2003–2004 were almost twice as likely to enroll into developmental mathematics as their four-year peers (59% versus 33%; Chen, 2016). However, only 45% of the students at two-year colleges who enroll in developmental mathematics eventually earn college-level mathematics credits (Chen, 2016). Since students beginning remediation are disproportionately more likely to be Black or Hispanic, low-income, and the first of their family in college (Chen, 2016), this reflects yet another educational inequity that merits attention. Furthermore, given the massive disruptions to K-12 and higher education precipitated by COVID-19

in 2020, the number of underprepared students may still increase, exacerbating these challenges even further.

In recent years, a substantial amount of scholarship has sought to evaluate remediation in higher education as a *process*, one that starts with placement, followed by developmental coursework, college coursework, and (perhaps) degree completion and successful entry into the workforce. Some longitudinal research by Bailey, Jeong, and Cho (2010) finds that the more remediation students required, the less likely they were to make it through their remedial sequence. While this result itself may be unsurprising, the magnitude of the cumulative effect of additional remediation is dramatic. For students needing one, two, and three or more levels of remedial mathematics to enter credit courses, the percentages who complete all remedial requirements are 44%, 29%, and 16%, respectively. Effectively, each additional level of remediation required cuts the overall likelihood of getting to a credit-level mathematics course nearly in half.

There is some reason to doubt how generalizable these figures in Bailey et al. are, given that the colleges in the study had lower instructional expenditure per student and higher rates of at-risk populations. Nevertheless, such research has prompted some skepticism toward remediation, with some scholars exploring the hypothesis that the poor outcomes of developmentally placed students are *caused* by remedial placement itself (Martorell & McFarlin, 2011; Scott-Clayton & Rodriguez, 2015). According to this line of reasoning, placement into remediation sends a signal that a student is unprepared for college, leading the student to feel stigmatized. This discouragement is amplified by the fact that these students' first experience with mathematics in college is much the same as the curriculum that they struggled to master in middle and high school. Furthermore, being assigned to remediation delays the return on investment in a college education, as transfer or degree attainment becomes a semester or more out of reach. Finally, if students receive developmental education that is ineffective or poorly aligned with college-level learning objectives, the minimal gains may not justify the costs in time and money to students or colleges.

While the low success rates in Bailey et al. (2010) might be interpreted as saying that developmental mathematics is a barrier to student success or even a programmatic failure, the matter merits a more detailed discussion. These figures leave open several questions: how much developmental coursework impacts success in future courses, what happens to success rates when marginally prepared students bypass remediation, and whether alternative placement mechanisms might improve the num-

ber of students succeeding in college coursework. This article explores in detail the methods that recent quantitative scholarship has utilized to measure the impact of developmental placement and coursework on student outcomes. The piece notes some limitations of such methods and how they affect how to interpret this research to inform developmental mathematics educator reforms.

Quantitative Estimates of Developmental Mathematics on Student Outcomes

For decades, researchers in developmental education have sought to measure the effectiveness of receiving remediation using quantitative methods, but much of the early quantitative research had serious methodological flaws (O'Hear & MacDonald, 1995). Martorell and McFarlin (2011) note that the most prominent of these issues is the inability of such studies to account for differences in remediated and nonremediated populations. In other words, directly comparing these two groups on various measurable outcomes (e.g., credits earned, GPA, etc.) fails to account for differences between the groups that could drive variation in student outcomes. Factors that may lead to placement into remediation, such as academic preparation or challenging life circumstances, may also impact measurable student success. To be able to rigorously measure the impact of remediation itself, it is critical to reduce the effect of confounding variables. One statistical technique that attempts to remove observable sources of bias is the use of ordinary least squares (OLS) regressions.

In an influential study on developmental placement measures using OLS regression, Ngo and Kwon (2015) estimate the effects of "multiple measures" placement into credit-level mathematics in California community colleges. These multiple measures gave students a bonus to their placement test score based on their high school GPA or whether they had taken more advanced mathematics coursework. At two colleges explored in detail, these measures boosted 4.2% of students into the next level of coursework. The authors use an OLS regression to address the question of whether these boosted students (who otherwise would have fallen just short of placement into credit-level mathematics) performed differently from their peers who were directly placed based on placement test score alone. An OLS regression improves upon a direct comparison of the success rates of the two groups by including placement test scores as a control variable, measuring the amount that placement test scores covary with outcome variables. Additionally, the model includes other demographic factors that may have an association with student success, such as gender, race, and age. The model then uses the data to measure how much

variation in outcomes is correlated with placement test scores and demographic factors, and how much is correlated with the boost itself.

Using this model, the authors find that students boosted into higher levels of mathematics performed similarly to higher scoring peers in terms of pass rates and credit completion. This research provides some evidence towards the conclusion that at least some students are unnecessarily assigned to remedial coursework by inaccurate high-stakes placement testing. However, when using studies based on OLS regression to inform policy decisions, it is worth noting the methodological limitations of such research, most notably that of selection bias. That is, the mechanism that assigns individuals to treatment (in this case, being boosted by GPA or coursework into credit-level coursework) may not be random. Consequently, the observed differences between treatment groups may be driven by unobserved factors rather than the treatment itself. For example, students with higher GPA who were boosted into the credit-level course may have earned a higher GPA because they were highly motivated, were able to advocate for higher grades, or had access to more educational resources. Since such factors could also have an impact on placement outcomes and success rates in college, they cannot be ruled out as explanatory variables. Furthermore, no amount of additional control variables can guarantee that selection bias has been eliminated in OLS regression (Angrist & Pischke, 2014). As a result, the method of OLS regression does not provide *causal* effects of treatment, but rather an estimate of how effects *correlate* with treatment by adjusting for the impacts of other predictive factors.

Quasi-Experimental Estimates of Traditional Developmental Education

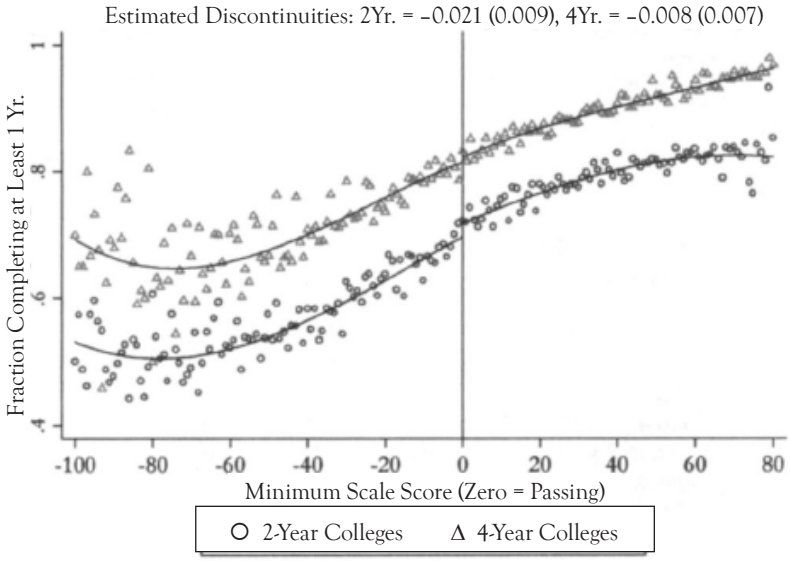
Given the limitations of OLS regression, researchers in the past decade have increasingly employed quasi-experimental methods such as regression discontinuity (RD) designs and instrumental variables (IV) to estimate the causal effect of developmental education on student outcomes. RD designs, like all quasi-experimental quantitative research, are an attempt to address the *fundamental problem of causal inference* (Holland, 1986). That is, for any given individual, it is impossible to observe both the effects of receiving treatment and not receiving treatment. In this case, an individual is either assigned to receive remediation or not; there is no data on the counterfactual scenario in which the individual receives the opposite assignment. Experimental research addresses this problem using randomization to create statistically equivalent groups and assign one to a particular treatment. However, in many circumstances it is im-

practical or unethical to design randomized groups. Quasi-experimental research uses various statistical techniques to reduce bias (such as selection bias) present in the naïve estimate of treatment—the difference in outcomes between treated and nontreated groups.

The way that RD designs overcome this bias is to utilize score cutoffs that determine treatment assignment. Colleges that make placement decisions based on whether a student receives above or below a given score on a placement test are an excellent opportunity to estimate treatment effects using an RD. Measurement error of standardized tests, the fact that an individual's score varies from attempt to attempt, serve as a sort of randomizing process. That is, an individual who avoided remedial placement by a single point could have by chance missed one additional question and been placed into developmental coursework instead. As a result, individuals clustered near the cut score are effectively sorted randomly to one side of the score or another. This means that the students immediately above and below the cutoff differ on little other than their assignment to remediation. When placement test score is included along with other covariates, the difference in outcomes predicted by regression curves at either side of the cutoff can be interpreted as the effect of treatment (Jacob et al., 2012)—that is, provided certain assumptions are met.

These studies are perhaps best understood by looking at a figure from one such study. In Figure 1 below, there is a score cutoff at zero, below which students are placed into remediation and above which students go directly into credit-level coursework. The reason these studies are called regression discontinuity is because of the sharp separation in the variable that determines treatment, and which can also sometimes be seen in the measurable outcomes. As in the figure below, there may be a visual jump in the outcome variable when moving from the left side of the cutoff to the right. The magnitude of the discontinuity yields the size of the effect of treatment. In the figure below, the jump for the data at two-year schools estimates that students who score just under the cutoff and are placed into remediation are slightly less likely to complete at least one year of college. After accounting for covariates, Martorell and McFarlin estimate assignment to remediation decreases the probability of completing at least a year by about 6%, though this result is only significant when accounting for covariates.

Figure 1. *(Martorell & McFarlin, 2011, p.* 446)



However, in other circumstances, there is not a visible jump from one side of the score cutoff to the other. This can be seen with the four-year college data above, and in other results in Martorell and McFarlin. For instance, they find that placement into remediation has no statistically significant impact on the proportion of students graduating within six years. Other authors using an RD, such as Scott-Clayton and Rodriguez (2015), find no impact, either positive or negative, of assignment to developmental mathematics on overall degree completion, credits earned, or grades in subsequent mathematics coursework. However, the use of this method is complicated by the fact that some individuals do not receive their assigned treatment. That is, some students elect to enroll in developmental mathematics despite placement into credit-level or may petition to start in credit-level coursework despite a lower placement. When individuals do not receive the treatment assigned, a problem known as *crossover*, an RD gives effects of *assignment* to treatment, which may not correspond to the treatment itself (Jacob et al., 2012).

Other researchers have dealt with this problem of crossover by using a "fuzzy" RD, also known as an RD-IV design. Such studies combine RD with instrumental variables. Instrumental variables use a two-stage least squares regression to predict the likelihood of taking up treatment based on an instrument (often the assignment to treatment). When certain assumptions are satisfied, instrumental variables produce *treatment on treated* effects: the effect that a treatment has on the group who take it up,

known as *compliers* (Imbens & Lemieux, 2008; Jacob et al., 2012). When coupled with an RD, instrumental variables can remove the bias on effect estimates introduced by noncompliers. In this case, the compliers enroll into the course they are placed; noncompliers choose not to take remediation when it is assigned, or who enroll into remedial coursework despite qualifying into college-level mathematics.

Calcagno and Long (2008) use an RD-IV and find that students receiving math remediation were two to four percentage points more likely to persist from fall-to-fall, but no more likely to pass their first college-level algebra course, earn a certificate or associate's degree, or transfer. Boatman and Long (2010) find negative but statistically insignificant results of developmental math on year-to-year retention, passing college-level mathematics, and college credits completed within three years. In another RD-IV study using data from colleges in the 2004 cohort of the VCCS, Xu and Dadgar (2018) examine the effects of the lowest level of remediation, prealgebra, with the middle level of remediation, basic algebra. The authors find negative but statistically insignificant effects of receiving remediation in prealgebra on receiving credentials within four years or passing the first credit-level mathematics course.

Relevant Limitations of Regression Discontinuity Designs

Estimates from RD designs show mixed results on the impact of mathematics remediation on student outcomes. However, some degree of caution is needed when interpreting these estimates due to the limitations and assumptions of such designs. First, an RD gives local area treatment effects, meaning that the population of causal inference for an RD is only those individuals scoring near the cutoff (Imbens & Lemieux, 2008; Jacob et al., 2012). In this case, these estimates apply to students who might plausibly place into either remedial or college-level coursework, according to wherever the cut scores have been set. There is no fixed interpretation in RD research for what exactly it means to be "near" the cutoff (Jacob et al., 2012), but the effect is certainly not internally valid for the entire population of study. Goudas and Boylan (2012) claim that such research has been misinterpreted to characterize developmental education as a failing enterprise. In a direct response, Bailey, Jaggars, and Scott-Clayton (2013) point out that different studies using RDs have looked at colleges using the same placement test (COMPASS) but different thresholds for the cut score. These scores, ranging from 27 through 81, represent a significant number of the scores (out of a possible 1 to 99). Consequently, they argue,

when these RDs are taken together, they provide evidence that these effect sizes apply to a rather broad range of students.

A second caution about RD designs is that they measure the effects of treatment on the treated. That is, the effects represent the impacts on students receiving remediation, not the impacts that remediation *would* have had on students who *did not* receive it. To explore that possibility among students scoring just above these score cutoffs, Moss, Yeaton, and Lloyd (2014) used a randomized experiment embedded within an RD. The authors imposed two separate score cutoffs, with the lowest scores assigned to developmental, the highest to credit-level, and finally a middle score range within which individuals were randomly assigned to either developmental or credit-level mathematics. The authors find that those students in the middle group who were randomly assigned to developmental mathematics performed approximately one-third of a letter grade better in their subsequent first credit-level mathematics course than those directly placed into credit-level coursework.

Another pair of considerations regarding RD research relate to potential issues regarding placement. The first issue is the uncertainty about whether the score threshold for placement into remediation is chosen in a way that accurately identifies those who would benefit from it. If the cut score were too high, the placement test might inaccurately place well-prepared students into a developmental mathematics they do not need. Were this the case, an RD comparing performance at the two sides of the cutoff would suggest that remediation was not beneficial. However, it would be more accurate to infer that remediation practices misidentified the potential beneficiaries.

While these RDs do not provide an indication as to whether the cut score is chosen accurately, research from Scott-Clayton, Crosta, and Belfield (2014) and Scott-Clayton and Rodriguez (2015) suggests that placement testing may indeed be unnecessarily assigning some students to remediation. Both studies estimate that upwards of one-quarter of students assigned to developmental mathematics could have otherwise earned a B or better in a gatekeeper mathematics course, based on statistical analysis of factors predictive of success.

The second issue with placement has to do with the methodological assumption required for RD of the *exogeneity of discontinuity*. This holds that there is not some nonrandom mechanism that is endogenously (nonrandomly) sorting individuals to one side of the score cutoff or another. For instance, the practice of allowing students the opportunity to retake a placement test after being assigned to remediation (as in some colleges

in Xu & Dadgar, 2018) could result in endogenous variation. That is, if more motivated students were more likely to retake the test, the populations at either side of the cutoff would differ along characteristics predictive of outcomes. Xu and Dadgar's (2018) estimates did not significantly change when they removed colleges whose data suggested they allowed students to retake placement tests. Nevertheless, variation in the specific placement practices of institutions may threaten this assumption.

A final issue with these quasi-experimental estimates is dependent upon the untestable assumption referred to as the *stable unit treatment value assumption* (Angrist, Imbens, & Rubin, 1996). This assumption holds that an individual's outcome is not dependent on the treatment status of other individuals. However, other research (e.g., Carrell, Fullerton, & West, 2009) notes that placing underprepared students into college-level mathematics can have negative peer effects, lowering the achievement of better-prepared students. Scott-Clayton and Rodriguez (2015) discuss three potential purposes of remediation: *development*, *discouragement*, or *diversion*. As discussed, their findings suggest that remediation does not significantly develop skills, nor discourage students from enrolling or persisting. Nevertheless, they argue that developmental courses could still serve the function of diverting less-prepared students away from credit-level courses.

Summary and a Vision for Reform

The quantitative literature offers evidence that prerequisite developmental coursework may be of limited benefit to some students, particularly those at the margins of requiring remediation. Utilizing data beyond placement test scores, such as high school GPA and coursework as in Ngo and Kwon (2015), may identify additional students who may be successful in college-level mathematics. However, the methodological limitations of RDs mean that these studies do not support the conclusion that prerequisite remedial coursework is completely unnecessary or of no value. Because of the issue of crossover, the effects produced by RD-IV designs cannot provide a meaningful estimate of the impact of remediation for students who choose to start remedial coursework despite a higher placement. Nor do they estimate the benefits such coursework might have had for students who placed into remediation but nevertheless started in credit-level coursework. Finally, the practice of mainstreaming the least prepared students directly into college-level coursework could have unintended negative peer effects that are unaccounted for in RD designs.

Taken together, this quasi-experimental research and its limitations indicates the need for reforms to developmental mathematics education

to balance the needs of various student groups assigned to remediation. Such reforms must address the common problem of improper alignment between developmental and credit-level coursework (Goldwasser, Martin & Harris, 2017). Additionally, these reforms ought to offer avenues for students of preparation, as well as those needing dedicated focused support on foundational topics, and accurately distinguish between these populations through informed placement measures.

One illustration of reforms can be seen in those currently being piloted by the Virginia Community College System (see Beamer, 2020). These reforms incorporate lessons learned from recent research into corequisite remediation, in which students receive remediation simultaneously with college-level coursework. Recent research suggests that such remediation may significantly increase student success rates in credit-level mathematics while reducing the amount of time required to complete remediation (Kashyap & Mathew, 2017; Logue, Watanabe-Rose, & Douglas, 2016; Ran & Lin, 2019; Royer & Baker, 2018). The primary aspects of the reforms currently taking place in Virginia are highlighted below.

- reform placement measures to reduce unnecessary remedial coursework
  - incorporate measures beyond high-stakes placement testing (HS GPA, mathematics coursework)
  - increase direct placement into credit-level coursework
- offer pathways for marginally prepared students that include targeted remediation
  - offer corequisite support for students just below placement thresholds into credit-level coursework
  - pair support courses to target remediation specific to credit-level courses
- ensure that pathways exist for the least prepared students to enter college mathematics
  - provide foundational support courses in arithmetic and algebra to offer a bridge to all levels of college mathematics
  - require different levels of remediation for entry into transfer-level courses and mathematics-intensive programs

These reforms, along with other corequisite reforms currently under study (see Daugherty et al., 2018), merit further exploration. Additional quantitative research, particularly of an experimental nature, would be able to improve upon the issues present in the quasi-experimental studies discussed here. Such research would be particularly beneficial in determining how to distinguish between students who (a) do not require

any remedial support to be successful in college-level mathematics, (b) students who can succeed in college-level mathematics with corequisite support, and (c) students who would benefit most from prerequisite developmental coursework. The design methods in Moss et al. (2014), for example, could compare the effects of using different cutoffs to determine eligibility into the randomized group. Doing so would offer insights into what sort of GPA or modified placement test score might be appropriate to set policy guidelines for remedial placement. Finally, qualitative research is necessary to provide further insights into the impacts of design and implementation on outcomes. There is considerable variation within and across institutions on how ongoing reforms such as corequisite instruction is being implemented (Daugherty et al., 2018). The final essential aspect needed to interpret the validity and generalizability of this quantitative research is a detailed description of how reforms have been implemented and the lessons learned during the process of implementation. Such details are crucial to illustrate the mechanisms by which various structures increase student success in college mathematics.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Angrist, J. D., & Pischke, J. S. (2014). *Mastering 'metrics: The path from cause to effect*. Princeton, NJ: Princeton University Press.

Bailey, T., Jaggars, S. S., & Scott-Clayton, J. (2013). *Characterizing the effectiveness of developmental education: A response to recent criticism*. New York, NY: Columbia University, Teachers College, Community College Research Center.

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270.

Beamer, Z. (2020). Mathematics corequisite remediation and direct enrollment: Addressing misconceptions and concerns. *VCCS Inquiry 23*(1). Retrieved from https://commons.vccs.edu/inquiry/vol23/iss1/9

Belfield, C., Jenkins, P. D., & Lahr, H. E. (2016). *Is corequisite remediation cost-effective? Early findings from Tennessee.* New York, NY: Columbia University, Teachers College, Community College Research Center.

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. *Educational Evaluation and Policy Analysis*, *40*, 29–58.

Calcagno, J. C., & Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance* (NCPR Working Paper). New York, NY: National Center for Postsecondary Research.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Chen, X. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes* (Statistical Analysis Report. NCES 2016-405).

Daugherty, L., Gomez, C. J., Carew, D. G., Mendoza-Graf, A., & Miller, T. (2018). *Designing and implementing corequisite models of developmental education.* RAND Corporation. Santa Monica, CA: Author.

Goldwasser, M., Martin, K., & Harris, E. (2017). A framework for assessing developmental education programs. *Journal of Developmental Education, 40*(2), 10–17.

Goudas, A. M., & Boylan, H. R. (2012). Addressing flawed research in developmental education. *Journal of Developmental Education*, *36*(1), 2–13.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*, 615–635.

Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. (2012). *A Practical guide to regression discontinuity* (MDRC Working Paper). New York, NY: MDRC.

Kashyap, U., & Mathew, S. (2017). Corequisite model: An effective strategy for remediation in freshmen level quantitative reasoning course. *Journal of STEM Education: Innovations & Research*, *18*(2), 23–29.

Logue, A. W., Watanabe-Rose, M., & Douglas, D. (2016). Should students assessed as needing remedial mathematics take college-level quantitative courses instead? A randomized controlled trial. *Educational Evaluation and Policy Analysis*, *38*(3), 578–598.

Martorell, P., & McFarlin, I, Jr. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, *93*(2), 436–454.

Moss, B. G., Yeaton, W. H., & LIoyd, J. E. (2014). Evaluating the effectiveness of developmental mathematics by embedding a randomized experiment within a regression discontinuity design. *Educational Evaluation and Policy Analysis*, *36*(2), 170–185.

Ngo, F., & Kwon, W. W. (2015). Using multiple measures to make math placement decisions: Implications for access and success in community colleges. *Research in Higher Education*, *56*(5), 442–470.

O'Hear, M., & MacDonald, R. (1995). A critical review of research in developmental education, part I. *Journal of Developmental Education* 19(1995), 2–6.

Ran, F. X., & Lin, Y. (2019). *The effects of corequisite remediation: Evidence from a statewide reform in Tennessee* (CCRC Working Paper No. 115). New York, NY: Community College Research Center, Columbia University, Teachers College.

Royer, D. W., & Baker, R. D. (2018). Student success in developmental math education: Connecting the content at Ivy Tech Community College. *New Directions for Community Colleges*, *2018*(182), 31–38.

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis*, *36*, 371–393.

Scott-Clayton, J., & Rodriguez, O. (2015). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy*, *10*, 4–45.

Xu, D., & Dadgar, M. (2018). How effective are community college remedial math courses for students with the lowest math skills? *Community College Review*, *46*, 62–81.