A Brief Response to Bailey, Jaggars, and Scott-Clayton

By Alexandros M. Goudas and Hunter R. Boylan

Abstract: ***Shortly after we published "Addressing Flawed Research in Developmental Education" (2012) in the* Journal of Developmental Education*, Thomas Bailey, Shanna Smith Jaggars, and Judith Scott-Clayton from the Community College Research Center (CCRC) wrote a response rebutting several of our claims. Though their response corrects some confusion and clarifies a few of their positions on the debate, Bailey et al. appear to persist in a lack of understanding of the content and function of developmental education courses. Compounding the problem is that they solely rely on a relatively new and imperfect method for analysis, the quasi-experimental regression discontinuity design study. Based on these studies, Bailey et al. have consistently argued that developmental education as a whole is ineffective. In this brief response to Bailey et al.'s counterarguments, we elaborate on one of our original paper's main points and discuss what we consider to be a fundamental flaw in their interpretation of data. The flaw apparently stems from a misunderstanding of what actually happens in remedial courses. As a result, they assume these courses should make remedial students perform better than statistically equivalent nonremedial students. We moreover point out other possible errors in the regression discontinuity approach and its application in developmental education.***

Regrettably, organizations have cited the CCRC and others' research to promote radical changes to developmental education.

Alexandros M. Goudas
Developmental Education Director
Assistant Professor of English
alexandrosgoudas@delta.edu

Delta College
1961 Delta Road
University Center, MI 48710,

Dr. Hunter R. Boylan
Director
boylanhr@appstate.edu

National Center for Developmental Education
Appalachian State University
ASU Box 32098
331 Duncan Hall
Boone, NC 28608

The intent of this paper is to respond to refutations set forth in Bailey, Jaggars, and Scott-Clayton's paper, "Characterizing the Effectiveness of Developmental Education: A Response to Recent Criticism" (2013), which is a response (see also,COMMENTARY, pg. 18) to our initial paper, "Addressing Flawed Research in Developmental Education" (2012). However, before we go any further, we would like to take the time to make an important distinction that may not have been explicitly clear in our initial work.

The Community College Research Center (CCRC) of Columbia University, established in 1996, is an organization whose mission is to "contribute to the development of practice and policy that expands access to higher education and promotes success for all students" (CCRC, 2013, para. 2). Even though we may disagree with a few of the CCRC's conclusions regarding developmental education as a whole, we do in fact agree with and support not only their mission but the overwhelming majority of their research as well. Their intent is genuine, their research is sound, and thus there should be no question of their integrity.

Regrettably, organizations have cited the CCRC and others' research to promote radical changes to developmental education, such as abolishing prerequisite remedial courses and replacing them with corequisite courses, much like a law passed recently by the Connecticut State Legislature (An Act Concerning, 2012). Fortunately the CCRC does not agree with this course of action. In their response, Bailey et al. (2013) clearly state, "We do not advocate—nor do we believe that the results of our research support—the elimination of developmental education, the placing of all students into college courses, or the wholesale conversion of developmental education into a corequisite model" (p. 2).

As researchers and academics, we believe in a healthy and critical exchange of ideas, and it is in this spirit that we wish to simultaneously support our colleagues at the CCRC yet add to this important dialogue. The following comments therefore attempt to define our disagreements more clearly so that readers may be more informed about our position. More importantly, with this paper and others, we hope students, parents, educators, legislators, and researchers can learn about and build upon the corpus of community college research so that all constituents can accomplish and maintain the shared goal of student success at the postsecondary level.

A Basic RDD Flaw: What the Treatment Should Accomplish

The CCRC's researchers have contributed to and continually reference a body of work they rely upon to argue that developmental education is ineffective overall for students in community colleges and four-year institutions. One problem with this assessment is that it defines developmental education exclusively in terms of remedial courses. Scholars in the field (Arendale, 2012; Boylan, 1999) generally define developmental education as the integration of courses and academic support services guided by the principles of adult learning and development. However, criticisms leveled against developmental

education may instead actually focus only on remedial courses, which may or may not be guided by the principles of adult learning and development. It is incorrect, therefore, to say that developmental education as a whole is ineffective.

Nevertheless, their sole evidence for claims of remediation's inefficacy are the results from quasi-experimental analyses of various longitudinal datasets. These studies are called regression discontinuity design (RDD) studies, and since approximately seven or eight RDD studies from varying levels of cutoffs show mostly null or negative results, then the CCRC argues that developmental education has no positive effect overall. We previously addressed their use of what we believe to be an incorrect definition of success. The CCRC's response dismissed our definition and reasserted their original position. However, we argue that once remedial courses are defined further, an important distinction becomes clear that should make scholars more carefully consider the suitability of the RD design as applied to remediation and conclusions based on this approach.

The only way an RDD approach can be correctly applied to remediation is if the "treatment," in this case, remedial courses, is designed to affect students so that after its application, students are in a "better condition" than the students who do not receive the treatment. That is, researchers would ideally have two statistically equivalent groups of students (those below and above the cutoff for college-level work), they would administer the treatment (remedial courses to those just below the cutoff), and they would analyze whether the treatment had any effect (if the remedial students benefited from the treatment, then they would perform "better" in later outcomes). Imagine a medicinal treatment provided to two randomly assigned groups, and this is the fundamental theoretical framework upon which RDD researchers base their studies.

Referring to two statistically equivalent groups of students immediately above and below a placement test's cutoff for college-level courses, Bailey et al. (2013) state this assumption quite clearly: "Because the two groups are identical prior to remedial assignment, if remediation has any beneficial effect, it would show up as a positive difference in outcomes" (p. 3). Thus RDD researchers and those who support its use look upon remediation as a treatment, much like a remedy one would take if one were ill. In fact, in a 2012 working paper for the National Bureau of Economic Research, Scott-Clayton, Crosta, and Belfield compare remediation to exactly that: "Like a costly medical treatment with non-trivial side effects, the value of remediation overall depends upon whether those most likely to benefit can be identified in advance" (Abstract).

The problem, however, is that remediation cannot quite be compared to a pill administered in the field of me3dicine. First, unlike medication, remediation's "treatment" is by no means universal or even very similar. Within a single institution there are wide variations in such areas as teaching methods and styles, student-teacher ratios, curricula, textbooks, exercises, lab work, homework, and so on that might confound RDD results significantly. Only very small groups of students among thousands studied are therefore receiving the exact same pills as a treatment. Shadish, Cook, and Campbell (2002) stress the importance of having a single treatment the experimenter can control when conducting a proper experiment. This in itself is enough to call into question the heavy reliance on RDD analyses for examining the impact of remediation.

There also appears to be some confusion as to what the "treatment" is actually designed to do. Hypothesizing that the treatment will improve student outcomes of those who place into remediation to a level that is higher than students slightly above the cutoff seems to assume that remedial courses are designed to develop student skills above the minimum entry-level cutoff for gatekeeper courses. Such an assumption might even include the underlying tenet that remediation is designed as a set of courses whose curriculum is identical to, albeit easier than, their respective gatekeeper courses. In other words, any given highest level remedial math course, for example, would contain all the concepts taught in the subsequent gatekeeper math course, but they would be taught in a simplified manner. That way, if a student took a remedial math course and reviewed all the gatekeeper material prior to taking the gatekeeper math course, then that student would be expected to perform better in the gatekeeper course than another student who did not take the remedial treatment. This makes sense if it were the case, and we can understand why researchers unfamiliar with the actual content of developmental courses might believe this. It is our opinion, however, that this belief is erroneous.

The fact is that remedial courses only cover a small portion of gatekeeper courses' curricula, and the overlapping material are concepts which are reviewed only at the beginning of or briefly throughout the gatekeeper course. Effective math curricula in particular are tiered so that one set of knowledge is dependent prerequisite information for the next set (Schmidt et al., 2005). And remedial math courses are no different along the continuum. Thus students taking the highest level remedial math course will simply become prepared for the start of a gatekeeper course.

To a slightly lesser degree, the same applies to English remediation. Students who take the highest level remedial English writing course should be prepared to write college-level compositions, but not much more so than students who do not remediate. This is because many highest level remedial English courses not only focus on basic composition forms, but they also stress grammar exercises and leave complicated writing procedures for subsequent college-level courses.

RDD researchers and those who support its use look upon remediation as a treatment, much like a remedy one would take if one were ill.

To put it another way, remedial course content is very similar to traditional secondary or high school courses. If college students were to go back to high school and retake courses to prepare for college gatekeepers instead of taking college remediation, no one would expect that they would perform better than nonremedial students on more difficult college-level concepts presented in gatekeeper courses. They would simply have a better grounding in the basic prerequisite concepts often repeated at the beginning of a gatekeeper course, much as we expect from traditional high school graduates' preparation. Their participation in high school courses or remedial college courses does not ensure a greater success rate at the middle or end of a gatekeeper course; this is because the new material presented at those times is new for everyone, former remedial students or not.

Before conducting a study on how a common postsecondary strategy such as remedial courses might affect students, it would make sense to investigate what exactly transpires in those courses. The precise treatment should be defined before an experimenter applies it and expects it to work. However, we have found no research which has delved into a detailed curricula of remedial courses and compared them with the curricula of their respective gatekeeper courses. Rather, there are many who incorrectly assume that remedial courses present a "light" version of the exact material gatekeeper courses offer, which is by and large not the case. Making this assumption is the only way one could predict that remediation should increase success rates in gatekeeper courses. When the RD approach repeatedly fails to show an increase in success rates for remedial students, researchers, based on this misunderstanding of what remedial courses' curricula are, therefore conclude that the courses are ineffective. This indeed is a simple misapplication of a theoretical framework (RDD) based on a treatment that is not thoroughly understood.

A Brief History and Critique of the RDD Approach in Education

The most important critique a scholar could make about the applicability of the RDD approach in developmental education is that the results can and should only be applied to the specific students directly under a particular placement test's cutoff used in that specific institution. Even though Bailey et al. (2013) extrapolate their findings to all remedial courses' efficacy because different RDD studies use varied cut-scores farther up and down placement tests' continuums, the accepted practice is that there is not much generality from single cutoff RDD studies (Cook, 2008; Cook & Wong, 2008). CCRC researchers and others who create RDD studies are quick to point out this fact, and we have already cited their reservations in applying their results to any students beyond the immediate cut-score range. In fact, referring to the null or negative effect of remediation on students in RDD studies, Bailey et al. (2013) support the fact that the results can only be applied to those at the cutoff: "This is the appropriate conclusion for *the students who were included in the comparisons*, all of whom scored within *the same narrow range on a placement test* [emphases added]" (pp. 4-5).

Beyond this, many scholars over the years since the RD design was invented (1958) have elaborated extensively on the problems and limitations of quasi-experimental studies (Calcagno & Long, 2008; Cook, 2008; Cook & Wong, 2008; Lee, 2008; McCrary, 2008; Shadish et al., 2002). Cook (2008), in a very thorough overview of RDD's history and its current applications, goes so far as to say that its former use in education had "ossified" by the mid-1990s, even though some of its original statistical problems had been overcome by that point. He states moreover that RDD is currently waiting for its day as a legitimately used analysis in education. In fact, Cook points out that the RDD approach was only "recently given special status at the Institute for Educational Sciences (IES) for when an experiment is not possible" (p. 643). Yet Cook also states, "RDD has not been widely adopted in Psychology or Education" (p. 642).

The fact that most results are null suggests that fully half of the students are performing above average.

It appears, then, that current scholars associated with the application of RDD to developmental education are a solitary group when they apply it routinely in that field. Scholars in statistics have yet to use this design regularly due to its limitations (Cook, 2008). Only in the field of economics is RDD becoming established, recognized, and growing in use (Cook, 2008; Lee & Lemieux, 2010). Although Cook argues that the RDD's validity is sound, he still cautions researchers:

> No advocates of RDD have seen it as superior to the randomized experiment or even equivalent to it in terms of warranting causal claims. RDD is less statistically powerful; it involves less transparent assumptions about functional form; its implementation is less well empirically understood; and methods for improving its implementation are less developed. (p. 652)

On the one hand, we should commend the researchers at CCRC and others for importing alternative methodologies to the study developmental education. On the other hand, this methodology still does not meet the "gold standard" for empirical research.

Aggregated Data: Why Not Drill Down?

It must be remembered that the RDD approach aggregates thousands of data points and creates an average to form the line which researchers ultimately assess for continuity or discontinuity. Implicit in this average is that there are wide variations in student outcomes, all of which are averaged to make it appear like one result. Researchers use averages and large numbers of student data points to avoid bias in their studies. However, a natural side effect is that there is little nuance to the regression data and thus no one can tell which educators, courses, divisions, or institutions are doing well and which are not. Perhaps in addition to the overall unbiased data, a more in-depth look at institutional scores in the regression statistics may suggest a successful model for other institutions to emulate.

Unfortunately, with respect to the thousands of student data points pulled for each regression discontinuity study, neither the CCRC nor other researchers has revealed any particular institutions whose scores are higher than the average so that individuals may judge for themselves whether certain developmental programs are indeed better than average. The CCRC has instead focused on a few particular promising programs around the nation—such as accelerated learning programs, multiple placement measures, bridge programs, and so forth—but none of these appears to come directly from current modes of instruction in remedial classrooms in the nation that are performing above average. This further contributes to the oft-stated and flawed assumption that all remediation is "broken."

Even though some researchers are disappointed with remediation's apparent lack of boost for students at the cutoffs, the fact that most results are null suggests that fully half of the students are performing above average. Could those above-average students be localized in a few classrooms, divisions, or institutions? Could five community colleges in a study of twenty-three, for example, be raising the average significantly? If so, which particular institutions are raising the bar and what exactly are they doing? These are questions left unanswered by large statistical analyses. Perhaps an in-depth examination of the above-average student base would reveal keys about what developmental educators are currently implementing to enhance student success.

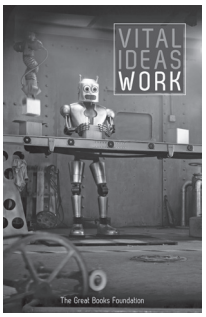
The Overall Evidence is Mixed

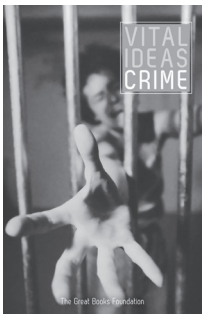
Because a few RDD studies have found positive results, many have found null results, and several have found negative results, and because these RDD studies analyze a wide range of placement test cut-offs, Bailey et al. (2013) argue that developmental education overall is ineffective. CCRC researchers and others have made a statement such as this repeatedly in numerous articles. After considering the same evidence with a contextualized understanding of remediation's purpose, however, we believe it would be best to make a more cautious statement instead: "When applied to developmental education, the RDD approach finds mixed results."

Indeed, some scholars associated with the CCRC do state the evidence as such. For instance, the CCRC's Michelle Hodara (2011), in a working paper on reforming math pedagogy, summarizes the same RDD evidence in this manner: "A majority of community college students enroll in developmental education (Bailey, Jeong, & Cho, 2010), but evidence of its effectiveness in promoting student progression and degree completion is mixed" (p. 1), and with this she goes on to cite the four most common RDD studies examined in our original paper (Bettinger & Long, 2009; Boatman & Long, 2010; Calcagno & Long, 2008; Martorell & McFarlin, 2008). We believe this is the most apt characterization of the RDD body of studies, and the use of such language will help tamp down the inherently inflammatory nature of wording such as "ineffective," a phrase which may lead to more legislators attempting to pass laws which limit or remove developmental education courses or funding.

CONTINUED ON PAGE 32

Money, Sex, Work, and Crime...
Great Ideas are VITAL IDEAS!





Vital Ideas, the newest series from the Great Books Foundation, answers your need for rich content-based material for developmental reading courses. The Great Books Foundation, a non-profit educational organization, advances the critical, reflective thinking skills of students in and out of the classroom. Since 1947, Great Books readers have been using Shared Inquiry™ to discuss works and ideas of enduring value.

Each volume in *Vital Ideas—Sex, Crime, Money,* and *Work*— focuses on a provocative topic with some of the best classic and contemporary literature. Questions accompanying each selection prompt lively discussions that will keep everyone talking long after the formal discussion ends.

Selections include works by Flannery O'Connor, Amy Tan, John Cheever, David Sedaris, Billy Collins, Sigmund Freud, Margaret Atwood, Jack London, and Martin Luther King, Jr.

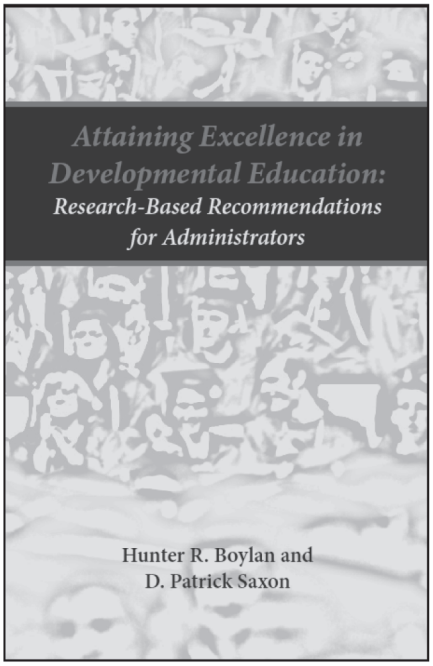
To learn more about Vital Ideas, teacher training, to request an exam copy, or for information about Shared Inquiry™ call us at (312) 646-7108 or visit our bookstore at www.greatbooks.org.

GreatBooks Foundation

read.think.**discuss**.grow.

Attaining Excellence in Developmental Education: Research-Based Recommendations for Administrators

By Hunter R. Boylan, Ph.D. and D. Patrick Saxon, Ed.D.



Applying lessons learned from years of studying research and best practices in developmental education at high performing institutions, ***Attaining Excellence*** is designed to provide recommendations to administrators that will contribute to excellence in the developmental education classroom. It is organized into two sections. Section One recommends actions that cost little or nothing to implement. Section Two recommends actions that involve the expenditure of resources and provides justification for doing so. Appendices are provided which include noncognitive assessment instruments, recommended readings for developmental educators, and a checklist for administrators to use in determining the extent to which they have made decisions and assigned the resources necessary for excellence in developmental education.

AVAILABLE NOW: $29.99 each plus $3 shipping and handling.

Order your copy today from www.ncde.appstate. edu or by calling 828-262-3057

CONTINUED FROM PAGE 30

Conclusion

Nowhere in Bailey et al.'s (2013) response did they refute our arguments about remediation as a barrier or question our section on the cost of developmental education. We must infer then that they do not disagree strongly with our findings that indeed remediation serves more to assist students than to hold them back, and that the overall cost of remediation is not significant when viewed contextually. Of course, they may still disagree with our findings but have chosen instead to focus on other arguments more important to their thesis. It is notable, however, that these issues were not addressed in their rebuttal.

In spite of our disagreements, however, we wish to reaffirm our collegiality with the researchers of the CCRC. Their data-based research in the field of community college postsecondary education is much needed. We simply diverge at various points along their line of reasoning regarding the use and interpretation of RDD studies and their assessment of the efficacy of developmental education. We also wish that they and others would identify the positive approaches many instructors and institutions are currently implementing, research their current developmental education curricula more in depth, and use the RDD datasets to perform these analyses.

Nonetheless, the CCRC has continually and rightly stressed that pedagogical and organizational reforms are necessary in improving postsecondary education. We feel honored to be colleagues in this pursuit. Most of all, we wholeheartedly concur with Bailey et al. (2013) in their rebuttal's concise conclusion, and we highly recommend this course of action for postsecondary educators:

> Reformers must thoughtfully design models that not only shorten developmental sequences and use corequisites when appropriate, but also strengthen curricular alignment, leverage noncognitive measures as part of the placement system, integrate strong academic and nonacademic supports, and tie developmental education more closely to college-level programs. (p. 13)

References

Arendale, D. R. (2012). Access at the crossroads: Learning assistance in higher education. *ASHE Higher Education Report*, *35*(6). San Francisco, CA: Jossey-Bass.

An act concerning college readiness and completion, Substitute Senate Bill No. 40, Public Act No. 40, §1b (2012). Retrieved from http://cga.ct.gov/2012/ACT/PA/2012PA-00040-R00SB-00040-PA.htm

Bailey, T. R., Jaggars, S. S., & Scott-Clayton, J. (2013). *Characterizing the effectiveness of developmental education: A response to recent criticism*. New York, NY: Community College Research Center, Columbia University, Teachers College. Retrieved from http://ccrc.tc.columbia.edu/publications/characterizing-effectiveness-of-developmental-education.html

Bailey, T. R., Jeong, D.W., & Cho, S.W. (2010). *Student progression through developmental sequences in community colleges* (CCRC Brief No. 45). New York, NY: Community College Research Center, Columbia University, Teachers College.

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, *44*(3), 736–771.

Boatman, A. ,& Long. B. T. (2010). *Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation* (An NCPR Working Paper). New York, NY: National Center for Postsecondary Research, Columbia University, Teachers College.

Boylan, H.R. (1999, July). *Developmental education 101: An introduction to the field*. Presented at the Kellogg Institute for the Training and Certification of Developmental Educators, Boone, NC.

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

Community College Research Center (CCRC). (2013). *About Us*. Retrieved from http://ccrc.tc.columbia.edu/About-Us.html

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, *142*(2), 636-654.

Cook, T. D., & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity design: Implications for its theory and its use in research practice. *Annales d'Economie et de Statistique*, *91/92*, 127-150.

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

Hodara, M. (2011). *Reforming mathematics classroom pedagogy: Evidence-based findings and recommendations for the developmental math classroom* (CCRC Working Paper No. 27). New York, NY: Community College Research Center, Columbia University, Teachers College.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House Elections. *Journal of Econometrics*, *142*(2), 675-697.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281-355.

Martorell, P., & McFarlin, I. (2008). *Help or hindrance? The effect of college remediation on academic and labor market outcomes*. Unpublished manuscript. RAND Corporation, Santa Monica, CA, and University of Michigan.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37*(5), 525-559.

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2012). *Improving the targeting of treatment: Evidence from college remediation* (NBER Working Paper No. 18457). Cambridge, MA: National Bureau of Economic Research.

Shadish W.R., Cook T.D., & Campbell D.T. (2002). *Experimental quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.