What Happens to Students Placed Into Developmental Education? A Meta-Analysis of Regression Discontinuity Studies

Jeffrey C. Valentine
University of Louisville

Spyros Konstantopoulos
Michigan State University

Sara Goldrick-Rab
Temple University

This article reports a systematic review and meta-analysis of studies that use regression discontinuity to examine the effects of placement into developmental education. Results suggest that placement into developmental education is associated with effects that are negative, statistically significant, and substantively large for three outcomes: (a) the probability of passing the college-level course in which remediation was needed, (b) college credits earned, and (c) attainment. Several sensitivity analyses suggest these results are not a function of particular stylized studies or the choices made in assembling the meta-analytic database. Two exploratory moderator analyses suggest that the negative effects of placement into developmental education are stronger for university students than for community college students and worse for students placed in reading or writing than in math. This work can inform debate and research on postsecondary policies and on alternative mechanisms for ensuring that college students have the skills needed to meet their goals.

Almost two in five beginning college students are placed in developmental education (Snyder, de Brey, & Dillow, 2016). Broadly speaking, the term "developmental education" connotes a set of policies and practices designed for students who are underprepared to do college-level work in a given area. The goal of this experience is to give students the knowledge, skills, and habits that will help them

be successful in the college-level version of the course (Bailey et al., 2016). The growing use of developmental education reflects an increasingly normative transition from high school to college, which while predicated on completion of secondary schooling, does not necessarily imply adequate preparation for what is deemed "postsecondary" work.

The specific mechanisms for deciding which students should be placed into developmental education vary. These policies are sometimes set at the state level (as is true in Florida), sometimes set at the system level (as in the California State University system), and sometimes set by individual institutions. Community colleges and other open access institutions generally require all students to take placement exams. Institutions that require an entrance exam like the Scholastic Aptitude Test or American College Testing (ACT) often use a tiered system. For example, in Tennessee (Boatman & Long, 2010), students scoring below 26 on the ACT's math subtest (approximately two thirds of all test takers score below this threshold) are required to take the COMPASS Algebra test, a placement exam developed by ACT. Students scoring 50 and above are placed into college algebra, while students scoring below 50 are placed into intermediate algebra. Depending on the specific policy in place, students may or may not be able to retake the placement exam.

Nationally, about 60% of students taking a placement exam are recommended for placement into developmental education, but not all students recommended for placement actually end up in the courses (see Bailey, 2009). According to the National Center for Educational Statistics (Snyder et al., 2016), rates of developmental course taking are somewhat higher in community colleges (about 42%) than in public and private doctoral degree–granting institutions (about 25% and 22%, respectively), but even in these latter institution types, developmental course taking is common. Math is the most common subject in which remediation is needed, with participation rates (about 15%) that are two to two and a half times the participation rates in English, reading, and writing (which range from 6% to 7%).

Placement into developmental education adds costs and, critically, time to a student's journey to a degree or certificate. With respect to student costs, Barry and Dannenberg (2016) estimate that each developmental course costs students $3,000 and adds $1,000 in student loan debt (and this analysis did not include the opportunity costs that students experienced). In addition, states are increasingly concerned about "paying twice" for courses taken both in high school and in college. Nationwide, Breneman and Haarlow (1998) estimate cost of development education to be $1 billion at public postsecondary institutions in 1996 dollars, while Pretlow and Wathington (2012), using similar methodologies, arrived at about $1.13 billion (again in 1996 dollars) for the 2004–2005 year. More recently, Barry and Dannenberg (2016) put the estimate at $1.5 billion (in 2011 dollars, or about $1.05 billion in 1996 dollars).

The research reported in this article examines the predictors of success at the postsecondary level (e.g., Credé, Roch, & Kieszczynka, 2010). Specifically, we are interested in understanding the effects of utilizing developmental education when it comes to college academic outcomes. Given the high personal and societal costs of developmental education, the effectiveness of developmental

education has become an important public policy question that has spurred both research and reform efforts (e.g., Complete College America, 2012). Most simple comparisons of students assigned to developmental education relative to those not assigned suggest that assignment to developmental education is associated with several negative outcomes, not least of which is a much lower likelihood of post-secondary attainment (i.e., graduation or certification). For example, using data from the National Educational Longitudinal Survey from the 1992 high school class, Attewell, Lavin, Domina, and Levey (2006) found that for students attending 2-year colleges, graduation rates were about 30% higher for students who did not enroll in at least one developmental education course than students who did (36% vs. 28%). For students attending 4-year institutions, the picture is even bleaker, with students enrolling in at least one developmental course graduating at a much lower rate (52%) than students not enrolling in developmental courses (77%). But it is far from clear whether these lower completion rates are *caused* by developmental education. There have been no formal, high-quality systematic reviews on the effects of student placement into developmental education. In part due to the difficulty of studying these effects, existing reviews have generated conflicting conclusions and have been contentious (e.g., see Bailey, Jaggars, & Scott-Clayton, 2013; Goudas & Boylan, 2012). This article offers a state-of-the-art systematic review and meta-analysis on the effects of placement into developmental education. We examine the effects of placement on four indicators of college attainment, including credit accumulation and degree or certificate completion.

Studying the Effects of Placement Into Developmental Education

Some of the observed differences in outcomes between students placed into developmental education in at least one subject and students not placed into developmental education are real in the sense that they reflect different levels of academic opportunities, preparation, and motivation. However, the raw statistics do little to untangle the causal effects of being placed into developmental education. There are two aspects to this problem. One is the distinction between enrollment and assignment. Attewell et al.'s (2006) data point to the negative association between enrollment in developmental education and attainment, but some students assigned to developmental education never take a developmental education course, either because they somehow avoid the placement decision and go directly into the college-level course, or because they take assignment to developmental education as a signal that they are unlikely to succeed in college and drop out (see Bailey, Jeong, & Cho, 2010; Scott-Clayton & Rodriguez, 2015); if true, this suggests that Attewell et al.'s (2006) analysis understates the negative impact of assignment to developmental education.

The second part of the problem is untangling the causal relationships (Goldrick-Rab, 2010). To test the effect of assignment to developmental education, researchers could identify a group of students for whom an institution's policy suggests developmental education is needed and randomly recommend students for placement into either the developmental course or into the college-level course in the subject in which remediation is needed. For example, Aiken, West, Schwalm, Carroll, and Hsiung (1998) randomly assigned students to either placement into

Freshman Composition or into a developmental writing course followed by Freshman Composition. However, their analyses were conditional on either passing the first assigned course (either developmental writing or Freshman Composition, depending on assignment) or on passing Freshman Composition, depending on the specific analysis[1] (see also Moss, Yeaton, & Lloyd, 2014). Much more common are nonrandomized experiments that adopt a similar approach of conditioning, in one way or another, on success in the developmental course. A randomized experiment that followed students regardless of whether they actually enrolled in the developmental course (or even enrolled in college) and assessed an outcome that is not dependent on course participation (e.g., whether or not students ultimately passed the college-level course in which remediation was needed) would provide a fair test of whether placement into developmental education is helpful to students. Given the scarcity of randomized trials in this area, it seems likely that institutions are reluctant to randomly assign students to developmental education or not. But because students are typically assigned to developmental education on the basis of a test score, regression discontinuity is a viable option for studying the effects of assignment to developmental education.

The Regression Discontinuity Design

The basic requirement of a regression discontinuity (RD) design study is that assignment to conditions is done using a score on a continuous variable. An example is the Tennessee process described by Boatman and Long (2010) above. Students are assigned to college algebra if they score 50 and above on the COMPASS Algebra placement test, and to developmental algebra if they score below 50. Thus, in an RD study, (a) groups are formed by design and (b) the assignment mechanism is completely known if (c) the cut score is adhered to (all of these features are shared with randomized experiments). The fact that the assignment mechanism is known allows for unbiased inferences if the assumptions of RD are met and if the data are analyzed properly.

Relative to randomized experiments, RD studies have lower statistical power (Schochet, 2009) and are dependent on more assumptions, some of which are untestable (Valentine & Thompson, 2013). Despite these drawbacks, RD is growing in popularity as researchers become more familiar with its strengths and the conditions under which it is particularly useful. Google Scholar (as of February 26, 2017) lists about 30 hits for "regression discontinuity" in 1995, about 250 in 2005, 1,320 in 2010, and 3,920 in 2016. Good primers on RD are available (e.g., Jacob, Zhu, Somers, & Bloom, 2012; Shadish, Cook, & Campbell, 2002; What Works Clearinghouse [WWC], 2015), but the basic logic underlying RD is easy to visualize. When the placement test score is presented on the x-axis of a graph and the dependent variable on the y-axis, a treatment effect can be seen as a visual break (or discontinuity) *at the cut point* between the scores of students who do and do not receive the intervention. Therefore, RD is similar to an interrupted time series approach, except that assignment is based on a score instead of time.

All studies should be evaluated for the rigor with which they were designed and analyzed (Valentine & Cooper, 2008), and this statement is especially true for RD studies. While still a developing field of study, the WWC's (2015) RD

standards provide a good example of how a quality assessment of an RD study might be carried out. The WWC articulates five quality markers for RD studies. These are (a) the variable used to create groups cannot be manipulated, (b) data loss due to attrition should be minimal, (c) there must be no evidence of a discontinuity anywhere other than at the cutoff variable, (d) the functional form of the relationship between the variable used to create groups and the outcome is properly specified (i.e., if the relationship is quadratic it should be modeled as such), and (e) the analyses are constrained to a proper "bandwidth" around the cutoff variable.

Systematic Review and Meta-Analysis

Systematic reviewing and meta-analysis are now the standard set of tools that researchers use to investigate the effectiveness of policies, procedures, and practices when multiple studies pertaining to the specific research question exist. As described below, we found 11 studies using RD that examine the effects of placement into developmental education. Though meta-analysis of RD studies is rare, it is not unknown (Deke, Dragoset, Bogan, & Gill, 2012; Quinn, Lynch, & Kim, 2014), and we anticipate that it will become more common in the future. In the Method section, we discuss some of the important considerations needed to support a meta-analysis of RD studies, using the set of RD studies we located as examples. We begin with a description of how we located, assessed for inclusion, and coded the studies in our analyses. After discussing meta-analysis in the context of RD studies, we present our findings on the effects of placement into developmental education on four outcomes: (a) college-level credits earned, (b) whether or not students eventually passed the college-level course in which remediation was needed, (c) student grades in the course in which remediation was needed, and (d) whether students earned a degree or certification. As will be seen, the data mostly suggest statistically significant and potentially important negative impacts on these outcomes. We conclude with suggestions about how placement into developmental education might be improved, and a discussion of the cautions and limitations that go along with our work.

Method

Literature Search

This review is part of a larger project examining interventions for developmental education students. Included studies used RD to examine the effects of placement into developmental education. We did not set inclusion or exclusion criteria around other parameters (e.g., outcomes measured) and searched for both published and unpublished studies. The electronic literature search was initially conducted in ProQuest (ProQuest Education journals and ProQuest dissertations), EBSCO (Education Research Complete and Academic Search Premier), ERIC, Wilson Education Full Text, the Social Science Citation Index, and PsycInfo, from 1993 to March 2013. Search terms were divided into three groups: (a) terms that identified the document as a study involving developmental education (developmental OR noncredit OR basic skills OR compensatory OR under achievement OR underachiev* OR remedia*); (b) terms that identified the context as postsecondary

education (e.g., universit* OR "institution of higher learning" OR "community college" OR "technical college" OR "junior college" OR "institutions of higher learning" OR "community colleges" OR "technical colleges" OR "junior colleges" OR "liberal arts" OR "Historically Black colleges and universities" OR "Hispanic Serving Institutions"); and (c) a term that identified the document as a study that used RD (discontin*). Documents with at least one search term from each these categories were screened for relevance by at least two trained individuals who worked independently. Disagreements were resolved by a third screener. We included only studies that examined the effects of placement into developmental education relative to placement directly into the college-level course (and not, e.g., studies that examined the effects of placement into different levels of developmental education; see Melguizo, Bos, Ngo, Mills, & Prather, 2016).

We also conducted ancillary searches to find studies of the effects of placement in developmental education. First, because the *Journal of Higher Education* does not publish abstracts, we hand searched that journal from 1993 forward. In addition, we conducted Google Scholar searches for relevant studies, and forward citation searches on the researchers who authored relevant papers. The last literature searches were run in November 2015.

Coding

Two reviewers working independently coded studies identified as potentially relevant. We coded characteristics related to study context, the developmental education placement process, the sample, and the study's outcomes. These characteristics included institution type (community college, university), the number of institutions in the study, whether the study was published, the process used to place students into developmental education (e.g., the specific placement test used and the cutoff for placement), and information about the students in the sample (e.g., whether the study included only first-time, full-time students).

Analytic Model and Analysis Issues

We used standard meta-analytic techniques to synthesize the results of eligible studies. These techniques included inverse variance weighting, which allocates proportionally more weight to larger studies. Many studies presented the results of multiple models (e.g., models with more or fewer covariates). Rather than adopting a robust variance estimation approach, we chose models in a deliberate attempt to maximize the conceptual similarity of the studies in the analysis. Therefore, when we had a choice, we always selected the model with (a) the largest number of control variables in it, (b) the narrowest bandwidth, and (c) results that were as close to 3 years from the time of assignment as possible (except for attainment, for which we selected the longest follow-up point).

Researchers undertaking a meta-analysis need to consider whether to employ a fixed-effect or a random-effects analytic model. Using the fixed-effect model, study effect sizes can be thought of as estimating a single population value, and therefore any differences in effect sizes across studies are treated as solely due to random sampling and identifiable covariates. Using the random-effects model, reviewers assume that studies do not in fact share a single population value but instead come from a distribution of effect sizes. Therefore, any differences in

effect sizes across studies are due to random sampling error, any identifiable covariates, and other random factors that cannot be identified.

The choice between fixed-effect and random-effects models can be an important one, because the confidence intervals arising from a random-effects analysis will never be smaller and are often larger than their fixed-effect counterparts; this has implications for both the statistical significance tests and interpreting the likely range of population effects. Often, the random-effects model is thought to be the most defensible choice, in part due to its somewhat better generalization properties (Hedges & Vevea, 1998). However, one issue with the random-effects model is that if the number of studies is small, the estimate of the between-studies variance component (i.e., the extent to which population effect sizes differ from one another) is both highly uncertain and highly unstable. That is, the between-studies variance component is estimated with a great deal of error, and it can be very sensitive to the inclusion of new information (e.g., a new study in an updated review). Due to these considerations, we report both the fixed-effect and the random-effects models in this review. In addition, we report several sensitivity analyses as robustness checks.

Finally, we should note that three studies examined the effects of placement into developmental education in multiple subjects (Boatman & Long, 2010; Calcagno & Long, 2008; Scott-Clayton & Rodriguez, 2015). Within each study, we treat these effects as independent. However, it is possible that some students could have been placed into developmental education in multiple subjects, and therefore be in our analyses more than once. For example, a student in Calcagno and Long's (2008) study could have been placed into developmental math and developmental reading and might have appeared in both of their bandwidth-constrained analyses; this would violate the statistical assumption of independence. We do not know the extent to which this combination of events happened. However, in Boatman and Long (2010), 17% of students were recommended for placement into developmental education in two subjects, and 5% were recommended for placement into three subjects. Therefore, in that study, the maximum overlap is 22%, but the overlap within the optimal bandwidth (i.e., students who scored between 47 and 52 on the math placement test and between 65 and 70 on the reading placement test) is likely much smaller (though probably not zero).

Meta-Analyzing RD Studies: RD Bandwidth

Randomized experiments can be thought of as estimating an average treatment effect. That is, in a simple randomized experiment the comparison of interest is the mean of one group relative to a mean of another group. RD studies are often thought of as estimating a *local* average treatment effect, with "local" defined as the group of participants who are relatively close to the cutoff. In the context of studying the effect of placement into developmental education, RD can be thought of as comparing students at the margin of college readiness, some of whom were assigned to developmental education and some of whom were assigned to college-level courses. Statistical procedures can be used to determine the optimal bandwidth within which treatment effects should be estimated (Imbens & Kalyanaraman, 2012). Some RD researchers use the entire sample instead of a bandwidth sample. This is probably reasonable only under extreme circumstances (e.g., the treatment has a constant effect on participants regardless of how far they

are from the cutoff). Researchers interested in using RD studies in meta-analysis should code for the presence or absence of bandwidths in the studies in their meta-analytic data set and if so, whether a statistical procedure to determine the optimal bandwidth was used. Furthermore, if possible researchers should statistically test whether effect sizes vary as a function of bandwidth (e.g., by conducting within study comparisons of the effects observed for wide and narrow bandwidths).

Meta-Analyzing RD Studies: Cutpoints

As noted, all RD studies use one or more cutpoints to assign students to conditions. This means that, in many real-world applications that feature some degree of local control over the assignment process, a somewhat different sample is being used across the studies in the meta-analysis. This problem is analogous to a meta-analysis of randomized trials in which some of the trials are targeted at very low-achieving students and others target students who are average achieving. Assuming that treatment effects are not homogeneous, this between-study variability associated with the cutpoints should be taken into account, and one way to do that is by incorporating the between sample variability via a random-effects approach, which we have done here. In some applications of meta-analysis with RD designs, it may be possible to control for the cutpoints used in each study (though this that was not possible here as many different placement tests were used, and we do not have enough information about either the tests or the samples to equate the test scores across studies).

Meta-Analyzing RD Studies: Adherence and Selective Sorting

In RD studies, some individuals may not comply with their condition assignments, especially if one condition is generally seen as more desirable than the other. This issue is similar to crossovers in a randomized experiment. And as in randomized experiment, crossovers could result in a bias in the estimate of the effect of being placed into developmental education. In examining the effects of placement into developmental education, college student counselors can sometimes override course placement recommendations, but probably the biggest threat to adherence is the student sorting associated with retesting. That is, institutions vary in the extent to which they allow students to retake the placement test. If students are allowed to retake the test, this has the potential to create a selection issue and carries with it the potential for bias. The main issue is that if retesting is allowed, students scoring in the developmental range are much more likely to take the test again than are students who score in the college range (for whom the probability of retaking the test is essentially zero). Thus, retesting means that some students originally assigned to the developmental group will end up in the college-level group. In Tables 1 to 4, to the extent that we are able we document retesting policies, and in most cases, retesting does not seem to be a problem. But, some authors were silent on this point. Calcagno and Long (2008) provided separate results for a subgroup of institutions that appeared to either not allow or to severely limit retesting, and we used the results from the "no retesting" group in our analyses below.

Meta-Analyzing RD Studies: RD Is a Model-Based Approach

RD researchers very often use additional control variables to help increase the precision of the estimates. One side effect of this practice is that researchers can end up using fairly different estimation models. The magnitude of a regression

coefficient arising from a model depends on the other variables in the model. For example, the regression coefficient describing the relationship between self-efficacy for algebra performance and actual algebra performance will change if a potential confounding variable, like math anxiety, is entered into the model (because math anxiety is correlated with self-efficacy for algebra). As can be in Tables 1 to 4, the models in our analyses varied somewhat. Most controlled for a robust set of student background variables including prior academic achievement and socioeconomic status, but two studies (Lesik, 2006; Moss & Yeaton, 2006) only included the placement test score in their estimation models (which is a requirement of an RD analysis). As a result, even if all of the studies were estimating the same population parameter, we would expect that their individual estimates would vary. Therefore, the fact that we observed somewhat different models across studies likely contributed to additional between-study heterogeneity.

One could ask whether it is even sensible to combine meta-analytically effects that arise from different model specifications. In this case, we believe that models are generally similar in that they share the most important covariate (the placement score) and tend to employ other covariates in similar domains. For example, in the analysis of the effects of placement into developmental education on college credits earned, there were 16 independent samples. All of the courses use the placement score. Gender, race, and SES were also common; hence, the core covariates tended to be similar. As a result, we believe that our models were sufficiently similar to support meta-analysis, though this will be an important consideration for future researchers thinking about conducting meta-analysis with RD studies.

Results

The literature search uncovered 11 reports, with a total of 21 independent samples, that use RD to investigate the effects of placement into developmental education (henceforth, we refer to independent samples as "studies"). However, Harmon (2011) does not appear in our analyses, as that study did not examine one of our four primary outcomes.[2] The studies varied widely in size. The within-study sample sizes of the analyses we used ranged from 185 to 59,334, with a median sample size of about 1,000 students (in all, well over 100,000 students are represented in the meta-analytic database).

Credits Earned

Sixteen analyses examined the effect of placement into developmental education on college credits earned. As can be seen in Table 1, credits earned were typically examined about 3 years after assignment. The mean effect size under fixed-effect assumptions was −1.86 credits, $p < .001$. The homogeneity test was statistically significant, $Q(15) = 43.18$, $p < .001$ ($I^2 = 68\%$), indicating more variability in the observed effect sizes than would be expected if sampling error alone drove variation in effect sizes. The mean effect size under random-effects assumptions was −3.00 credits, $p = .002$. Below, we report two sensitivity analyses and two exploratory moderator analyses on this data set.

Ever Pass College-Level Course for Subject?

Table 2 houses the effect size estimates for the six analyses involving whether students ever passed the college-level course in which remediation was needed.

TABLE 1

Study characteristics and outcomes: credits earned

Author(s) (year)	Institution type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model	Effect size (*SE*)
Boatman (2012) (APSU)	4	ACT	1	Allowed	Math	928	2 Years	Gender; race/ethnicity; high school achievement	0.662 (7.098)
Boatman (2012) (CSCC)	2	ACT	1	Allowed	Math	489	2 Years	Gender; race/ethnicity; high school achievement	1.724 (2.305)
Boatman (2012) (JSCC)	2	ACT	1	Allowed	Math	624	2 Years	Gender; race/ethnicity; high school achievement	1.969 (0.242)
Boatman and Long (2010)	4	COMPASS	1	Rare	Math	263	3 Years	Gender; race/ethnicity; SES	−4.8965 (3.4042)
Boatman and Long (2010)	2	COMPASS	1	Rare	Math	227	3 Years	Gender; race/ethnicity; SES	−8.3323 (4.178)
Boatman and Long (2010)	4	COMPASS	1	Rare	Reading	559	3 Years	Gender; race/ethnicity; SES	−4.1206 (3.0274)
Boatman and Long (2010)	2	COMPASS	1	Rare	Reading	938	3 Years	Gender; race/ethnicity; SES	−9.5066 (1.8544)
Boatman and Long (2010)	4	COMPASS	1	Rare	Writing	336	3 Years	Gender; race/ethnicity; SES	−2.0349 (3.1653)
Boatman and Long (2010)	2	COMPASS	1	Rare	Writing	622	3 Years	Gender; race/ethnicity; SES	−7.3279 (2.4366)

(continued)

TABLE 1 (continued)

Author(s) (year)	Institution type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model	Effect size (SE)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Math	9,593	6 Years	Gender; race/ethnicity	−0.244 (3.641)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Reading	8,755	6 Years	Gender; race/ethnicity	−1.59 (2.124)
Hodara (2012)	2	CUNY	Unknown	Unclear	Writing/ESL	12,773	3 Years	Gender; high school achievement; race/ethnicity; SES	−1.146 (0.925)
Martorell and McFarlin (2011)	2	TASP	Unknown	Allowed (but used first attempt score)	Reading or math	59,344	6 Years	Race/ethnicity; SES	−3.96 (2.15)
Martorell and McFarlin (2011)	4	TASP	Unknown	Allowed (but used first attempt score)	Reading or math	33,910	6 Years	Race/ethnicity; SES	−7.59 (2.71)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Math	17,641	3 Years	Gender; race/ethnicity	0.007 (0.796)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Reading	1,374	3 Years	Gender; race/ethnicity	−3.183 (3.023)

Note. APSU = Austin Peay State University; CSCC = Cleveland State Community College; JSCC = Jackson State Community College; *SE* = standard error; ACT = American College Testing; ESL = English as a second or foreign language; SES = socioeconomic status. Effect sizes are unstandardized regression coefficients, so represent the observed effect in terms of credit hours earned. Because all sample sizes are not small, a *z* test for each effect size can be given by the effect size ÷ standard error. For Martorell and McFarlin (2011), the regression coefficient represents the total number of college-level credits attempted over a 6-year period. For institution type, 2 = community college, 4 = university. For placement test, CPT = Florida College Entry Level Placement Test; CUNY = City University of New York's placement test; TASP = Texas Academic Skills Program.

TABLE 2

Study characteristics and outcomes: ever pass college-level course

Author(s) (year)	Institu-tion type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model?	Effect size (*SE*)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Reading	8,755	6 Years	Gender; race/ethnicity	−0.036 (0.017)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Math	9,593	6 Years	Gender; race/ethnicity	−0.002 (0.064)
Hodara (2012)	2	CUNY	1	Unknown	Writing/ESL	14,733	3 Years	Gender; high school achievement; race/ethnicity; SES	−0.147 (0.017)
Lesik (2006)	4	Unknown	1	Allowed but none in sample did	Math	212	4 Years	Placement score only	0.307 (0.668)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Math	17,641	3 Years	Gender; race/ethnicity	−0.059 (0.015)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Reading	1,374	3 Years	Gender; race/ethnicity	−0.143 (0.055)

Note. SE = standard error; ESL = English as a second or foreign language; SES = socioeconomic status. Effect sizes are ordinary least squares regression estimates with a binary dependent variable, so represent the observed effect in terms of percentages passing the college-level course (e.g., −0.147 means that students assigned to developmental education passed the first college-level course in which remediation was needed at a rate that was 14.7 percentage points less than the rate at which students assigned directly to the college-level course passed it). Because all sample sizes are not small, a *z* test for each effect size can be given by the effect size ÷ standard error. For Lesik (2006), we assumed that the base rate of passing was 50%, which resulted in the most optimistic effect size possible. The 0.307 effect size represents a translation of the logged odds ratio reported in Table 4 of 1.43. For institution type, 2 = community college, 4 = university. For placement test, CPT = Florida College Entry Level Placement Test; CUNY = City University of New York's placement test.

For both fixed- and random-effects models, the mean effect size was a 7.9–percentage point reduction in the proportion of students eventually passing the college-level course in which remediation was needed (e.g., from 75% to 68%), $p <$ .001 for the fixed-effect model and $p = .004$ for the random-effects model. The homogeneity test was statistically significant, $Q(5) = 27.31$, $p < .001$ ($I^2 = 81\%$).

Achievement in College-Level Course if Taken and Completed

Table 3 contains effect size estimates for nine analyses addressing the academic performance in the college-level course in which remediation was needed, conditional on students taking and completing that course. Of our four main outcomes, this is the one that is most likely to be biased by treatment-induced attrition, though the direction of this bias is difficult to predict. For the fixed-effect analysis, the estimated effect size is 0.00 ($p = .98$). For the random-effects model, the estimated effect size is +0.01 grade points ($p = .94$). The homogeneity test was not statistically significant, $Q(8) = 15.16$, $p = .06$ ($I^2 = 46\%$).

We should note here that while Scott-Clayton and Rodriguez (2015) also measured academic achievement in the college-level course in which remediation was needed, they did so by dummy coding achievement as whether students earned a B in the college-level course. In and of itself, this does not create a problem for our analysis, but Scott-Clayton and Rodriguez (2015) coded as "0" any student who either (a) earned less than a B or (b) never took the college-level course. Because this analysis conflates two aspects of the educational experience that we think should be kept separate, we did not use the two effect sizes from this study in our meta-analysis. Both were negative and statistically significant.

Degree Attainment

Thirteen studies examined the effect of placement into developmental education on degree or certificate attainment (see Table 4). For both fixed- and random-effects models, the mean effect size was a 1.5-percentage point reduction in the proportion of students eventually earning a degree (e.g., from 30% to 28.5%; $p =$ .03 for both models). The homogeneity test was not statistically significant, $Q(12) = 13.39$, $p = .34$ ($I^2 = 7\%$).

The raw magnitude of this effect depends on (a) the size of the incoming class and (b) the proportion of these students assigned to developmental education. At the institution level, in small institutions and in institutions with low developmental education placement rates, the negative effect of placement into developmental education will not matter much. But in larger institutions, and in institutions with higher placement rates, this effect might be large enough to matter. For example, imagine a typical mid-sized university with 6,000 incoming students, 25% of whom are assigned to developmental education. This institution could be expected to award 22 or 23 fewer degrees in that class than it would have if placement into developmental education had no effect on attainment (i.e., if the graduation rate among nondevelopmental students is 60%, then 58.5% of the 1,500 developmental students are expected to earn a degree, and the difference between the two attainment rates is 22.5 degrees).

Of course, at the policy level, the consequences are staggering. Assume that in a given year, 2.5 million students start their college careers in either a university

TABLE 3

Study characteristics and outcomes: achievement in college-level course

Author(s) (year)	Institution type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model?	Effect size (SE)
Boatman and Long (2010)	4	COMPASS	1	Rare	Math	185	3 Years	Gender; race/ethnicity; SES	0.2169 (0.2025)
Boatman and Long (2010)	2	COMPASS	1	Rare	Math	227	3 Years	Gender; race/ethnicity; SES	0.1451 (0.2422)
Boatman and Long (2010)	4	COMPASS	1	Rare	Reading	460	3 Years	Gender; race/ethnicity; SES	−0.1058 (0.1382)
Boatman and Long (2010)	2	COMPASS	1	Rare	Reading	596	3 Years	Gender; race/ethnicity; SES	0.0473 (0.1171)
Boatman and Long (2010)	4	COMPASS	1	Rare	Writing	315	3 Years	Gender; race/ethnicity; SES	0.2971 (0.1464)
Boatman and Long (2010)	2	COMPASS	1	Rare	Writing	467	3 Years	Gender; race/ethnicity; SES	−0.1438 (0.1356)
Horn, McCoy, Campbell, and Brock (2009)	2	COMPASS	1	Unknown	Reading	328	Unknown	Gender; race/ethnicity	−0.552 (0.216)
Moss and Yeaton (2006)	2	ASSET	Unknown	Unknown	Reading	1,473	6 Years	Placement score only	−0.02 (0.09)
Moss et al. (2014)	2	COMPASS	1		Math		1 Semester	Placement score only	0.34 (0.33)

Note. SE = standard error; SES = socioeconomic status. Effect sizes are unstandardized regression coefficients, so represent the observed effect in terms of student grades (e.g., 0.22 means that students assigned to developmental education scored 0.22 grade points higher on average than students assigned directly to the college-level course). Because all sample sizes are not small, a *z* test for each effect size can be given by the effect size ÷ standard error. Horn et al. (2009) did not report the standard error for the regression coefficient but did report that the coefficient's *p* value was less than .05 (but presumably larger than .01); 0.216 is the standard error that yields *p* = .011. The model without covariates has a standard error of 0.245 so 0.216 does seem reasonable. For institution type, 2 = community college, 4 = university.

TABLE 4

Study characteristics and outcomes: attainment

Author(s) (year)	Institution type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model	Effect size (*SE*)
Boatman and Long (2010)	4	COMPASS	1	Rare	Math	263	6 years	Gender; race/ethnicity; SES	−0.1176 (0.1687)
Boatman and Long (2010)	2	COMPASS	1	Rare	Math	227	6 years	Gender; race/ethnicity; SES	−0.4397 (0.2080)
Boatman and Long (2010)	4	COMPASS	1	Rare	Reading	559	6 years	Gender; race/ethnicity; SES	−0.1424 (0.1506)
Boatman and Long (2010)	2	COMPASS	1	Rare	Reading	938	6 years	Gender; race/ethnicity; SES	−0.1526 (0.1157)
Boatman and Long (2010)	4	COMPASS	1	Rare	Writing	366	6 years	Gender; race/ethnicity; SES	0.1982 (0.1626)
Boatman and Long (2010)	2	COMPASS	1	Rare	Writing	652	6 years	Gender; race/ethnicity; SES	0.0035 (0.1488)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Math	9,593	6 years	Gender; race/ethnicity	−0.027 (0.015)
Calcagno and Long (2008)	2	CPT	Unknown	Used "no retake" sample	Reading	8,755	6 years	Gender; race/ethnicity	−0.031 (0.026)

(continued)

TABLE 4 (continued)

Author(s) (year)	Institution type	Placement test	Levels of developmental education	Retake policy	Subject	Analytic sample size	Timing (from placement semester)	Covariates included in the model	Effect size (SE)
Hodara (2012)	2	CUNY	Unknown	Unclear	Reading and writing	12,773	3 years	Gender; high school achievement; race/ethnicity; SES	−0.001 (0.016)
Martorell and McFarlin (2011)	2	TASP	Unknown	Allowed (but used first attempt score)	Reading or math	59,344	6 years	Race/ethnicity; SES	−0.023 (0.016)
Martorell and McFarlin (2011)	4	TASP	Unknown	Allowed (but used first attempt score)	Reading or math	33,910	6 years	Race/ethnicity; SES	−0.040 (0.028)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Math	17,641	3 years	Gender; race/ethnicity	−0.001 (0.010)
Scott-Clayton and Rodriguez (2015)	2	COMPASS	Unknown	Allowed but strict	Reading	1,374	3 years	Gender; race/ethnicity	−0.029 (0.039)

Note. SE = standard error; SES = socioeconomic status. Effect sizes are unstandardized regression coefficients, so represent the observed effect in terms of percentage of the sample earning a certificate or degree. Because all sample sizes are not small, a *z* test for each effect size can be given by the effect size ÷ standard error. For institution type, 2 = community college, 4 = university. For placement test, CPT = Florida College Entry Level Placement Test; CUNY = City University of New York's placement test; TASP = Texas Academic Skills Program.

or a community college setting, that one third of these students are placed into developmental education, and that the overall 6-year graduation rate is 34%. The 1.5–percentage point reduction can be thought of as suggesting that 35% of students not placed into developmental education and 33.5% of students placed into developmental education will graduate in 6 years. This works out to a loss of about 12,500 certificates or degrees for that year's cohort of students.

Exploratory Moderator Analyses

Our data set of studies examining the effects of placement into developmental education on credits earned is the only one large enough to support even tentative moderator analyses; we report two of these analyses below. The first examines the effects observed in community colleges relative to universities, and the second examines effects observed separately for reading, writing, and math. Even though we approached these hypothesis tests with specific predictions in mind, we believe that they are best conceptualized as exploratory analyses because, as Lipsey and Wilson (2001) observed, studies have personalities in the sense that their traits tend to cluster together. For a meta-analysis, this means that study characteristics tend to correlate with one another, confounding univariate analyses of the relationship between study characteristics and outcomes. As a result, moderator analyses in meta-analysis should generally be multivariate so that study characteristics can be examined net of other characteristics in the model. However, meta-regression (the meta-analytic analog to multiple regression) generally requires a large number of studies for both reasonable statistical power and stable estimates. The analyses below are univariate and as such warrant an extra level of caution when interpreting them.

Effects for 2- Versus 4-Year Institutions (Credit Accumulation Only)

In our meta-analytic data set, we have five estimates of the effects of placement into developmental education on college credit accumulation that are based on 4-year institutions and 11 estimates that are based on 2-year institutions. For universities, the fixed-effect and random-effects mean effect size is −4.64 credits, $p = .002$. The homogeneity test within these five estimates was not statistically significant, $Q(4) = 2.46$, $p = .65$. For community colleges, a somewhat different picture emerges. The mean effect size under fixed-effect assumptions was −1.56 credits, $p = .001$. The homogeneity test was statistically significant, $Q(10) = 36.84$, $p < .001$. The mean effect size under random-effects assumptions was −2.62 credits, $p = .03$.

Effects for Different Subjects

Our meta-analytic data set includes four analyses of developmental education for reading, three analyses of developmental education for writing, and seven analyses of developmental education for math. For math, the fixed-effect and random-effects mean effect size is −0.08 credits, $p = .90$. The homogeneity test within these seven studies was not statistically significant, $Q(6) = 7.38$, $p = .29$. For reading, the mean effect size under fixed-effect assumptions was −5.45 credits, $p < .001$. The homogeneity test was statistically significant, $Q(3) = 8.58$, $p = .04$. The mean effect size under random-effects assumptions was −4.87

credits, $p = .01$. For writing, the mean effect size under fixed-effect assumptions was −1.93 credits, $p = .02$. The homogeneity test was not statistically significant, $Q(2) = 5.63$, $p = .06$. The mean effect size under random-effects assumptions was −3.18 credits, $p = .11$.

Sensitivity Analyses

Because we have the most information on credits earned, we used this dataset to conduct several sensitivity analyses. First, we Winsorize the meta-analytic weights and next, we drop studies one at a time from the analysis. Both of these strategies are intended to ensure that our results are not being driven by a single study. Finally, five studies allow us to tentatively test the extent to which study results are sensitive to the bandwidth that was used.

Influence Analyses

As can be seen in Table 5, under fixed-effect assumptions, two studies (Hodara, 2012 and Scott-Clayton & Rodriguez's, 2015 math analysis) have relative weights of 25% and 33%, suggesting that these studies are large relative to the other studies in the data set. Perhaps more important, Boatman's (2012) community college reading analysis is very influential. By this, we mean that the analysis' weight (which is above the mean) and effect size (the absolute value of which is the largest in the database) combine to exert a large influence on the fixed-effect analysis of college credits earned.

In Table 6, we first report our primary analyses for comparison. Then, we present the results of the primary analysis with two outlying weights Winsorized. We defined an outlier using Tukey's (1977) rule (i.e., an outlier is an observation that is more than two standard deviations beyond the 75th percentile). As we suspected, Hodara (2012) and Scott-Clayton and Rodriguez's (2015) math analysis were identified as outliers. We then trimmed the weights iteratively (recoding the weights so that they were no longer outliers, then rechecking for outliers) until no outliers were identified. This process had the effect of inflating the standard errors for these two studies (from 0.925 to 1.462 for Hodara, and from 0.796 to 1.462 for Scott-Clayton & Rodriguez, 2015). As can be seen, the patterns of statistical significance were unchanged across the mean effect size under fixed-effect assumptions, the mean effect size under random-effects assumptions, and the homogeneity analysis. Winsorizing resulted in a much larger point estimate for the fixed-effect analysis and had virtually no effect on the random effects analysis. The estimate of between study heterogeneity dropped somewhat with Winsorized weights (I^2 values were 68% for the main specification vs. 59% for the Winsorized analysis).

Next, we addressed potentially influential studies by dropping one study at a time from the main analysis of the effects of placement into developmental education on college credits earned. Again, most of the changes are minor, but dropping Boatman's (2012) community college reading effect results in a large change to the fixed estimate (from −1.86 to −1.36 credits) and to the random-effects estimate (−3.04 to −2.25 credits). Dropping both Hodara (2012) and Scott-Clayton and Rodriguez's (2015) math analysis resulted in less dramatic increases to both fixed-effect and random-effects estimates. Across these "drop one study" analyses, the statistical conclusions did not change (i.e., the mean effect was negative

TABLE 5

Relative weights and relative influence (college credits earned analysis)

Study	Fixed effect		Random effect	
	Relative weight	Relative influence	Relative weight	Relative influence
Boatman and Long (2010); University, Math	1.8%	1.8%	1.1%	1.3%
Boatman and Long (2010); Community College, Math	1.2%	5.4%	0.8%	8.2%
Boatman and Long (2010); University, Reading	2.3%	1.3%	1.2%	0.5%
Boatman and Long (2010); Community College, Reading	6.1%	38.3%	1.8%	26.6%
Boatman and Long (2010); University, Writing	2.1%	0.0%	1.1%	0.4%
Boatman and Long (2010); Community College, Writing	3.5%	11.3%	1.5%	9.7%
Hodara (2012)	24.5%	1.8%	2.3%	3.1%
Martorell and McFarlin (2011); Community College	4.5%	2.1%	1.6%	0.5%
Martorell and McFarlin (2011); University	2.9%	10.1%	1.3%	10.0%
Scott-Clayton and Rodriguez (2015); Reading	3.6%	2.1%	1.5%	0.8%
Scott-Clayton and Rodriguez (2015); Math	33.1%	13.1%	2.3%	7.7%
Calcagno and Long (2008); Math	1.6%	0.4%	1.0%	2.6%
Calcagno and Long (2008); Reading	4.3%	0.0%	1.6%	1.1%
Boatman (2012); APSU	0.4%	0.3%	0.4%	1.7%
Boatman (2012); CSCC	3.9%	5.4%	1.5%	12.1%
Boatman (2012); JSCC	4.2%	6.6%	1.6%	13.7%

Note. For each study, Relative Weight is the percentage of the total weight contributed by the study, and Relative Influence is defined as the weight times the square of the distance from each study's effect size to the grand mean divided by the sum of the weights (i.e., $w_i(X_i - X_{GM})^2/\Sigma w_i$). APSU = Austin Peay State University, CSCC = Cleveland State Community College, JSCC = Jackson State Community College.

TABLE 6
Sensitivity analyses (credits earned)

Study	Fixed effect			Homogeneity statistics			Random effects		
	ES	SE	p	Q(df)	p	I²	ES	SE	p
Main specification	−1.86	0.46	<.001	43.18 (15)	<.001	68%	−3.00	0.97	.002
Winsorized weights	−2.72	0.59	<.001	36.82 (15)	.001	59%	−3.06	0.98	.002
Drop Boatman and Long (2010); University, Math	−1.80	0.47	<.001	42.37 (14)	<.001	70%	−2.91	1.01	.004
Drop Boatman and Long (2010); Community College, Math	−1.78	0.46	<.001	40.75 (14)	.002	69%	−2.79	0.99	.005
Drop Boatman and Long (2010); University, Reading	−1.80	0.47	.001	42.61 (14)	<.001	70%	−2.94	1.02	.004
Drop Boatman and Long (2010); Community College, Reading	−1.35	0.48	.005	25.03 (14)	.034	51%	−2.25	0.83	.007
Drop Boatman and Long (2010); University, Writing	−1.85	0.46	<.001	43.18 (14)	<.001	71%	−3.06	1.02	.003
Drop Boatman and Long (2010); Community College, Writing	−1.65	0.47	<.001	37.95 (14)	<.001	67%	−2.68	0.99	.007
Drop Hodara (2012)	−1.86	0.46	<.001	43.18 (14)	<.001	67%	−3.00	0.97	.002
Drop Martorell and McFarlin (2011); Community College	−1.75	0.47	<.001	42.18 (14)	<.001	70%	−2.94	1.04	.005
Drop Martorell and McFarlin (2011); University	−1.68	0.47	<.001	38.57 (14)	<.001	67%	−2.69	0.98	.006
Drop Scott-Clayton and Rodriguez (2015); Reading	−1.82	0.47	<.001	42.98 (14)	<.001	71%	−3.00	1.03	.004
Drop Scott-Clayton and Rodriguez (2015); Math	−2.79	0.57	<.001	34.95 (14)	.002	60%	−3.34	1.02	.001
Drop Calcagno and Long (2008); Math	−1.88	0.47	<.001	42.98 (14)	<.001	70%	−3.13	1.01	.002
Drop Calcagno and Long (2008); Reading	−1.87	0.47	<.001	43.16 (14)	<.001	70%	−3.12	1.04	.003
Drop Boatman (2012); APSU	−1.87	0.46	<.001	43.05 (14)	<.001	70%	−3.06	0.99	.002
Drop Boatman (2012); CSCC	−2.00	0.47	<.001	40.67 (14)	<.001	67%	−3.34	0.99	<.001
Drop Boatman (2012); JSCC	−2.02	0.47	<.001	40.14 (14)	<.001	66%	−3.36	0.98	<.001

Note. ES = effect size (expressed as college credits earned), *SE* = standard error, *p* = probability, *Q (df)* = homogeneity test statistic value and degrees of freedom. The statistic *I²* expresses the proportion of variability in effect sizes that is attributable to differences across studies (as opposed to sampling error). APSU = Austin Peay State University, CSCC = Cleveland State Community College, JSCC = Jackson State Community College.

and statistically significant under fixed- and random-effects assumptions, and the homogeneity test was statistically significant), and the substantive interpretations of the effects were highly similar.

RD Assumptions

Two studies (Calcagno & Long, 2008; Martorell & McFarlin, 2011) provide effect sizes both for all students in the analysis and for a specific bandwidth. Similarly, three studies in our meta-analytic database (Hodara, 2012; Moss et al., 2014; Scott-Clayton & Rodriguez, 2015) used at least two bandwidths as a sensitivity check. In Table 7, we present the effects observed in these five studies across a total of 13 analyses and also provide a statistical test for the difference in the effect sizes across each of these analyses. The statistical tests are *z* tests using procedures described in Borenstein, Hedges, Higgins, and Rothstein (2009) for computing the variance of two correlated variables. This procedure requires that researchers know or estimate the extent to which the standard errors are based on independent information. Though not realistic, we chose zero for this value because doing so yields the smallest possible standard error. This means that the statistical tests in Table 7 are more likely to result in a rejection of the null hypothesis of no difference, even when there is no actual difference between the estimates and as such represent a "worst case" scenario.

As can be seen, only 1 of the 13 tests resulted in a rejection of the null hypothesis (*p* = .048). Correcting for multiple comparisons using any common procedure (e.g., a Bonferroni correction or the Benjamini–Hochberg correction) yields nonsignificant results for all tests. Furthermore, there was no consistency in the direction of the differences, and the median *p* value across these 13 analysis is .41. As such, we cannot find evidence in these studies that the observed effect sizes were unduly influenced by our decision to use the most narrow bandwidth given in the studies.

Discussion

This article reviewed evidence on the effects of placement into developmental education as evaluated with RD designs, and as such represents the most rigorous review of the effects of placement into developmental education to date. If the causal inferences are correct and our effect sizes are reasonably accurately estimated, the meta-analyses of studies using RD to investigate the effects of placement into developmental education suggest that placement into developmental education results in statistically significant and substantively sizable negative impacts. Relative to their peers who are also on the margin of college readiness but who were placed into college-level courses, students placed into developmental education earned fewer college credits after about 3 years (our estimates ranged from about 2 to 3 credit hours, depending on model specification), were about 8 percentage points less likely to eventually pass the college-level course in which remediation was needed, and were about 1.5 percentage points less likely to earn a certificate or degree. We cannot reject the null hypothesis that marginal students placed into developmental education perform similarly (i.e., earn similar grades) in the college-level course in which remediation was needed relative to marginal students placed into the college-level course. The results for college credits earned were not sensitive to either outlier effect sizes (there were none) or outlier weights.

TABLE 7

Sensitivity analysis: narrow bandwidth

Study	Analysis	% Sample overlap	Global ES	Global SE	Narrow ES	Narrow SE	SE difference	Mean difference	z	p
Comparisons of the full sample to the narrow bandwidth										
Calcagno and Long (2008)	Complete college-level course	36%	−0.049	0.016	−0.036	0.017	0.012	−0.013	−1.11	0.265
Calcagno and Long (2008)	Attainment	36%	−0.031	0.01	−0.031	0.026	0.014	0	0.00	1.000
Calcagno and Long (2008)	College-level credits earned	36%	−2.225	0.749	−1.59	2.214	1.169	−0.635	−0.54	0.587
Martorell and McFarlin (2011)	College-level credits earned—2 year colleges		−6.068	1.89	−3.956	2.151	1.432	−2.112	−1.48	0.140
Martorell and McFarlin (2011)	College-level credits earned—4 year colleges		−4.475	1.998	−7.558	2.712	1.684	3.083	1.83	0.067
Martorell and McFarlin (2011)	Attainment 6 years—2 year colleges		−0.023	0.016	−0.029	0.017	0.012	0.006	0.51	0.607
Martorell and McFarlin (2011)	Attainment 6 years—4 year colleges		−0.023	0.02	−0.04	0.028	0.017	0.017	0.99	0.323
Comparisons of the widest bandwidth to the narrowest bandwidth										
Hodara (2012)	Ever passed college-level course	75%	−0.127	0.011	−0.147	0.017	0.010	0.02	1.98	0.048
Hodara (2012)	College-level credits earned	75%	−1.247	0.622	−1.146	0.925	0.557	−0.101	−0.18	0.856
Hodara (2012)	Attainment 5 years	75%	−0.009	0.011	−0.001	0.016	0.010	−0.008	−0.82	0.410
Moss et al. (2014)	Grade in college-level course		0.26	0.14	0.34	0.33	0.179	−0.08	−0.45	0.655
Scott-Clayton and Rodriguez (2015)	Ever passed college-level course	36%	−0.048	0.013	−0.059	0.015	0.010	0.011	1.11	0.268
Scott-Clayton and Rodriguez (2015)	College-level credits earned	36%	−0.029	0.683	0.007	0.796	0.524	−0.036	−0.07	0.945

Note. ES = effect size; *SE* = standard error.

Influential studies similarly did not affect the statistical significance of the results, though in the fixed-effect model, there was some variation in the effect sizes observed depending on which studies were in the analysis (effect sizes in the random effects model were very similar regardless of which studies were in the analysis). There is no evidence that the observed effect sizes were influenced by the decision to focus on the narrowest bandwidth presented in the studies in the review.

The exploratory moderator analyses using the studies that assessed college credits earned suggest that the negative effects of placement into developmental education are stronger for university students (but still statistically significant and negative for community college students), and for students placed into developmental education in reading and writing (recall that for writing, the fixed-effect estimate was statistically significant but the random-effects estimate was not, $p = .11$), but not math (the fixed- and random-effects estimates were close to zero and were not statistically significant). This latter point merits additional research attention. Using a sample of community college students enrolled in college-level English, Roksa, Jenkins, Jaggars, Zeidenberg, and Cho (2009) found that the probability of passing that course was unrelated to placement test scores. Though just one study, this finding raises questions about the adequacy of placement test scores as a basis for assigning students to developmental education.

How Can Educational Systems and Institutions Improve the Situation?

This study was designed to assess—across multiple studies in many contexts—if placement into developmental education helps students be successful in college. It was not designed to address how or why any positive or negative effects might have occurred. That said, because students were about 8 percentage points less likely to eventually pass the college-level course in which remediation was required, it is reasonable to speculate that the college-level course in which remediation was required represents an important roadblock for students assigned to developmental education. For example, Bailey (2009) concluded that developmental education is "not every effective . . . partly because the majority of students referred to developmental education do not finish the sequences to which they are referred" (p. 12; see also Bailey et al., 2010). Furthermore, much of the national conversation on developmental education has focused on misplacement rates. As mentioned earlier, placement is generally based on a single test. No one believes—or at least, no one should believe—that these tests are perfect indicators of college readiness (see Armstrong, 2000). A general principle of psychological measurement is when a construct (like college readiness) is measured imperfectly, one way to improve measurement is to measure the construct in multiple ways. Incorporating information that many institutions already have—such as high school grade point average and scores on standardized entrance tests—into placement decisions is a relatively easy way to modify the placement rubric that has the potential to reduce misplacement rates. Title 5 §55502 of the California Code of Regulations explicitly recognizes this by requiring institutions to use multiple measures for placement into developmental education, and even placement test developers recommend that institutions use multiple measures for placement (Westrick & Allen, 2014). If we were responsible for running an institution, attempting to reduce misplacement rates by using multiple measures would be where we would start reform efforts.

Furthermore, it is not clear that all students need a semester-long course to achieve college readiness, and researchers have been experimenting with other ways to accomplish this goal. For example, Logue, Watanabe-Rose, and Douglas (2016) conducted a randomized experiment in which algebra instruction was embedded into a college-level statistics course supplemented with weekly workshops that focused on algebra. Compared with students who took the usual developmental algebra course, or that course supplemented with weekly workshops, students taking the college statistics course earned more college-level credits over three semesters (21 to about 15 in the other two groups), and were more likely to pass the course to which they were assigned. Other possible ways of remediating deficits include summer bridge programs, targeted one credit presemester tune-up courses, and by providing additional supports (e.g., mandatory tutoring sessions during the semester). The important point is that educational leaders should think carefully about who gets placed into developmental education and develop flexible systems to help students develop the skills that they need to be successful in college (see Bailey, 2009).

Limitations and Conclusion

An important conceptual limitation is that this study did not address the effect of placement at different levels of developmental education (e.g., elementary vs. intermediate algebra). Due to the relatively small numbers of students placed at the lowest levels of developmental education, and the fact that all else being equal statistical power in RD is much lower than in a randomized experiment, it is likely that a series of randomized experiments will be needed to address this question.

With respect to the questions that we were able to address, perhaps the greatest threat to the conclusions we draw in this article is that our analyses are based on studies with characteristics that differ in fundamental and probably important ways. Most obviously, we included studies that examined the effects of placement into developmental reading, writing, and math, and studies that occurred in both community colleges and universities. We were only able to test these two potential modifiers of the effects of placement in developmental education for one outcome (credits earned) because we had too few studies to support parallel analyses for the other outcomes. Those analyses did suggest that there is reason to suspect heterogeneous effects (e.g., placement into developmental education appears to have more negative effects on university students than on community college students). However, these analyses were not multivariate, and therefore could confound the effects of other study characteristics with the ones we were examining. Readers can draw some reassurance from our extensive sensitivity analyses, which suggest that our results are not unduly influenced by exceptional studies or by some of the important decisions we made when assembling our meta-analytic data set.

Even exercising appropriate caution in drawing causal conclusions from our research, based on the studies we review, it is very difficult to walk away with the conclusion that placement into developmental education *helps* students. More than 75% of the estimates in our meta-analytic database are negative, and the meta-analytic estimates for the probability of passing the college-level course in which remediation was needed, college credits earned, and attainment are all negative, statistically significant, and large enough to be meaningful. Our hope is that this work spurs thoughtful debate and research on placement policies and on

alternative mechanisms for ensuring that college students have the skills needed to meet their goals.

Notes

The work reported herein was supported in part by the U.S. Department of Education's Institute of Education Sciences (ED-IES-12-C-0084). However, the views expressed are the authors' and do not necessarily represent the positions or policies of the Institute of Education Sciences or the U.S. Department of Education. Attendees at the *Society for Research Synthesis Methods* annual meeting (July 2015) provided helpful feedback on portions of this article.

[1]Of the approximately 100 students who passed their assigned course in the fall semester, students in the college class scored statistically significantly higher on a writing sample than did students placed into developmental education. There was not a statistically significant difference between groups on the Test of Standard Written English. Among approximately 90 students who passed Freshman Composition, this pattern was reversed: The remediation group scored statistically significantly higher on the Test of Standard Written English, while there was no statistically significant difference between the two groups on the writing sample.

[2]This study focused on the overall grade point average of students who successfully exited developmental education relative to students not assigned to developmental education.

References

References marked with an asterisk indicate studies included in the research synthesis.

Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J. L., & Hsiung, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation efficacy of a university-level remedial writing program. *Evaluation Review*, *22*, 207–244. doi:10.1177/0193841x9802200203

Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research and Practice*, *28*, 681–695. doi:10.1080/10668920 050140837

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*, 886–924. doi:10.1353/jhe.2006.0037

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions for Community Colleges*, *145*, 11–30. doi:10.1002/cc.352

Bailey, T., Bashford, J., Boatman, A., Squires, J., Weiss, M., Doyle, W., . . .Young, S. H. (2016). *Strategies for postsecondary students in developmental education: A practice guide for college and university administrators, advisors, and faculty*. Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

Bailey, T., Jaggars, S. S., & Scott-Clayton, J. (2013). *Characterizing the effects of developmental education: A response to recent criticism*. Retrieved from http://ccrc.tc.columbia.edu/publications/characterizing-effectiveness-of-developmental-education.html

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*, 255–270. doi:10.1016/j.econedurev.2009.09.002

Barry, M. N., & Dannenberg, M. (2016). *Out of pocket: The high costs of inadequate high schools and high school student achievement on college affordability*.

Washington, DC: Education Reform Now. Retrieved from https://edreformnow.org/app/uploads/2016/04/EdReformNow-O-O-P-Embargoed-Final.pdf

*Boatman, A. (2012). *Evaluating institutional efforts to streamline postsecondary remediation: The causal effects of the Tennessee developmental course redesign initiative on early student academic success*. New York, NY: National Center for Postsecondary Research.

*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*. New York, NY: National Center for Postsecondary Research.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley. doi:10.1002/9780470743386

Breneman, D. W., & Haarlow, W. N. (1998). *Remediation in higher education: A symposium featuring developmental education: Costs and consequences. Fordham Report*, *2*. Washington, DC: Thomas B. Fordham Foundation.

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

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college a meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*, 272–295. doi:10.3102/0034654310362998

Complete College America. (2012). *Remediation: Higher education's bridge to nowhere*. Retrieved from http://www.completecollege.org/docs/CCA-Remediation-final.pdf

Deke, J., Dragoset, L., Bogan, K., & Gill, B. (2012). *Impacts of Title I supplemental educational services on student achievement* (NCEE 2012-4053). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldrick-Rab, S. (2010). Challenges and opportunities for improving community college student success. *Review of Educational Research*, *80*, 437–469. doi:10.3102/0034654310370163

Goudas, A. M., & Boylan, H. R. (2012). Addressing flawed research in developmental education. *Journal of Developmental Education*, *36*, 2–13. doi:10.4018/978-1-4666-2621-8.ch021

Harmon, T. B. (2011). *Remedial policy in the California State University system: An analysis*. Available from Proquest Digital Dissertations (UMI Number 3468568)

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504. doi:10.1037//1082-989x.3.4.486

*Hodara, M. (2012). *Language minority students at community college: How do developmental education and English as a second language affect their educational outcomes*? Available from Proquest Digital Dissertations (UMI Number 3505981)

*Horn, C., McCoy, Z., Campbell, L., & Brock, C. (2009). Remedial testing and placement in community colleges. *Community College Journal of Research and Practice*, *33*, 510–526. doi:10.1080/10668920802662412

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, *79*, 933–959. doi:10.1093/restud/rdr043

Jacob, R. T., Zhu, P., Somers, M. A., & Bloom, H. S. (2012). *A practical guide to regression discontinuity*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.671.4723&rep=rep1&type=pdf

*Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *Review of Higher Education*, *30*, 1–19. doi:10.1353/rhe.2006.0055

Lipsey, M. W., & Wilson, D. B. (2001). The way in which intervention studies have "personality" and why it is important to meta-analysis. *Evaluation & the Health Professions*, *24*, 236–254. doi:10.1177/01632780122034902

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*, 578–598. doi:10.3102/0162373716649056

*Martorell, P., & McFarlin, Jr., I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *Review of Economics and Statistics*, *93*, 436–454. doi:10.1162/rest_a_00098

Melguizo, T., Bos, J. M., Ngo, F., Mills, N., & Prather, G. (2016). Using a regression discontinuity design to estimate the impact of placement decisions in developmental math. *Research in Higher Education*, *57*, 123–151. doi:10.1007/s11162-015-9382-y

*Moss, B. G., & Yeaton, W. H. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, *28*, 215–229. doi:10.3102/01623737028003215

*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*, 170–185. doi:10.3102/0162373713504988

Pretlow, J., III, & Wathington, H. D. (2012). Cost of developmental education: An update of Breneman and Haarlow. *Journal of Developmental Education*, *36*, 4–14.

Quinn, D. M., Lynch, K., & Kim, J. S. (2014, March). *Replicating the moderating role of income status on summer school effects across subject areas: A meta-analysis*. Paper presented at the Spring Conference of the Society for Research on Educational Effectiveness, Washington, DC.

Roksa, J., Jenkins, D., Jaggars, S. S., Zeidenberg, M., & Cho, S. (2009). *Strategies for promoting gatekeeper success among students needing remediation: Research report for the Virginia Community College System*. Retrieved from http://67.205.94.182/publications/gatekeeper-course-success-virginia.html

Schochet, P. Z. (2009). Statistical power for regression discontinuity designs in education evaluations. *Journal of Educational and Behavioral Statistics*, *34*, 238–266. doi:10.3102/1076998609332748

*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. doi:10.1162/edfp_a_00150

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of Education Statistics 2014* (NCES 2016-006). Retrieved from https://nces.ed.gov/pubs2016/2016006.pdf

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods*, *13*, 130–149. doi:10.1037/1082-989x.13.2.130

Valentine, J. C., & Thompson, S. G. (2013). Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods*, *4*, 26–35. doi:10.1002/jrsm.1064

Westrick, P. A., & Allen, J. (2014). *Validity evidence for ACT COMPASS tests*. Retrieved from http://files.eric.ed.gov/fulltext/ED546849.pdf

What Works Clearinghouse. (2015). *Preview of regression discontinuity design standards*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_rdd_standards_122315.pdf

Authors

JEFFREY C. VALENTINE is a professor in and the coordinator of the Educational Psychology, Measurement, and Evaluation program in the College of Education and Human Development at the University of Louisville. A social psychologist by training (University of Missouri-Columbia, 2001), most of his work involves using, explaining, and seeking to improve meta-analytic techniques as a means of helping policymakers and practitioners identify effective interventions that improve the health, well-being, and educational outcomes of children, young adults, and families. He is co-editor, with Harris Cooper and Larry Hedges, of the second edition of the *Handbook of Research Synthesis and Meta-Analysis*; associate editor of the peer-reviewed journal *Research Synthesis Methods*; and statistical editor for the Cochrane Collaboration's *Psychological, Developmental, and Learning Problems Group*.

SPYROS KONSTANTOPOULOS is professor of measurement and quantitative methods at the department of counseling, educational psychology, and special education at the college of education at Michigan State University. He received his BA from the University of Athens in Education, his first MS from Purdue University in Educational Psychology and Research Methods, his second MS from the University of Chicago in Statistics, and his PhD from the University of Chicago in Research Methods. His research interests include the extension and application of statistical methods to issues in education, social science, and policy studies. His methodological work involves statistical methods for multilevel data structures. His substantive work encompasses research on class size effects, teacher and school effects, and the social distribution of academic achievement.

SARA GOLDRICK-RAB is professor of higher education policy & sociology at Temple University, and founder at the Wisconsin HOPE Lab, the nation's only translational research laboratory seeking ways to make college more affordable. Dr. Goldrick-Rab's commitment to scholar-activism is evidenced by her broad profile of research and writing dissecting the intended and unintended consequences of the college-for-all movement in the United States. In more than a dozen experimental, longitudinal, and mixed-methods studies, she has examined the efficacy and distributional implications of financial aid policies, welfare reform, transfer practices, and a range of interventions aimed at increasing college attainment among marginalized populations. She provides extensive service to local, state, and national communities, working directly with governors and state legislators to craft policies to make college more affordable, collaborating with non-profit organizations seeking to examine the effects of their practices, and providing technical assistance to Congressional staff, think tanks, and membership organizations throughout Washington, DC.