The Effect of English Language Learner Reclassification on Student ACT Scores, High School Graduation, and Postsecondary Enrollment: Regression Discontinuity Evidence from Wisconsin

Deven Carlson
Jared E. Knowles

Abstract

The recent increase in the number of students classified as English language learners (ELLs) has focused significant attention on reclassification policy, which governs the process by which ELLs move toward, and are deemed to reach, full English proficiency. In this paper, we draw on a data set containing annual individual-level records for every Wisconsin student ever classified as an ELL between the 2006–07 and 2012–13 school years to estimate the effects of being reclassified at the end of 10th grade—a crucial period on the pathway to postsecondary education—on several measures related to students' postsecondary attainments. We estimate these effects in a regression discontinuity framework, exploiting Wisconsin's policy rule that automatically reclassifies ELLs who score above a specified cutoff on the state's English language proficiency exam. Our analysis indicates that being reclassified as fully English proficient in 10th grade has a positive effect on students' ACT scores. It also provides some evidence of a positive effect on high school graduation and the probability of enrolling in a postsecondary institution the fall after graduation. Together, our analyses provide evidence on the effects of a policy directly relevant to the country's fastest growing student population, and we close the paper with a discussion of the implications for research and policy.

INTRODUCTION

The number of students classified as English language learners (ELLs) has increased by 14 percent in the past decade, and today these 4.7 million individuals comprise approximately 10 percent of all students attending public schools in the United States (Snyder & Dillow, 2013). Such rapid growth has increased the focus on policies surrounding the education of this student population, with particular attention paid to reclassification policy, which governs the process by which ELLs move toward—and are deemed to reach—full English proficiency. As states and districts are faced with the challenge of educating increasing numbers of ELLs, however, they are becoming increasingly aware that our understanding of reclassification policy is limited. Currently, relatively little is known about how reclassifying ELLs

Journal of Policy Analysis and Management, Vol. 35, No. 3, 559–586 (2016)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21908

as fully English proficient affects important student outcomes, particularly those related to postsecondary experiences.

Drawing on a data set containing annual individual-level records for every Wisconsin student ever classified as an ELL between the 2006–07 and 2012–13 school years, in this paper we estimate the causal effect of being reclassified at the end of 10th grade—a crucial point on the pathway to postsecondary education—on several outcomes related to students' postsecondary attainments, including ACT scores, high school graduation, and postsecondary enrollment. We estimate these effects in a regression discontinuity framework, exploiting Wisconsin's policy rule that automatically reclassifies ELLs who score above a specified cutoff on the state's English language proficiency assessment—the Assessing Comprehension and Communication in English State-to-State (ACCESS) exam—to identify the effect of reclassification. Together, our analyses provide evidence on several effects of a policy directly relevant to the education of the country's fastest growing student population.

Our analysis indicates that being reclassified as fully English proficient in 10th grade has a positive effect on students' composite ACT scores and the probability of enrolling in a postsecondary institution the fall after high school graduation. It also provides suggestive evidence of a positive effect on high school graduation. We demonstrate that the positive effect of reclassification on students' composite ACT score is primarily attributable to improved performance on the English and reading portions of the test. With respect to postsecondary enrollment, our analysis provides evidence that the positive effect of reclassification stems from increased enrollment at four-year institutions.

We proceed by providing background information on ELL reclassification and summarizing the limited previous research on this topic before describing the data we draw upon. Our empirical analysis proceeds by first demonstrating the validity of our regression discontinuity design before detailing our approach to estimating the effect of being reclassified in 10th grade on students' ACT scores, likelihood of high school graduation, and probability of postsecondary enrollment. After presenting the results of our analyses we close the paper with a discussion of their implications for research and policy.

ELL RECLASSIFICATION IN WISCONSIN

Within broad federal requirements laid out in Title III of the Elementary and Secondary Education Act, states are given substantial latitude in determining the details of ELL education, such as the process for identifying students as ELLs, the content of the proficiency standards, and the specific assessment used to gauge student mastery of those standards, among others. The emphasis of this paper, however, leads us to focus primarily on the processes and criteria related to reclassifying students as fully English proficient. Given the latitude granted them, it is perhaps not surprising that states employ a wide variety of approaches to determine whether to reclassify a student as English proficient. Recognizing this variation, three general patterns and commonalities do exist in reclassification practices across states. First, student performance on an English language proficiency exam informs the reclassification decision in most states.[1] Second, most states either require or recommend that additional criteria inform the reclassification decision, such as student performance on standard content assessments, student grades or other measures of academic

[1] According to a recent evaluation by the U.S. Department of Education (2012), 46 states and D.C. required districts to consider student performance on the proficiency assessment in the reclassification decision and another three states recommended it.

performance, teacher input, district-level committee recommendations, and parental consultation (U.S. Department of Education, 2012).[2] Finally, a majority of states provide districts with at least some discretion in determining whether to reclassify a student as English proficient—only 18 states specify a process or set of criteria that districts are compelled to follow while the remaining 32 provide districts with some flexibility over the process.

Like the nation as a whole, Wisconsin has experienced a marked increase in the number of ELL students in recent years. In the ten years from 2004–2005 to 2013–2014, the number of ELL students in Wisconsin grew by 26 percent, which corresponds to an addition of over 10,000 ELL students during this time period.[3] There are also notable differences, however, between the ELL context in Wisconsin and that in states where ELLs are typically studied. First, in contrast to the typical context, one where ELLs are concentrated in a small number of schools in a state's urban areas (Kanno & Kangas, 2014), ELLs in Wisconsin are spread quite evenly across urban, suburban, and rural areas. For example, in our analyses below, only about 20 percent of our sample resides in Milwaukee or Madison, which are the two major urban areas in Wisconsin. Second, unlike states that have a single, dominant language minority group, Wisconsin has two language minority groups that are approximately equal in size. Slightly more than 40 percent of our sample consists of students who speak Hmong as their native language, with another 40 percent consisting of students who are native Spanish speakers.

In Wisconsin, ELL students can be reclassified as fully English proficient both automatically—solely on the basis of their performance on the state's English language proficiency assessment—and manually, on the basis of both their assessment performance and subjective evaluation by district personnel. Like over 25 other states, Wisconsin uses the ACCESS exam to gauge students' English proficiency and it is administered annually to every student classified as an ELL. The ACCESS exam consists of four language domains—listening, speaking, reading, writing—and students receive scale scores in each of these domains as well as in four composite areas—oral language, literacy, comprehension, and overall—constructed from students' performance in various combinations of the four domain areas.[4] Students' scale scores are then mapped to one of six proficiency levels and it is on the basis of students' assigned proficiency levels that reclassification occurs.[5] The ACCESS exam is administered in November or December of each year, and the corresponding reclassification decision occurs at the end of the school year—reclassified students are treated as fully proficient at the beginning of the following school year.

Wisconsin's current reclassification policy states that ELL students can be automatically reclassified in two ways. First, ELLs in grades K-12 are automatically reclassified as fully proficient if they achieve an overall composite proficiency level of six or greater—the highest level—on the ACCESS. Second, the policy states that ELLs in grades 4 through 12 are automatically reclassified as fully proficient if they achieve an overall composite proficiency level of five or greater on the ACCESS plus a proficiency level of five or greater on the literacy composite. Although students are currently automatically reclassified according to both criteria, the second set

[2] In particular, 15 states require consideration of content-area achievement and six states recommend it; 11 states require or recommend other academic performance measures (e.g., grades), and smaller numbers of states specify other factors (e.g., teacher input, parental consultation, etc.).

[3] For context, Wisconsin had approximately 870,000 students in the K-12 system during the 2013 to 2014 school year. See http://esea.dpi.wi.gov/sites/default/files/imce/esea/pdf/bul_0701.pdf for a description of the process for identifying a student for ELL status.

[4] See https://www.wida.us/assessment/access/scorereports/access_interpretive_guide11.pdf for the contribution of each language domain to each of the composite scores.

[5] The six proficiency levels are 1—Entering, 2—Beginning, 3—Developing, 4—Expanding, 5—Bridging, and 6—Reaching.

became operational only recently. Consequently, our empirical analysis of the effects of reclassification relies on the first set of automatic reclassification criteria.

Along with automatic reclassification, Wisconsin policy also allows for students to be manually reclassified. Specifically, ELL students who achieve an overall composite proficiency level of five on the ACCESS, but fail to achieve a proficiency level of five on the literacy composite, can be manually reclassified as fully proficient if the district determines that the student clearly demonstrates English proficiency. On the flip side, students who were automatically reclassified as fully proficient can be manually classified back into ELL status if the district determines via observation and analysis of academic performance that the student is not fully proficient.[6] The data indicate, however, that relatively few students are manually reclassified, either from ELL status to fully proficient or vice versa.

Compared to students classified as fully English proficient—who are typically mainstreamed into classes with their native English-speaking peers and receive no English language instruction—the educational context experienced by students classified as ELLs in Wisconsin is markedly different. Most notably, students classified as ELLs spend a substantial proportion of their school day receiving instruction in reading, writing, and speaking the English language, either through English as a Second Language (ESL) or bilingual-bicultural (BLBC) programs.[7] In Wisconsin, approximately two-thirds of ELLs are enrolled in BLBC programs while the other one-third receives ESL instruction. And although districts have substantial autonomy in how they administer their BLBC or ESL programs, they generally involve removing ELLs from mainstream classrooms, at least to some degree and often during reading or English class. An additional difference in educational context on the basis of ELL status stems from the availability of instructional accommodations for students classified as ELLs. These classroom practices and procedures are designed to allow students to better access instructional content by mitigating obstacles associated with students' lack of proficiency in the English language. The set of possible instructional accommodations are categorized into three groups, including those that provide: (1) direct linguistic support in English; (2) direct linguistic support in the student's native language; and (3) indirect linguistic support.[8]

RECLASSIFICATION AND STUDENT OUTCOMES

Reclassifying a student as fully English proficient represents the culmination of a process designed to develop ELLs' mastery of the English language.[9] As such, both policymakers and scholars have historically treated reclassification as a goal to work toward—an outcome for students, schools, and districts to achieve. In the policy community, the outcome-oriented view of reclassification is reflected in the content of policy governing the education of ELLs—many states specify annual increases in the proportion of students attaining reclassification as explicit goals—and by the use of reclassification rates to gauge the success of a particular policy or

[6] Guidance accompanying the manual reclassification policy states that districts should have at least two pieces of evidence in support of the reclassification decision. See http://dpi.wi.gov/sites/default/files/imce/esea/pdf/bul_0702.pdf for more detail.
[7] Whether an ELL student receives instruction via an ESL or BLBC program depends in part on the concentration of ELLs in the student's district. See Wisconsin state statute 115.96-115.97 for more detail.
[8] ELLs are also eligible to receive assessment accommodations on the Wisconsin Knowledge and Concepts Examination (WKCE). See http://ell.dpi.wi.gov/files/ell/doc/ell-accommodations-guide.doc for a discussion of accommodations for ELLs in Wisconsin.
[9] Many studies examine the effectiveness of English-only versus bilingual approaches to educating ELLs—see Greene (1997) and Slavin and Cheung (2005) for reviews. Studies also address issues such as full-day kindergarten (Cannon, Jacknowitz, & Painter, 2011) or coursetaking patterns (Conger, Long, & Iatarola, 2009) among the ELL population.

practice relevant to the education of ELLs (e.g., Grissom, 2004; Parrish et al., 2006). Among scholars, a substantial amount of work has been devoted to developing, cataloging, and analyzing criteria relevant to both the reclassification decision and corresponding reclassification rates. For example, a number of studies examine the determinants of school or district reclassification rates (e.g., Hill, Weston, & Hayes, 2014; Jepsen & de Alth, 2005) or assess the validity of the reclassification process (e.g., Abedi, 2008; Kim & Herman, 2010; Mahoney & MacSwan, 2005).

Although the outcome-oriented view of reclassification has historically represented the dominant conception of the issue, a growing number of studies recognize—sometimes implicitly—that reclassification can also be considered a policy intervention. Reclassification has the potential to affect student outcomes because it changes several features of the environment in which students are educated (Robinson, 2011). Most visibly, reclassification typically results in a decline in—or even a complete elimination of—instruction designed to develop English language skills. Such a change alters both the mix of teachers and the peer group to which students are exposed on a day-to-day basis. At the high school level, reclassification has the potential to put students on an educational track that may differ greatly from the one they would have experienced had they remained classified as an ELL (Callahan & Gándara, 2004; Schiller & Muller, 2000). These differences could have both positive and negative effects on students' postsecondary-related outcomes. On one hand, reclassification as fully proficient may expose students to coursework that better prepares them for the ACT (Callahan, Wilkinson, & Muller, 2010). More generally, English proficiency status may result in students being guided toward an educational track where they receive greater support in pursuing postsecondary education—such as assistance with the financial aid or application process—relative to students classified as ELLs (Callahan & Gándara, 2004). On the other hand, ELL status provides students with access to instructional accommodations that may allow them to better access, engage, and understand the instructional content presented in the classroom, which could translate into improved performance on the ACT and, perhaps, better postsecondary prospects. Ideally, the transition from ELL status to classification as fully English proficient should be smooth, with no effect on student outcomes. However, the changes in students' educational environment induced by reclassification result in a scenario where that might not be the case.

A limited, although increasing, number of studies have conducted analyses designed to gain insight into the effects of reclassification, most often on content-area achievement (e.g., Flores et al., 2009; Grissom, 2004; Robinson, 2011) but also on other outcomes such as dropping out of school (Kim, 2011), attendance (Robinson, 2011), or enrollment in college preparatory coursework (Callahan, Wilkinson, & Muller, 2010). The most common design of these analyses involves comparing the outcomes of students classified as ELLs to those of students who have been reclassified as proficient (Flores et al., 2009; Grissom, 2004) or to observably similar students not classified as ELLs due to variation in policies used to classify students as such (Callahan, Wilkinson, & Muller, 2010). Two of these studies demonstrate that reclassification is related to improved academic performance (Flores et al., 2009; Grissom, 2004) while Callahan, Wilkinson, and Muller (2010) return evidence of a heterogeneous relationship—relatively recent immigrants and those with low levels of English proficiency appear to benefit from ELL status whereas students who have retained their ELL classification for a relatively long time do not benefit, and may actually be harmed by ELL classification. Perhaps most relevant to our analysis, however, is Callahan, Wilkinson, and Muller's (2010) finding that students classified as ELLs are about 50 percent less likely to enroll in college preparatory coursework than students who are similarly proficient in English but not classified as ELLs. Although Callahan, Wilkinson, and Muller (2010) only explicitly study college preparatory coursework, their results provide a basis for expecting differences on the basis of ELL classification for other postsecondary-related outcomes.

The studies reviewed above provide valuable information on the relationship between ELL status and important student outcomes, but none of them employ designs that convincingly remove the threat of bias from unobservable factors.[10] Such a critique cannot be leveled against Robinson's (2011) study, which exploits the policy rule governing reclassification of ELLs in a single California district to estimate the effect of reclassification on content-area achievement in a regression discontinuity framework.[11] This study finds reclassification to have a negative effect on the language arts achievement of high school students, but no effect on the scores of elementary or middle school students. Robinson-Cimpian and Thompson (2015) extend this study by using a "differences-in-discontinuities" design to estimate the effect of increasing the stringency of reclassification criteria, which makes classification as fully proficient more difficult to obtain on students' achievement and high school graduation outcomes. The study provides evidence that this policy change had a significant positive effect on students' English language arts achievement as well as their probability of high school graduation. The authors attribute this effect to the fact that reclassification negatively affected these outcomes under the old criteria, but had no effect under the new, more rigorous reclassification criteria.

We build on the work of Robinson (2011) and Robinson-Cimpian and Thompson (2015) by estimating the causal effect of reclassification on several outcomes related to students' postsecondary attainments, including ACT scores, high school graduation, and postsecondary enrollment. In doing so, we focus on the effects of reclassification that occurred at the end of 10th grade, which induces a change in ELL status that takes effect at the beginning of students' junior year. We focus on the effect of reclassification at this point in time for multiple reasons, both methodological and substantive. Methodologically, the nature of the reclassification process—coupled with our research design—is best suited for focusing on the effects of reclassification in a given year, particularly one where the timing of reclassification is likely to be consequential. As we describe in greater detail below, our regression discontinuity approach compares future outcomes for students just above the reclassification threshold to outcomes for students just below the threshold. However, because students are eligible for reclassification each school year, the difference in ELL status on the basis of a student's ACCESS score is likely to only exist for a single year, as students just below the reclassification threshold in any given year are likely to score above the threshold the following year and thus achieve classification as fully proficient.

Substantively, there are three main reasons we believe that reclassification occurring at the end of students' 10th-grade year—affecting their ELL status throughout 11th grade—is likely to be particularly consequential for students' future postsecondary outcomes. First, most students who take the ACT or SAT do so at the end of their junior year. In Wisconsin, many districts have long administered the ACT to all juniors, and have thus structured their 11th-grade curriculum—or at least the curriculum in their college preparation track—in a manner designed to prepare students for this college entrance exam. However, because many schools and districts do not consider all students to be on a college track, it is possible that students classified as fully English proficient at the beginning of their junior year could be treated

[10] Flores et al. (2009) and Callahan, Wilkinson, and Muller (2010) employ approaches that adjust for a variety of observable characteristics; Grissom (2004) presents unadjusted differences.
[11] Using a similar design and data from a California district, Matsudaira (2005) finds negligible achievement differences. However, the cutoff on the English proficiency assessment for determining ELL status is also the cutoff for determining bilingual education versus English-only instruction, complicating the interpretation.

very differently from a curricular perspective than students who remain classified as ELLs (e.g., Callahan & Gándara, 2004; Callahan, Wilkinson, & Muller, 2010); these curricular differences could manifest themselves in students' ACT performance. Second, to the extent that ACT performance—itself potentially affected by 10th-grade reclassification—influences other postsecondary outcomes, such as college enrollment or type of institution attended, it is natural to maintain the focus on 10th-grade reclassification when examining these additional outcomes. Third, students' junior year is typically when high schools begin college counseling in earnest. Indeed, it is often during 11th grade that counselors first discuss topics such as financial aid, college applications, and entrance exams with students—guidance counselors have informally described the timing of postsecondary counseling as "seniors in the fall; juniors in the spring" (Fazekas & Warren, 2011).[12] And although counselors typically meet with all students to discuss postsecondary plans, the nature of the conversation could differ significantly on the basis of students' ELL status (Callahan & Gándara, 2004). Together, these three considerations suggest the importance of students' ELL status in their junior year with respect to postsecondary preparation and, in doing so, motivate our focus on estimating the effects of 10th-grade reclassification.

DATA

We constructed a broad data set from annual student-level records maintained by the Wisconsin Department of Public Instruction (DPI) to serve as the basis of estimating the effects of 10th-grade reclassification. The first step in constructing this data set involved identifying every Wisconsin student ever classified as an ELL between the 2006–07 and 2012–13 school years. Having identified all such students, we then extracted all annual records for each of those students, regardless of a student's ELL status in a given year. As a result, our data set contains all available records from the 2006–07 school year through the 2012–13 school year for every student ever classified as an ELL during that time period.

The student-level records we extracted contain a wide variety of information on student demographics, English language proficiency, ACT performance, high school graduation, postsecondary enrollment, and other topics that are instrumental to the analyses below. Demographically, our data set contains standard measures such as district and school of attendance, age, grade, gender, and subsidized lunch eligibility, but also less common measures such as native language, migrant status, and homeless status. With respect to English proficiency, the data set contains a time-varying measure indicating whether a student is classified as an ELL, the number of years that the student has been classified as an ELL, and a measure indicating whether a student was reclassified as fully proficient in that year. For years in which a student is classified as an ELL, and thus took the ACCESS exam, the data set contains the scale score and associated proficiency level that students achieved in each of the four language domains and in each of the four composite areas.

Along with information on English language proficiency, our data also contain three sets of measures relevant to students' postsecondary outcomes. First, our data contain ACT scores for all students who took the assessment between the 2006–07 school year and the 2012–13 school year, which is the full time period our DPI records span. Students typically take the ACT—the primary college entrance exam taken by students in Wisconsin—in the spring of 11th grade and DPI receives the

[12] Many policy interventions are designed to improve the postsecondary outcomes of disadvantaged 11th graders, including the California Early Assessment Program (Howell, Kurlaender, & Grodsky, 2010), Career Beginnings (Cave & Quint, 1990), and College Possible (Avery, 2013).

scores of all tested students in the state directly from ACT. Our data contain students' composite scale score, as well as their scale score in each of the four subject domains comprising the assessment—reading, English, math, and science.[13] Second, our data contain a measure indicating whether a student graduated from high school by the end of the 2012–13 school year. Third, our data contain records of postsecondary enrollment for all students in our sample. DPI contracts with the National Student Clearinghouse (NSC) to obtain student postsecondary records and then matches the NSC records to the student records maintained by the agency.[14] As a result, our data set includes a measure indicating whether students enrolled in a postsecondary institution in the fall following their graduation from high school. It also contains measures indicating whether students were ever enrolled in a two-year institution and whether they were ever enrolled in a four-year institution.[15]

ESTIMATING THE EFFECTS OF RECLASSIFICATION: A REGRESSION DISCONTINUITY APPROACH

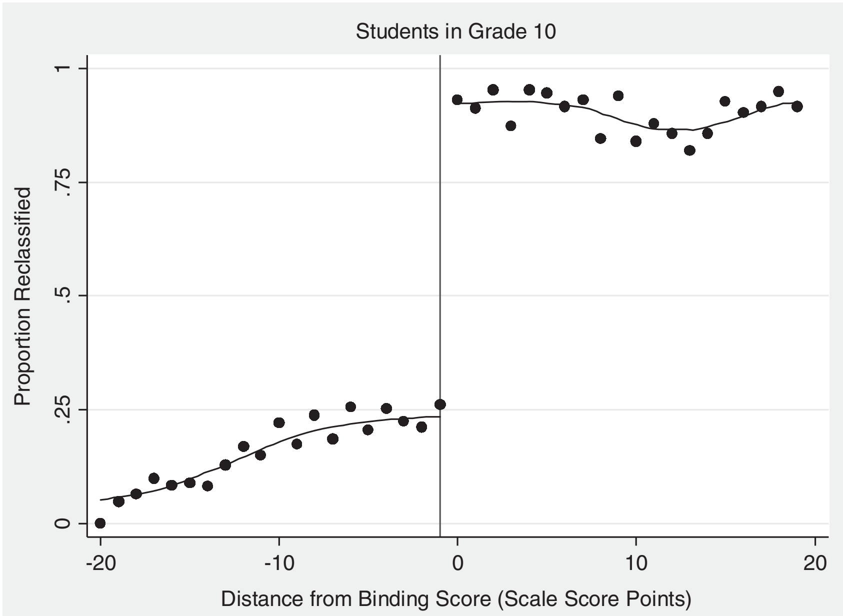
To estimate the effect of being reclassified as fully English proficient in 10th grade on ACT or postsecondary outcomes, we might begin with an empirical model of the form:

$$Y_i = \alpha + \lambda R_{iGrade10} + X_{iGrade10}\beta + \varepsilon_i \quad (1)$$

where y represents the outcome of interest—such as ACT score, postsecondary enrollment, or high school graduation—for student i; R indicates whether the student was reclassified as fully proficient; X represents a set of observable student, school, and district characteristics; and ε is the error term. The parameter of interest in this model is represented by λ, which can be straightforwardly estimated using ordinary least squares (OLS) techniques and data on students who were and were not reclassified as English proficient in 10th grade. The main obstacle to interpreting the OLS estimate of λ as the causal effect of reclassification stems from the potential endogeneity of R; there may be unobserved factors—such as home environment, motivation, or any number of other influences—related to both a student's reclassification status as well as the outcome of interest, thus resulting in a biased estimate of λ. This necessitates identifying a plausibly exogenous source of variation in R in order to obtain a credible causal estimate of the effects of reclassification.

We rely on the design of Wisconsin's reclassification policy—particularly the automatic reclassification provision—as a source of exogenous variation in ELL status that we leverage to estimate the causal effect of being reclassified on student achievement and attainment outcomes, at least for those students very close to the automatic reclassification threshold. As described above, Wisconsin automatically reclassifies ELLs as fully proficient if they achieve a proficiency level of six on the ACCESS exam while students who achieve a proficiency level of five or below remain classified as ELLs. Because ACCESS proficiency levels map directly to nonoverlapping scale score ranges, a single scale score point separates students who are automatically reclassified as English proficient from those who remain classified as ELLs. We exploit this policy rule to conduct a regression discontinuity analysis,

[13] For the small number of students who took the ACT multiple times, we include the scores associated with the highest composite score in our data.
[14] NSC data cover approximately 95 percent of students enrolled in two- and four-year institutions and the records contain information on students' enrollment begin and end dates. See Dynarski, Hemelt, and Hyman (2013) for in-depth description of the NSC data.
[15] The analyses of high school graduation and postsecondary outcomes are restricted to school years up to and including 2010 to 2011 in order to allow sufficient time to potentially graduate high school and enroll in a postsecondary institution.



Notes: The markers in the figure represent the mean reclassification rate at each scale score within 20 points of the automatic reclassification threshold on the ACCESS exam. The line is a smoothed, kernel-weighted mean calculated separately on each side of the cutoff. The mean smoothing was performed using an epanechnikov kernel and a rule-of-thumb bandwidth.

Figure 1. Proportion of 10th-Grade Students Reclassified as English Proficient, by Distance from Binding ACCESS Scale Score.

which bases identification of the effect of reclassification on the assumption that students who score just above the reclassification threshold are no different—on both observable and unobservable characteristics—from those who score just below the threshold. Below we provide strong evidence in support of the validity of this assumption and thus the causal nature of our estimates.

Validity of the Design

The first step in establishing the validity of our regression discontinuity design involves demonstrating a difference in the likelihood of reclassification for students who score on either side of the threshold established by Wisconsin's reclassification policy. Figure 1 provides such a demonstration. The markers in the figure represent the mean reclassification rate at each scale score within 20 points of the automatic reclassification threshold on the ACCESS exam while the line is a smoothed, kernel-weighted mean calculated separately on each side of the cutoff.[16] The figure makes clear that approximately 25 percent of students who score just below the automatic reclassification threshold receive a designation of fully English proficient while over 90 percent of students who score above that threshold receive such a designation—the graph clearly demonstrates a substantial jump in reclassification at the specified threshold. The difference in the probability of reclassification for students on either

[16] The mean smoothing was performed using an epanechnikov kernel and a rule-of-thumb bandwidth. Alternative kernel types or bandwidths produce similar results.

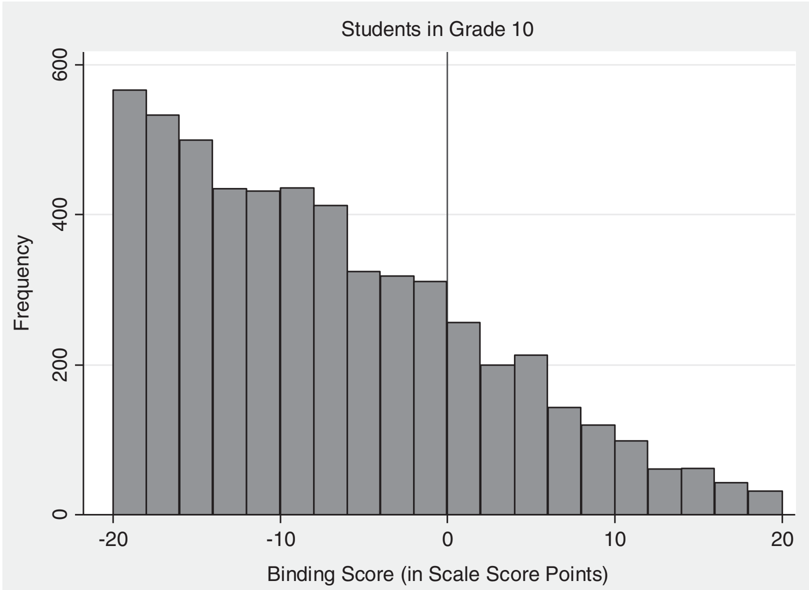


Figure 2. Distribution of 10th-Grade Students by Distance to Binding ACCESS Scale Score.

side of the threshold is subjected to formal statistical testing in our analyses below, and is found to be large and highly significant. The probabilistic, rather than deterministic, nature of the relationship between ACCESS achievement and reclassification is not problematic—it requires an analytic approach appropriate for a "fuzzy," as opposed to "sharp," regression discontinuity design.

It is clear that there is a large discontinuity in the probability of reclassification at the threshold specified by the policy rule, but the validity of the design could be threatened if students or schools are able to manipulate ACCESS results in a manner that systematically affects whether students score below or above the threshold. Such manipulation is unlikely given the policies and procedures surrounding ACCESS administration—proficiency thresholds are set by ACCESS developers and the assessments are scored by neither students nor school or district personnel—but we nonetheless undertook multiple tests designed to detect any manipulation. First, we simply plot the density of cases around the reclassification threshold to assess whether there is a disproportionate stacking of students on either side of the cutoff, which would be evidence of potential manipulation. Figure 2 provides no evidence of disproportionate stacking near the cutoff. The visual evidence that cases are not heaped near the cutoff is corroborated by the results of the statistical test proposed by McCrary (2008), which fails to reject the null hypothesis of no change in the density of cases at the threshold.[17]

Although seemingly unlikely given the relative smoothness in the density of cases around the reclassification threshold, it is possible that the characteristics of students just above the cutoff are systematically different from those just below—any such differences would threaten the validity of the design. To assess the likelihood of differential student composition we check for differences in observable student characteristics across the automatic reclassification threshold. Specifically, using

[17] The coefficient estimate produced by this test is −0.093 with a standard error of 0.048.

an analytic sample of all students with a composite ACCESS scale score within 10 points of the automatic reclassification threshold, we estimate:

$$O_{iGrade10} = f(A_{iGrade10}) + \tau H_{iGrade10} + C_{iGrade10}\theta + \varepsilon_{iGrade10} \quad (2)$$

where O represents an observable characteristic of student i in 10th grade, $f(A_i)$ is a flexible function of the distance in ACCESS scale score points from the automatic reclassification threshold (i.e., the running variable), H is an indicator for scoring above the threshold, C is a vector of school year fixed effects, and ε is the error term. In the results presented below we specify $f(A_i)$ to contain a linear term with different slopes below and above the cutoff, but substantively similar results are returned from other specifications, such as one that also includes a quadratic term with different slopes on each side of the cutoff. We estimate equation (2) separately for the following observable characteristics: sex, free lunch eligibility, race/ethnicity, disability status, 10th-grade reading and math achievement (taken prior to the ACCESS exam), initial English proficiency level, and composite scale score on the prior administration of the ACCESS exam.[18] Results from these models are presented in Table 1 and visual corroboration is provided by Figure A1 in Appendix A, which plots the means of each observable characteristic by the distance from the automatic reclassification threshold along with a line of best fit that is fitted separately on each side of the automatic reclassification threshold.[19] We combine the 11 separate tests in Table 1 into a single test statistic by estimating a Seemingly Unrelated Regression and conducting a chi-squared test of the hypothesis that the estimated coefficients for the indicator of scoring above the reclassification threshold across the 11 regressions are jointly equal to zero (see, e.g., Chin, Daysal, & Imberman, 2013). This test is unable to reject the null hypothesis that the coefficients are jointly equal to zero—the p-value for this test is 0.790—and thus provides further evidence on the validity of our design.[20]

Statistical Models

Considered together, the description of the assignment rule, the clear discontinuity in the proportion of students reclassified above and below the threshold, and our validity checks provide confidence in the ability of our design to return valid causal estimates. With this confidence, we estimate the effects of 10th-grade reclassification on outcomes related to students' postsecondary attainments using the following reduced-form model:

$$Y_i = f(A_{iGrade10}) + \delta H_{iGrade10} + X_{iGrade10}\beta + \varepsilon_i \quad (3)$$

where i indexes students and Y represents one of the following outcomes of interest:

- Taking the ACT in grade 11 or 12;
- ACT scores;
- Graduating high school by the end of the 2012–13 school year;

[18] To assess whether endogenous migration represents a potential validity threat, we estimated a variant of equation (2) where the outcome is an indicator of attending the same school the following school year. The results indicate that endogenous migration is not a concern.
[19] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[20] We estimated a version of the SUR that excludes the prior ACCESS composite score. We also estimated SURs for the ACT and postsecondary samples. All chi-squared tests are unable to reject the hypothesis that the coefficients for scoring above the cutoff are jointly equal to zero.

Table 1. Coefficients and standard errors on measure of scoring above reclassification threshold from reduced-form model predicting student background characteristics.

Outcome variable	*N*	Coefficients (standard errors)
Female	2,733	−0.036 (0.038)
Free lunch	2,733	−0.014 (0.037)
Asian	2,733	0.023 (0.038)
Black	2,733	0.003 (0.009)
Hispanic	2,733	−0.048 (0.038)
White	2,733	0.022 (0.019)
Disability	2,733	0.027* (0.014)
10th-grade reading achievement	2,719	0.043 (0.044)
10th-grade math achievement	2,717	0.089 (0.046)
Initial proficiency level	2,718	0.116 (0.083)
Prior ACCESS composite score	2,322	0.176 (1.078)

Notes: Robust standard errors in parentheses below coefficient estimates. All estimates based on analytic sample containing students within 10 scale score points of reclassification threshold. Reduced-form model controls for distance from the reclassification threshold using a linear term with different slopes below and above the cutoff. Model also contains school year fixed effects. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

- Enrolling in postsecondary schooling the fall after high school graduation;
- Enrolling in a four-year institution by the end of calendar year 2013; and
- Enrolling in a two-year institution by the end of calendar year 2013.

In this model, $f(A_{iGrade10})$ is a flexible function of the distance in scale score points from the automatic reclassification threshold on students' 10th-grade ACCESS exam, H is an indicator for scoring above the threshold on that 10th-grade ACCESS exam, $\boldsymbol{X}$ is a vector of background characteristics from students' 10th-grade year, and ε is the error term.[21] We present results from models where $f(A_{iGrade10})$ is specified as a linear term with different slopes below and above the reclassification threshold, as well as from models where $f(A_{iGrade10})$ is specified as linear and quadratic terms that are each allowed to have different slopes on each side of the cutoff.[22] In all models, we use an analytic sample of students with a composite ACCESS scale score within 10 points of the automatic reclassification threshold and we cluster standard errors by district. Table 2 presents descriptive

[21] The specific background measures include sex, subsidized lunch eligibility, disability status, and school year.
[22] Formally, the first specification of $f(A_{iGrade10})$ can be written as $\tau A_{iGrade10} + \pi(A_{iGrade10}\ H_{iGrade10})$. The second specification can be written as $\tau_1 A_{iGrade10} + \tau_2 A^2_{iGrade10} + \pi_1(A_{iGrade10}\ H_{iGrade10}) + \pi_2(A^2_{iGrade10}\ H_{iGrade10})$.

Table 2. Sample means for grade 10 observations within 10 scale score points of reclassification threshold.

Variable	All observations	Observations below reclassification threshold	Observations above reclassification threshold
	Background characteristics and taking ACT		
Female	0.474	0.480	0.462
Free lunch	0.634	0.649	0.605
Reduced-price lunch	0.141	0.139	0.145
No subsidized lunch	0.225	0.212	0.250
Asian	0.434	0.409	0.483
Black	0.016	0.015	0.018
Hispanic	0.488	0.517	0.431
White	0.059	0.056	0.065
Other race	0.002	0.002	0.002
Disability	0.035	0.038	0.031
Reclassified	0.461	0.222	0.922
Took ACT	0.528	0.498	0.585
N	*2,733*	*1,801*	*932*
	ACT scores		
ACT composite score	17.8	17.2	18.7
ACT English score	17.2	16.7	17.8
ACT reading score	15.8	15.2	16.9
ACT math score	18.8	18.2	19.9
ACT science score	18.7	18.3	19.4
N	*1,442*	*897*	*545*
	Postsecondary outcomes		
Graduated high school	0.895	0.888	0.907
Postsecondary enrollment—fall after high school graduation	0.433	0.405	0.485
Postsecondary enrollment— two-year institution	0.330	0.343	0.305
Postsecondary enrollment—four-year institution	0.390	0.348	0.471
N	*1,863*	*1,217*	*646*

statistics for the analytic samples, both as a whole and separately for students that score above and below the reclassification threshold.[23]

The reduced-form estimates of δ resulting from equation (3) can be interpreted as the effect of scoring above the automatic reclassification cutoff—the estimates represent the effect of increasing the probability that a student is reclassified. To obtain an estimate that can be interpreted as the effect of reclassification per se we employ an instrumental variables (IV) approach commonly used with fuzzy regression discontinuity designs. In this specific application, we use scoring above the

[23] Descriptive statistics in Table 2 reflect the different samples underlying the ACT and postsecondary analyses. The background characteristics are calculated from all observations (i.e., the ACT sample) while the postsecondary outcomes are calculated using observations up to and including the 2010 to 2011 school year.

reclassification threshold as an instrument for being reclassified as fully proficient. As with instruments more generally, scoring above the reclassification threshold is a valid instrument for being reclassified if it: (1) predicts reclassification, and (2) is uncorrelated with the outcomes of interest other than through its effect on reclassification. Figure 1 above clearly demonstrates that scoring above the reclassification threshold is highly correlated with being reclassified as fully English proficient, satisfying the first condition. The second condition—the exclusion restriction—is not directly testable, but will be met if the flexible function of distance from the reclassification threshold is properly specified and student ACCESS scores were not manipulated, which the results of the validity checks above indicate to be the case. We implement this IV approach in a two-stage least squares (2SLS) framework where the first stage predicts reclassification using the following model:

$$R_{iGrade10} = f(A_{iGrade10}) + \psi H_{iGrade10} + \boldsymbol{X_{iGrade10}\pi} + \omega_{iGrade10} \quad (4)$$

where R is an indicator for being reclassified as fully English proficient in 10th grade, $f(A_{it})$ is the flexible function of distance from the reclassification threshold described earlier, H is an indicator for scoring above the reclassification threshold, $\boldsymbol{X}$ is the vector of student background characteristics listed earlier, and ω is the error term. The predicted values of R resulting from estimation of equation (4)—denoted as $\hat{R}$ below—are then inserted into the second-stage equation, taking the place of the indicator for scoring above the reclassification threshold from the reduced-form model above. The second-stage model can be written as:

$$Y_i = f(A_{iGrade10}) + \lambda \hat{R}_{iGrade10} + \boldsymbol{X_{iGrade10}\beta} + \varepsilon_i. \quad (5)$$

Because $\hat{R}$ contains only the variation in reclassification attributable to scoring above the specified cutoff, it is uncorrelated with ε and the resulting estimate of λ thus represents the local average treatment effect (LATE) of reclassification on the outcome of interest. As with LATEs in fuzzy regression discontinuity contexts more generally, the estimate in this application is only generalizable to those students near the threshold for whom scoring above the threshold would have resulted in reclassification as fully English proficient.

Although the generalizability of the estimated effects of reclassification is somewhat limited, these effects still provide valuable information on several important policy issues. Most directly, by estimating the effect of reclassification for students who score just above the threshold, it provides evidence on whether these students are helped or hurt by reclassification, as well as whether students who score below the threshold might also be considered for reclassification. In addition, as Robinson (2011) notes, these results could inform discussions about the policies and procedures surrounding the reclassification decision, such as the menu of criteria or the appropriateness of specified thresholds on relevant assessments. Finally, the estimates are germane to discussions about how reclassification affects students' day-to-day educational experiences, such as their classroom environment, the postsecondary preparation or counseling they receive, or their eligibility for instructional accommodations. So, although we acknowledge the limitations of our LATE estimates, we also believe they are clearly relevant to several important policy issues.

RESULTS

Table 3 presents estimates of the effect of 10th-grade reclassification on ACT-related outcomes. The first column of the table presents the estimated effect on the probability of taking the ACT by the end of the 2012–13 school year while the remaining

Table 3. Coefficients and standard errors on measures of reclassification, by ACT-related outcome measure and model specification.

		ACT scores				
Model	Take ACT	Composite	English	Reading	Math	Science
	Running variable: Linear on both sides of cutoff					
Reduced form—scoring above reclassification threshold	0.042 (0.035)	0.684** (0.310)	0.872** (0.415)	0.962** (0.456)	0.488 (0.374)	0.275 (0.296)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.685*** (0.024)	0.746*** (0.033)	0.746*** (0.033)	0.746*** (0.033)	0.746*** (0.033)	0.746*** (0.033)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.062 (0.051)	0.917** (0.412)	1.169** (0.551)	1.289** (0.615)	0.654 (0.491)	0.369 (0.392)
N	2,733	1,442	1,442	1,442	1,442	1,442
	Running variable: Quadratic on both sides of cutoff					
Reduced form—scoring above reclassification threshold	0.064 (0.054)	0.701 (0.455)	0.897 (0.566)	1.264* (0.698)	0.417 (0.521)	0.214 (0.512)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.674*** (0.039)	0.713*** (0.057)	0.713*** (0.057)	0.713*** (0.057)	0.713*** (0.057)	0.713*** (0.057)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.096 (0.078)	0.983 (0.616)	1.258 (0.799)	1.774* (0.946)	0.585 (0.711)	0.300 (0.705)
N	2,733	1,442	1,442	1,442	1,442	1,442

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students in 10th grade within 10 scale score points of automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the top panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the bottom panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. The first stage of the 2SLS model predicts reclassification using a model identical in structure to the reduced-form model. Other than substitution of the predicted value of reclassification for the indicator of scoring above the reclassification threshold, the second stage of the 2SLS model is identical in structure to the reduced-form model.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

columns of the table present—for those students who take the ACT—the estimated effect on students' ACT score; it presents the estimated effect on the composite score, as well as on the score in each of the four subject areas. The table presents results from two model specifications, one where the flexible function of the running variable is specified as a linear term with different slopes on each side of the reclassification cutoff (top panel of the table) and a second where it is specified as a linear and quadratic term, each of which is allowed to have different slopes on each side of the reclassification threshold (bottom panel of table).

The reduced-form results in the first column of Table 3 demonstrate that scoring above the reclassification threshold has no significant effect on the probability of taking the ACT. The point estimates are broadly consistent across the two specifications—in the neighborhood of 0.05—but neither reaches statistical significance. The second column of Table 3, however, illustrates that scoring above the reclassification threshold has a positive effect on students' composite ACT score. In each specification, the reduced-form results demonstrate that scoring above the automatic reclassification cutoff is estimated to increase a student's composite ACT score by about 0.7 points. We note, however, that only the estimate from the model where the running variable is specified as linear on both sides is statistically significant at $p < 0.05$; the p-value for the estimate from the quadratic specification is 0.126.

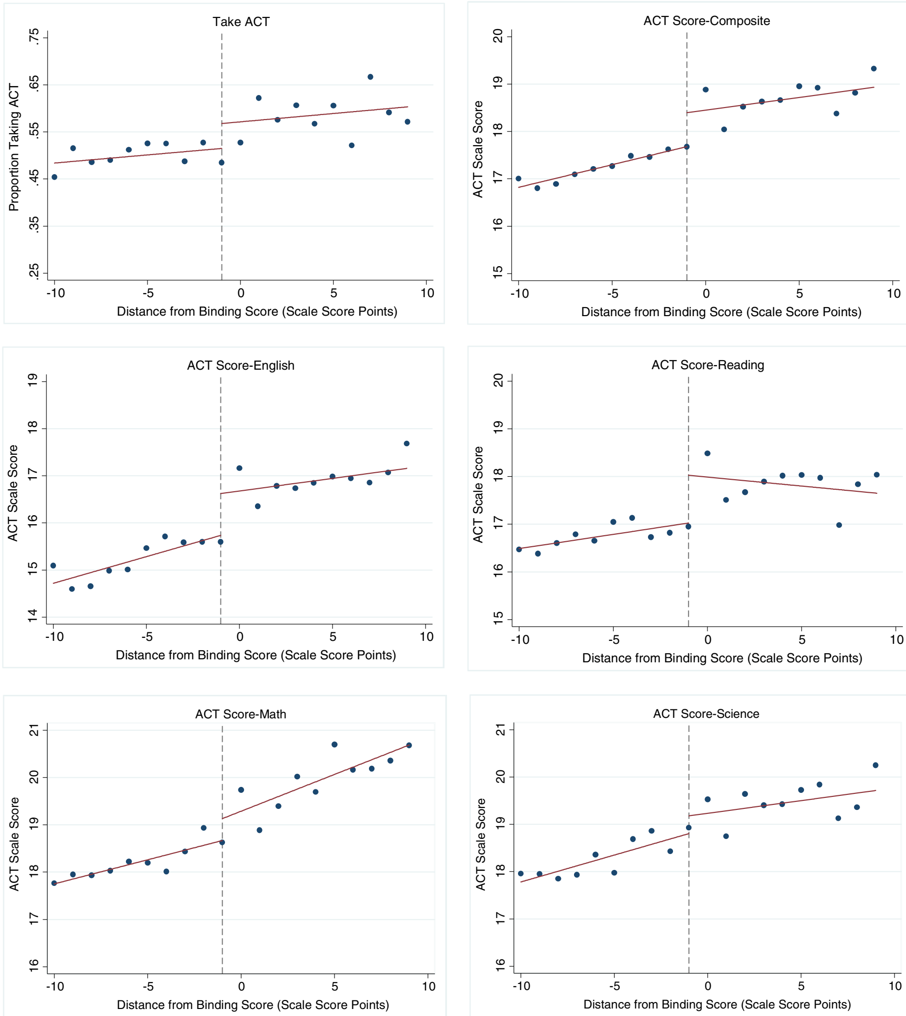
The results in the third and fourth columns of Table 3 indicate that the positive effect of reclassification on the composite ACT score is primarily attributable to improved performance of reclassified students on the English and reading portions of the assessment. In English, the results demonstrate that scoring above the reclassification threshold is estimated to increase a student's subscale score by a statistically significant 0.9 points. The estimated effects of reclassification on students' reading subscale score are even larger than the effects on English scores. Specifically, the reduced-form results indicate that scoring above the reclassification threshold is estimated to increase reading subscale scores by 0.96 to 1.26 points—depending on the model specification—with the estimates significant at $p < 0.05$ and $p < 0.10$, respectively. The fifth and sixth columns of Table 3 present the estimated effects of reclassification on students' scores on the math and science subscales. The results indicate that scoring above the reclassification threshold has no significant effect on scores in these subject areas, although the point estimates across both specifications and subjects are positive and generally in the range of one-quarter to one-half of a point.

The fact that the positive effect of reclassification on composite ACT scores largely operates through increased scores on the English and reading subscales lends a degree of face validity to the results. As we discussed previously, students who are classified as ELLs at the beginning of their 11th-grade year likely spend much of their reading/English instructional time receiving English language instruction while students who are classified as fully proficient are likely to spend that time in classes with native English-speaking peers receiving content instruction that is better aligned with college entrance exams (Callahan, Wilkinson, & Muller, 2010).

Visual corroboration of the reduced-form results in Table 3 is provided in Figure 3, which plots the mean of each outcome measure by the distance from the automatic reclassification threshold. Each panel also contains a separate line of best fit on each side of the automatic reclassification threshold. Figure A2 in Appendix A presents the plots in Figure 3 with a quadratic curve, rather than a line, fitted separately on each side of the automatic reclassification threshold.[24] The visual evidence is consistent with the results in Table 3, demonstrating significant discontinuities in students' ACT composite, English, and reading scores at the automatic reclassification cutoff.

As noted above, the reduced-form estimates represent the effect of scoring above the reclassification threshold—they represent the effect of increasing the probability of reclassification. Estimates that can be interpreted as the effect of reclassification per se are provided by the 2SLS results in Table 3. In each panel of the table, the first row of the 2SLS results presents the estimated coefficients for the measure of scoring above the reclassification threshold from the first-stage equation predicting reclassification. Consistent with the graphical evidence in Figure 1, the results

[24] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.



Notes: The markers in each panel of the figure represent the mean of the outcome variable at each scale score within 10 points of the automatic reclassification threshold on the ACCESS exam. Each panel also contains a line of best fit that is fitted separately on each side of the automatic reclassification cutoff.

Figure 3. Mean Proportion of Students Taking ACT and Mean ACT Scale Score, by Distance from Automatic Reclassification Threshold.

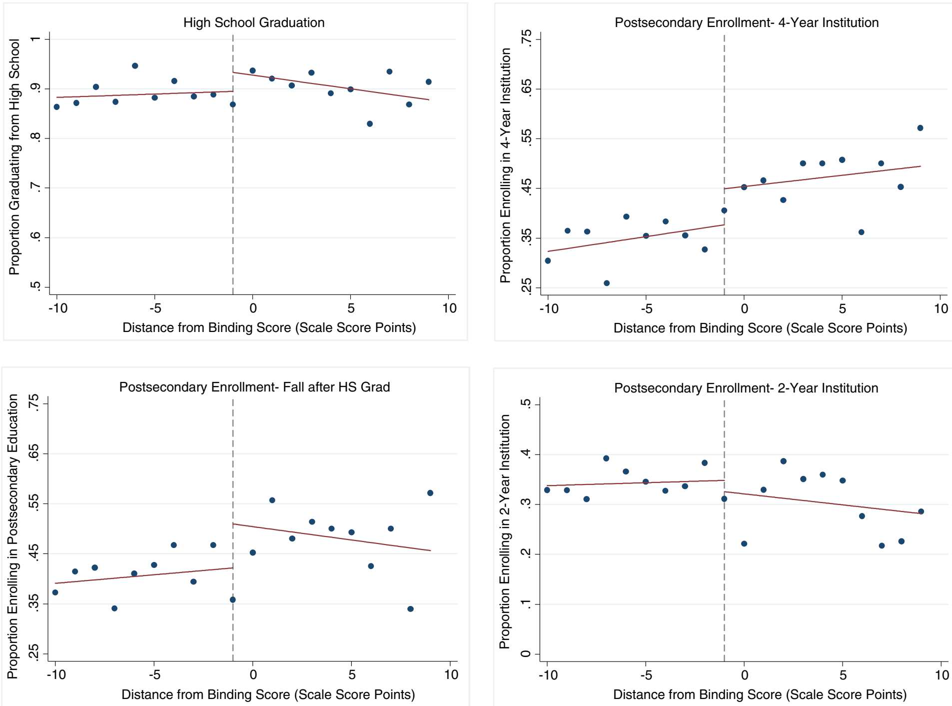
demonstrate a strong, significant relationship between scoring above the cutoff and being reclassified; scoring above the threshold is estimated to increase the probability of being reclassified by 0.70 to 0.75. The second row of the 2SLS results presents the estimated coefficients for the measure of reclassification, which has been instrumented with the indicator for scoring above the reclassification cutoff and thus represents the estimated LATE of reclassification. Complementing the reduced-form results, the LATE estimates reveal that the estimated positive effect of 10th-grade reclassification is about 1 point on students' composite ACT score and 1.2 to 1.7 points on the English and reading subscales—depending upon the particular specification—with the pattern of significance mirroring the reduced-form results.

Table 4. Coefficients and standard errors on measures of reclassification, by postsecondary-related outcome measure and model specification.

Model	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution
	Running variable: Linear on both sides of cutoff			
Reduced form—scoring above reclassification threshold	0.030 (0.025)	0.083* (0.045)	0.075 (0.046)	−0.027 (0.039)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.774*** (0.028)	0.774*** (0.028)	0.774*** (0.028)	0.774*** (0.028)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.039 (0.032)	0.108* (0.059)	0.097 (0.061)	−0.035 (0.050)
N	1,863	1,863	1,863	1,863
	Running variable: Quadratic on both sides of cutoff			
Reduced form—scoring above reclassification threshold	0.080** (0.033)	0.090 (0.065)	0.062 (0.071)	−0.059 (0.059)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.766*** (0.043)	0.766*** (0.043)	0.766*** (0.043)	0.766*** (0.043)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.104** (0.044)	0.118 (0.082)	0.081 (0.092)	−0.077 (0.077)
N	1,863	1,863	1,863	1,863

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students in 10th grade within 10 scale score points of automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the top panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the bottom panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. The first stage of the 2SLS model predicts reclassification using a model identical in structure to the reduced-form model. Other than substitution of the predicted value of reclassification for the indicator of scoring above the reclassification threshold, the second stage of the 2SLS model is identical in structure to the reduced-form model.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 3 provides strong evidence that 10th-grade reclassification has a significant positive effect on students' ACT performance. The question that naturally follows is whether positive effects of reclassification are also observed for other outcomes related to postsecondary attainment. Table 4 begins to address this question by presenting estimates of the effect of 10th-grade reclassification on high school graduation and multiple measures of postsecondary enrollment, including enrolling in a postsecondary institution the fall following high school graduation, ever enrolling in a four-year postsecondary institution by the end of 2013, and ever enrolling in a



Notes: The markers in each panel of the figure represent the mean of the outcome variable at each scale score within 10 points of the automatic reclassification threshold on the ACCESS exam. Each panel also contains a line of best fit that is fitted separately on each side of the automatic reclassification cutoff.

Figure 4. Mean Proportion of Students Graduating High School and Mean Proportion of Students Enrolling in Postsecondary Education, by Distance from Automatic Reclassification Threshold.

two-year institution by the end of 2013. As with the ACT-related outcomes, Figures 4 and A3 in Appendix A provide visual evidence to accompany the reduced-form results in Table 4.[25] These figures contain plots of the mean of each outcome variable by the distance from the automatic reclassification threshold, along with linear (Figure 4) and quadratic (Appendix Figure A3)[26] curves fitted on each side of the automatic reclassification cutoff.

The first column of Table 4 presents estimates of the effect of reclassification on graduating from high school, and although the point estimates are positive in both specifications of the reduced-form model, their magnitudes vary. In the first specification—where the flexible function of the running variable is specified as linear with different slopes on each side of the cutoff—scoring above the automatic reclassification threshold is estimated to increase the probability of high school graduation by an insignificant 0.03. In the second specification, however, the point estimate is 0.08 and significant at $p < 0.05$. Inspection of Figures 4 and A3 in

[25] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[26] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Appendix A[27] suggest that the quadratic specification is a better fit to the data, but we draw no definitive conclusions regarding the effect of reclassification on high school graduation.

The second column of Table 4 presents estimates of the effect of reclassification on enrolling in a postsecondary institution the fall after graduating from high school. Both specifications of the reduced-form model indicate that scoring above the reclassification threshold is estimated to increase the probability of postsecondary enrollment by 0.08 to 0.09, although only the first specification—where the running variable is specified as linear with different slopes on each side of the threshold—is significant at $p < 0.10$. The 2SLS results demonstrate that the corresponding LATE estimates range from 0.11 to 0.12 with the same pattern of significance as the reduced-form results.

The third and fourth columns of Table 4 are intended to gain preliminary insight into whether any estimated positive effect of reclassification on postsecondary enrollment is driven by increased enrollment in two-year institutions or four-year institutions, or perhaps both. Although the estimates are not statistically significant, the results suggest that reclassification may have a positive effect on ever attending a four-year institution and a null, or even slightly negative, effect on ever attending a two-year institution. These results are consistent with the positive effect of reclassification on postsecondary enrollment stemming primarily from increased enrollment at four-year institutions. In Appendix B, we demonstrate the robustness of all our results to alternative specifications, including different sample bandwidths and alternative estimators.[28] We also further assess the validity of our design through analysis of placebo cut scores.

INSIGHT ON POTENTIAL MECHANISMS: ANALYSIS OF DIFFERENT GRADES

To this point, our analyses indicate that 10th-grade reclassification has a positive effect on ACT scores, postsecondary enrollment, and perhaps high school graduation. However, the results provide little insight into the mechanisms potentially responsible for producing these effects. Here we perform a series of analyses that provide evidence on this issue, at least indirectly.[29]

Earlier we theorized that reclassification at the end of 10th grade may positively affect students' ACT scores and subsequent postsecondary enrollment because it puts them on a different educational track—one that places more emphasis on college preparation and counseling—than their peers who remain classified as ELLs. If such a mechanism is responsible for the observed effects, then we should not see similar effects of reclassification in ninth grade because students who fell just below the cutoff are likely to score above the threshold in 10th grade and thus still achieve reclassification prior to their pivotal junior year.

We provide evidence relevant to this conjecture in two ways. First, for each outcome analyzed above, we estimate both specifications of the reduced-form model in equation (3) over the sample of ninth graders who scored within 10 points of

[27] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[28] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[29] In addition, in Appendix C we present an analysis of the effect of reclassification on an additional outcome, content-area achievement. We estimate the effects of reclassification for students in three different grade ranges: (1) all grades; (2) grades 2 through 5; and (3) grades 6 through 9. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

the automatic reclassification threshold. Consistent with expectations, the results from these models, which are presented in Table 5, demonstrate that ninth-grade reclassification has no significant effect on student ACT scores or postsecondary enrollment outcomes. Such estimates are perhaps unsurprising, as our data indicate that over 60 percent of students who score just below the reclassification threshold in ninth grade achieve reclassification by the end of 10th grade. Although this number is smaller than the percentage of ninth graders scoring just above the cutoff who achieve reclassification by the end of 10th grade—over 90 percent do so—it reduces the disparity in 11th-grade ELL status to a level that renders the reduced-form estimates of scoring above the ninth-grade cutoff insignificant. We note, however, that the point estimates for postsecondary attendance and attending a four-year institution are positive, as are the estimates for ACT scores from the model where the running variable is specified as linear and quadratic terms on each side of the cutoff.

The takeaway from the first analysis is not that ninth-grade reclassification is unimportant, but rather that ninth-grade reclassification has no effect on students' postsecondary outcomes because the majority of students who just miss ninth-grade reclassification still end up classified as fully proficient prior to the pivotal junior year. Our second analysis provides further evidence on this issue by calculating—separately for each student—a variation of the running variable that we define as the maximum of students' ninth and 10th-grade ACCESS score, relative to the relevant automatic reclassification threshold. We then estimate the reduced-form and 2SLS models presented earlier over a sample of observations from the grade in which students achieved their maximum score. In doing so, we are effectively using our regression discontinuity approach to estimate the effect of being reclassified by the end of 10th grade, rather than in 10th grade as we do in our primary analysis. The results from this analysis are presented in Tables D1 (ACT-related outcomes) and D2 (postsecondary enrollment outcomes) in Appendix D.[30] Across all outcomes, the estimated effects in Tables D1 and D2 are remarkably similar in terms of sign and magnitude to the corresponding primary estimates in Tables 3 and 4. Moreover, because of the increased sample size due to the inclusion of ninth graders, the estimates in Tables D1 and D2 are more precise and thus exhibit higher levels of statistical significance. Together, these analyses provide additional, if indirect, evidence on the importance of students' ELL status at the beginning of their crucial junior year and hint at the roles that potential mechanisms such as exposure to college preparatory coursework or more general postsecondary counseling might play in generating the observed effects.

We use similar logic to develop conjectures regarding the effect of 11th-grade reclassification on postsecondary-related outcomes. In particular, we theorize that reclassification should have no effect on ACT scores because most students who will take the exam have already done so by that point. Similarly, for postsecondary enrollment outcomes, 11th-grade reclassification might be expected to have little effect because many students have established their postsecondary plans, at least generally. We test this proposition by estimating both specifications of the reduced-form model in equation (3) over the sample of 11th graders who scored within 10 points of the automatic reclassification threshold. The results are presented in Table 6. As expected, the results show that 11th-grade reclassification has no significant effect on student ACT scores. However, the postsecondary enrollment and high school graduation results exhibit inconsistency across model specifications.

[30] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 5. Coefficients and standard errors on measure of scoring above automatic reclassification threshold, by ACT subject score, grade of reclassification, and model specification: Ninth grade.

	ACT scores					Postsecondary attainments			
	Composite	English	Reading	Math	Science	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution
	Running variable: Linear on both sides of cutoff								
Scoring above reclassification threshold	−0.005 (0.208)	0.078 (0.393)	−0.221 (0.505)	0.108 (0.273)	−0.010 (0.273)	−0.003 (0.018)	0.030 (0.031)	0.015 (0.032)	−0.032 (0.030)
N	1,512	1,512	1,512	1,512	1,512	2,525	2,525	2,525	2,525
	Running variable: Quadratic on both sides of cutoff								
Scoring above reclassification threshold	0.405 (0.367)	0.480 (0.571)	0.096 (0.682)	0.915 (0.661)	0.059 (0.496)	0.025 (0.026)	0.037 (0.051)	0.009 (0.058)	−0.071 (0.051)
N	1,512	1,512	1,512	1,512	1,512	2,525	2,525	2,525	2,525

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within 10 points of the automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the top panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the bottom panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 6. Coefficients and standard errors on measure of scoring above automatic reclassification threshold, by ACT subject score, grade of reclassification, and model specification: 11th grade.

	ACT scores					Postsecondary attainments			
	Composite	English	Reading	Math	Science	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution
	Running variable: Linear on both sides of cutoff								
Scoring above reclassification threshold	−0.156	−0.299	−0.226	−0.107	−0.016	−0.033	−0.023	0.068	−0.075
	(0.278)	(0.357)	(0.375)	(0.374)	(0.417)	(0.028)	(0.059)	(0.059)	(0.048)
N	1,223	1,223	1,223	1,223	1,223	1,526	1,526	1,526	1,526
	Running variable: Quadratic on both sides of cutoff								
Scoring above reclassification threshold	−0.273	−0.637	−0.065	0.056	−0.361	−0.079*	−0.156*	0.028	−0.215***
	(0.545)	(0.623)	(0.564)	(0.621)	(0.684)	(0.042)	(0.093)	(0.075)	(0.076)
N	1,223	1,223	1,223	1,223	1,223	1,526	1,526	1,526	1,526

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within 10 points of the automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the top panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the bottom panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff.

***$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

In particular, 11th-grade reclassification is estimated to have no significant effect on these outcomes when the running variable is specified as linear on each side of the reclassification threshold. When the running variable is specified as linear and quadratic terms on each side of the cutoff, though, the results suggest a negative and significant effect of reclassification on high school graduation and postsecondary enrollment.

Even recognizing the inconsistency across model specifications, these results warrant further discussion given our a priori expectations. The results in Table 6 provide strong evidence that the negative effect of 11th-grade reclassification on postsecondary enrollment is driven by reduced attendance at two-year institutions. This result is generally consistent with the negative point estimates for the effect of reclassification on ever attending a two-year institution across the other two grades we analyze. However, we highlight two notable differences between the 11th grade result and those for ninth and 10th grade. First, the magnitude of the significant negative point estimate is notably larger in 11th grade than in ninth or 10th grade. Second, and perhaps more importantly, the negative point estimates for attendance at a two-year institution are not offset or exceeded by a positive point estimate for attendance at a four-year institution, as they are in ninth and 10th grade.[31] The fact that 11th-grade reclassification has little effect on attendance at a four-year institution is perhaps unsurprising as the process likely occurs too late in a student's high school career for the mechanisms by which reclassification might increase attendance at a four-year institution—enrollment in courses that improve ACT performance or exposure to postsecondary preparation activities, such as financial aid forms or the application process—to operate as they do in the context of 10th grade.

Finally, given that our results suggest 10th-grade reclassification to have a positive effect on both ACT scores and postsecondary enrollment, it is possible that the increases in postsecondary enrollment are primarily—or even entirely—a product of improved ACT performance. To examine this possibility, we estimate a variant of the reduced-form model in equation (3) where we specify the postsecondary enrollment measures as the outcomes and include a student's composite ACT score as a covariate. If the positive effects of reclassification on postsecondary enrollment were operating entirely through increased ACT scores, then the estimated effects from this model should be close to zero. The results from this model, which are presented in Table D3 in Appendix D,[32] demonstrate that the estimated effects are indeed attenuated compared to the primary results, but they still remain positive and meaningfully above zero, if insignificant.[33] This suggests that the effects of reclassification on postsecondary enrollment are not just operating through improved ACT performance. Other mechanisms—perhaps college counseling or other postsecondary preparation activities—also seem likely to be contributing to the observed effects.

[31] Recall that the measurement of the postsecondary enrollment variable measures whether a student enrolled in any postsecondary institution the fall following high school graduation while the measures of attendance at two and four years indicate whether a student was ever recorded doing so.
[32] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[33] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

DISCUSSION AND CONCLUSION

The rapid growth of ELL populations across the country is focusing significant attention on reclassification policy, which governs the process by which ELLs move toward—and are deemed to reach—full English proficiency. In this paper, we exploit Wisconsin's policy rule that automatically reclassifies a student as fully English proficient if he scores above a specified threshold on the ACCESS exam to identify the causal effect of reclassification. Our analysis indicates that being reclassified as fully English proficient during 10th grade has a positive effect on students' ACT scores and postsecondary enrollment. The results also provide suggestive evidence of a positive effect of 10th-grade reclassification on high school graduation. We demonstrate that the positive effect of reclassification on students' composite ACT score is primarily attributable to improved performance on the English and reading portions of the test. With respect to postsecondary enrollment, our analyses provide evidence that the positive effect of reclassification stems from increased enrollment at four-year institutions. These results have several implications for both research and policy.

At a basic level, our results suggest the importance of a student's ELL classification at the beginning of his or her junior year, at least for students who scored near the automatic reclassification threshold on the 10th-grade ACCESS exam. We theorize that the positive effects of scoring above the reclassification cutoff are attributable to students being exposed to different college preparation activities and resources than their peers who scored just below the cutoff and remain classified as ELLs—ACT preparation, college counseling, and assistance with the application or financial aid process are potential examples—and we provide indirect evidence in support of this explanation. And although additional work should further explore potential mechanisms at work, such evidence suggests that school and district personnel may do well to ensure that they provide the same postsecondary-related resources to students who score just below the reclassification threshold as they do to students scoring on the other side of that cutoff.

Given the heightened consequence of our sample members' ELL classification during their junior year, schools and districts may also do well to more closely consider whether to manually reclassify 10th-grade students who score a point or two below the automatic reclassification threshold. Figure 1 demonstrates that less than one-quarter of students who score just below the automatic reclassification threshold subsequently lose their ELL status but over 90 percent of students who score above the cutoff are reclassified as fully proficient. In light of the effects of 10th-grade reclassification on postsecondary outcomes, as well as other outcomes that future work could examine, it is reasonable to question whether a single scale score point on an English proficiency assessment should exert such a dramatic effect on the likelihood of being classified as an ELL.

At the same time, the estimated effects of 11th-grade reclassification on postsecondary enrollment may provide some cause for concern. Although the estimates exhibit inconsistency across model specifications and potential mechanisms are not fully clear, the results provide at least some evidence that 11th-grade reclassification reduces postsecondary attendance, particularly in two-year institutions. Additional work should examine whether these suggestive results are replicated in other contexts and, if they are, greater attention should be given to ensuring that reclassification does not limit postsecondary options, irrespective of the timing at which it occurs.

In having this whole discussion, we reiterate that the limited generalizability of our LATE estimates is a disadvantage when it comes to considering policy implications. As with LATEs in fuzzy regression discontinuity analyses more generally, our estimates are only generalizable to those students in a narrow band surrounding the automatic reclassification threshold for whom scoring above the cutoff would have resulted in reclassification as fully English proficient. Moreover, given our focus on students in high school grades—10th grade, in particular—near the reclassification cutoff, our sample is disproportionately composed of ELLs who have been in the school system for a relatively long period of time, compared to the broader ELL population in Wisconsin. Indeed, a full 40 percent of our sample has been in the United States for more than six years while only about 20 percent of the broader ELL population has been in the country for that length of time. Correspondingly, only about a quarter of our sample have been in the United States for less than four years; over half of the broader ELL student population in Wisconsin are relatively recent arrivals. It is perhaps unsurprising that relatively recent immigrants are underrepresented in our sample of ELLs, as it typically takes several years for students to score near the proficiency threshold on exams like the ACCESS. Assessing whether our estimated effects of reclassification are generalizable to a broader set of ELLs would certainly provide additional useful information to policymakers, but would require a wholly different research design from the one we employ here.

When interpreting the results in this paper, it is important to recognize that Wisconsin is a relatively uncommon setting for analyzing policies governing the education of ELLs. Our use of this context has implications for both the generalizability of our results and their comparability to the existing literature—we highlight two specific features of the Wisconsin context relevant to these considerations. First, the substantial growth in Wisconsin's ELL population has largely occurred since the turn of the century, resulting in the state only recently beginning to devote meaningful attention to policies governing ELL education. In light of this, our results may provide valuable information to states that have exhibited growth in ELL students similar to that in Wisconsin, but are unlikely to be as relevant to states with well-established ELL populations and policies, such as California, Texas, or Arizona. Second, the ELL population in Wisconsin is spread across all areas of the state—urban, suburban, and rural alike—and tends not to be concentrated in a small number of schools. This stands in stark contrast to the context in which ELLs are typically studied, one where students are concentrated in a small number of schools in urban areas (Kanno & Kangas, 2014). Moreover, our analyses draw on statewide data whereas much previous work relies on data from a single district (e.g., Robinson, 2011) or even school (e.g., Kanno & Kangas, 2014). Together, these features of the setting underlying our analysis suggest that conclusions from existing work regarding the postsecondary preparation of ELLs should be applied with caution to the Wisconsin context, and vice versa.

As the ELL population across the country continues to grow, policies surrounding their education are likely to attract significant attention from policymakers and educators. To this point, analysis of these policies has been relatively limited, but by estimating the effects of reclassifying students as fully English proficient this paper improves our understanding of one dimension of policy relevant to ELL education.

DEVEN CARLSON is an Assistant Professor of Political Science at the University of Oklahoma, 205 Dale Hall Tower, 455 W. Lindsey Street, Norman, OK 73019 (e-mail: decarlson@ou.edu).

JARED E. KNOWLES is a Research Analyst at the Wisconsin Department of Public Instruction, P.O. Box 7841, Madison, WI 53707–7841 (e-mail: jknowles@gmail.com).

ACKNOWLEDGMENTS

We thank the Wisconsin Department of Public Instruction for providing access to the data we use in this paper and for sharing their expertise on program administration in Wisconsin. We also thank the editor, the co-editor, and three anonymous reviewers for comments and suggestions that greatly improved the quality of this paper.

REFERENCES

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. Educational Measurement: Issues and Practice, 27, 17–31.

Avery, C. (2013). Evaluation of the College Possible program: Results from a randomized controlled trial. Cambridge, MA: National Bureau of Economic Research.

Callahan, R., & Gándara, P. (2004). On nobody's agenda: Improving English-language learners' access to higher education. In M. Sadowski (Ed.), Teaching immigrant and second language students: Strategies for success (pp. 107–127). Cambridge, MA: Harvard Education Press.

Callahan, R., Wilkinson, L., & Muller, C. (2010). Academic achievement and course taking among language minority youth in U.S. schools: Effects of ESL placement. Educational Evaluation and Policy Analysis, 32, 84–117.

Cannon, J. S., Jacknowitz, A., & Painter, G. (2011). The effect of attending full-day kindergarten on English learner students. Journal of Policy Analysis and Management, 30, 287–309.

Cave, G., & Quint, J. (1990). Career beginnings impact evaluation: Findings from a program for disadvantaged high school students. New York, NY: MDRC.

Chin, A., Daysal, N. M., & Imberman, S. A. (2013). Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas. Journal of Public Economics, 107, 63–78.

Conger, D., Long, M. C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. Journal of Policy Analysis and Management, 28, 555–576.

Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2013). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. Cambridge, MA: National Bureau of Economic Research.

Fazekas, A., & Warren, C. (2011). Building a pathway to the future: Maximizing high school guidance and advisory support. Washington, DC: U.S. Department of Education.

Flores, E., Painter, G., Harlow-Nash, Z., & Pachon, H. (2009). ¿Qué pasa? Are ELL students remaining in English learning classes too long? Los Angeles, CA: Tomás Rivera Policy Institute.

Greene, J. P. (1997). A meta-analysis of the Rossell and Baker review of bilingual education research. Bilingual Research Journal, 21, 103–122.

Grissom, J. B. (2004). Reclassification of English Learners. Education Policy Analysis Archives, 12. Retrieved March 13, 2014, from http://epaa.asu.edu/epaa/v12n36/ 1–38.

Hill, L. E., Weston, M., & Hayes, J. M. (2014). Reclassification of English learner students in California. San Francisco, CA: Public Policy Institute of California.

Howell, J. S., Kurlaender, M., & Grodsky, E. (2010). Postsecondary preparation and remediation: Examining the effect of the Early Assessment Program at California State University. Journal of Policy Analysis and Management, 29, 726–748.

Jepsen, C., & de Alth, S. (2005). English learners in California schools. San Francisco, CA: Public Policy Institute of California.

Kanno, Y., & Kangas, S. E. (2014). "I'm not going to be, like, for the AP" English language learners' limited access to advanced college-preparatory courses in high school. American Educational Research Journal, 51, 848–878.

Kim, J. (2011). Relationships among and between ELL status, demographic characteristics, enrollment history, and school persistence. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California.

Kim, J., & Herman, J. L. (2010). When to exit ELL students: Monitoring success and failure in mainstream classrooms after ELLs' reclassification. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California.

Mahoney, K. S., & MacSwan, J. (2005). Reexamining identification and reclassification of English language learners: A critical discussion of select state practices. Bilingual Research Journal, 29, 31–42.

Matsudaira, J. D. (2005). Sinking or swimming? Evaluating the impact of English immersion versus bilingual education. Working paper.

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

Parrish, T. B., Merickel, A., Perez, M., Linquanti, R., Socias, M., Spain, A., Speroni, C., Esra, P., Brock, L., & Delancey, D. (2006). Effects of the Implementation of Proposition 227 on the Education of English learners, K-12: Findings from a Five-Year Evaluation. Final Report for AB 56 and AB 1116. Washington, DC: American Institutes for Research; San Francisco, CA: WestEd.

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. Educational Evaluation and Policy Analysis, 33, 267–292.

Robinson-Cimpian, J. P., & Thompson, K. D. (2015). The effects of changing test-based policies for reclassifying English learners. Journal of Policy Analysis and Management, 35, 279–305.

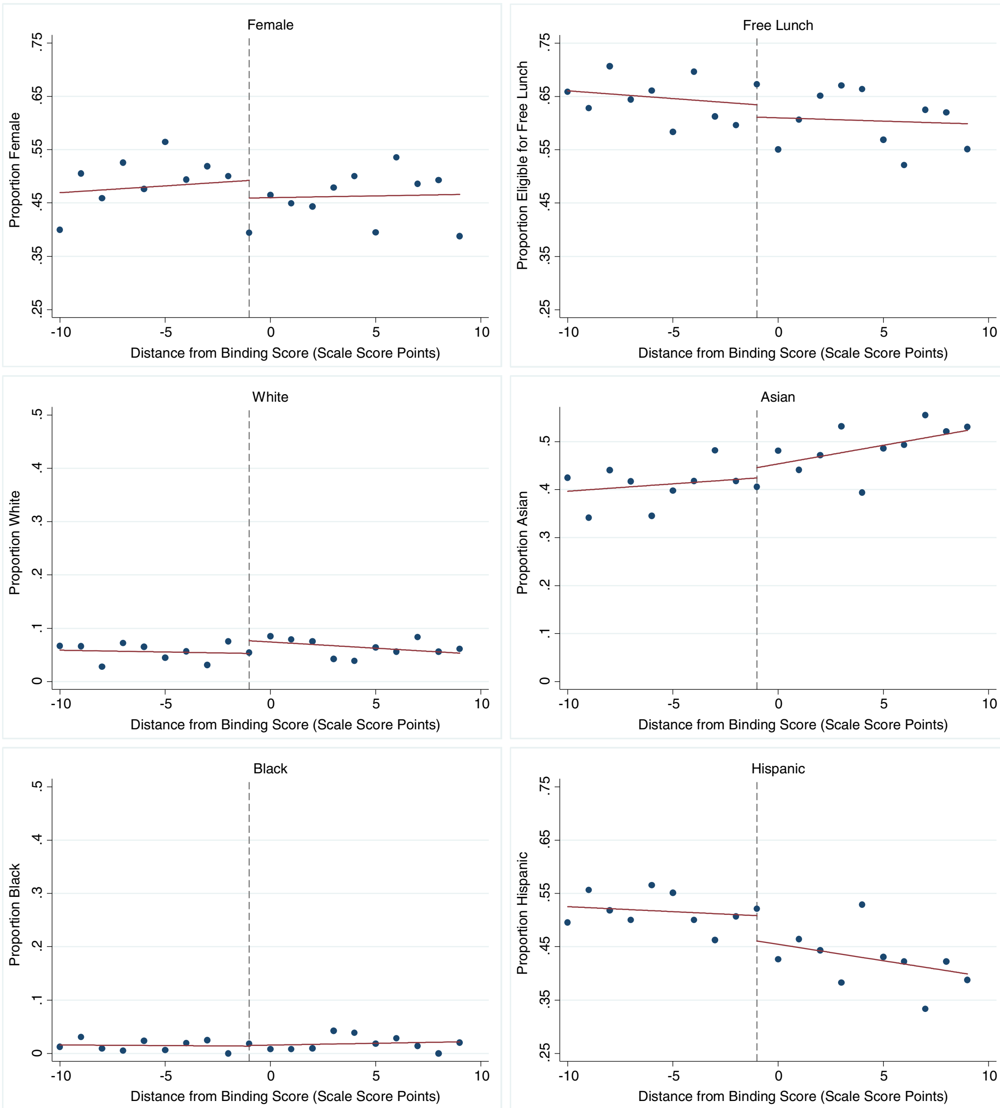
Schiller, K. S., & Muller, C. (2000). External examinations and accountability, educational expectations, and high school graduation. American Journal of Education, 108, 73–102.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. Review of Educational Research, 75, 247–284.

Snyder, T. D., & Dillow, S. A. (2013). Digest of Education Statistics 2012 (NCES 2014–015). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

U.S. Department of Education. (2012). National evaluation of Title III implementation—Report on state and local implementation. Washington, DC: Policy and Program Studies Service, Office of Planning, Evaluation, and Policy Development, U.S. Department of Education. Wisconsin State Statute 115.95-115.99.

APPENDIX A: Additional Figures for Validity Checks and Primary Results



Notes: The markers in each panel of the figure represent the mean of the observable characteristic at each scale score within 10 points of the automatic reclassification threshold on the ACCESS exam. Each panel also contains a line of best fit that is fitted separately on each side of the automatic reclassification cutoff.

Figure A1. Mean Observable Student Background Characteristics, by Distance from Automatic Reclassification Threshold.

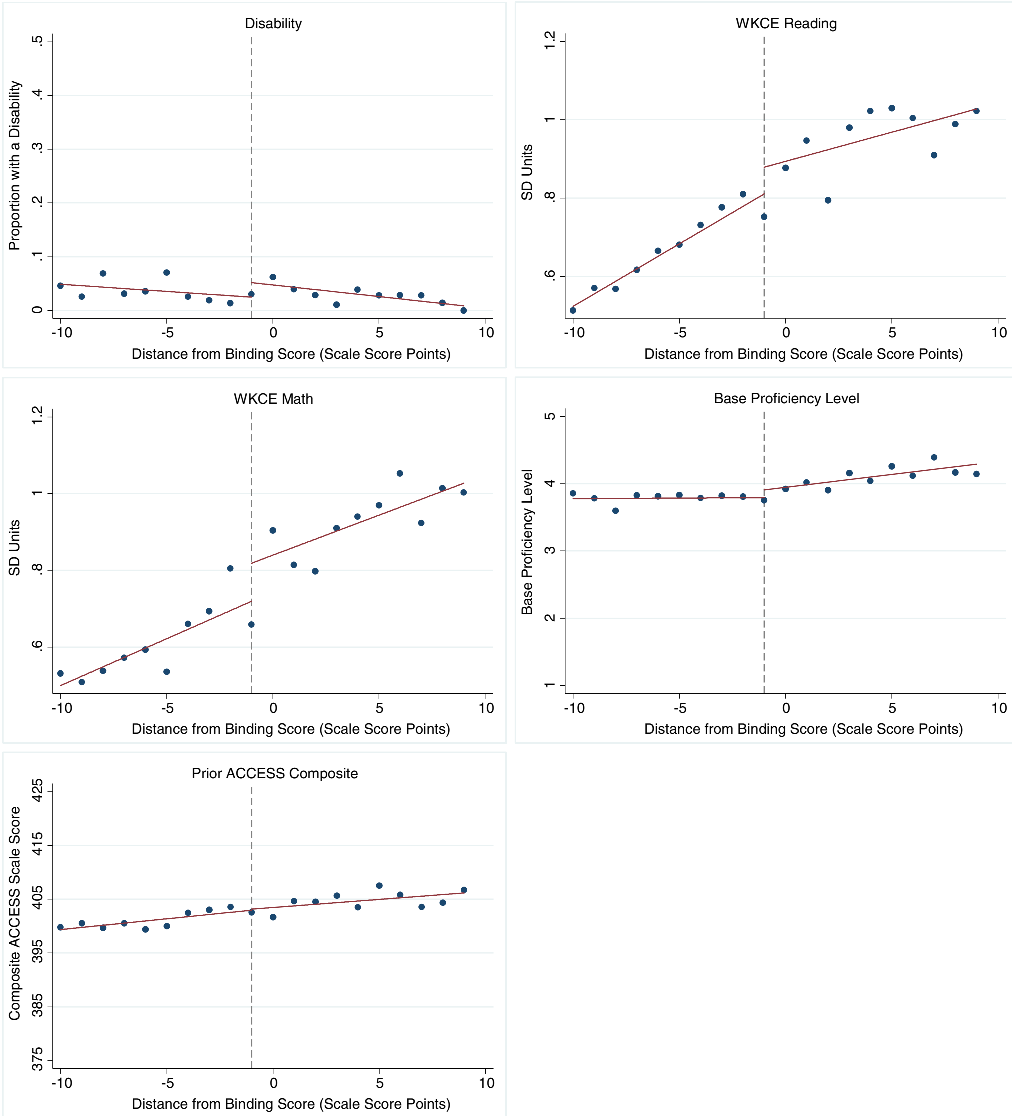
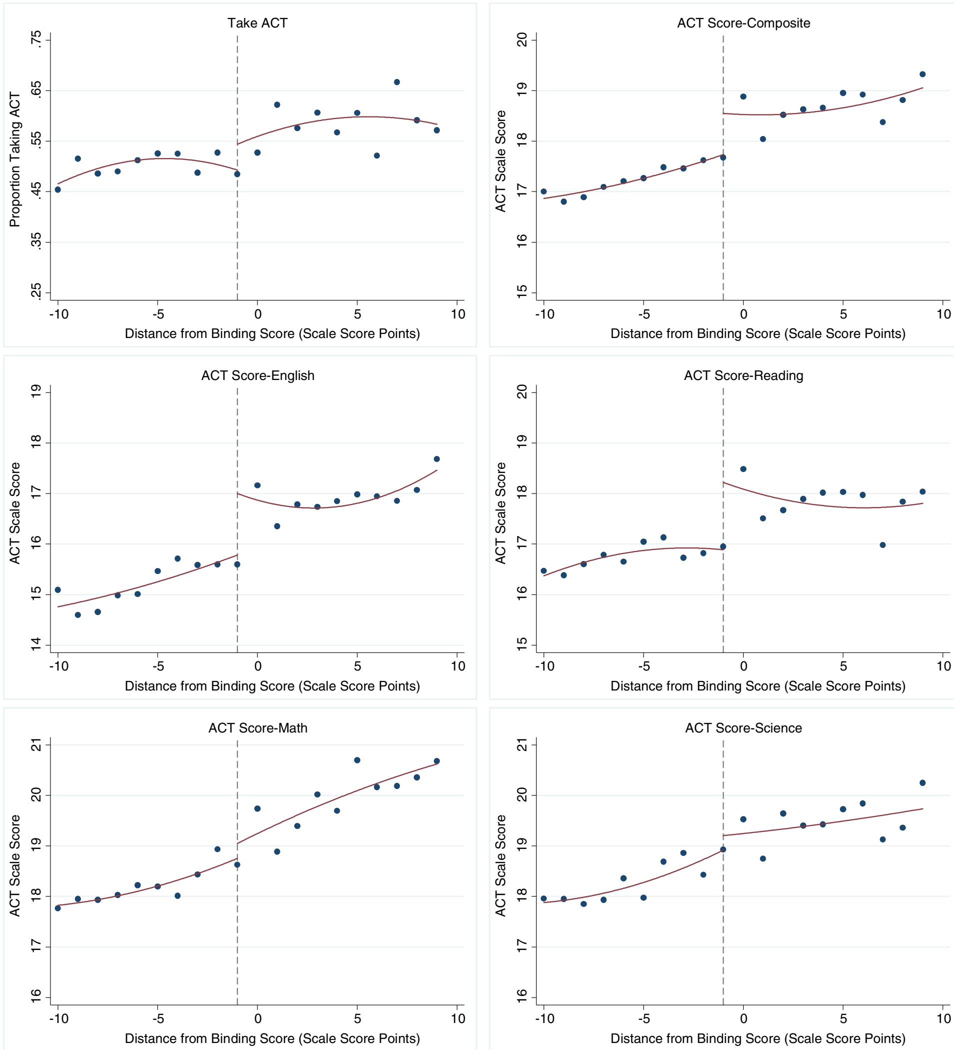
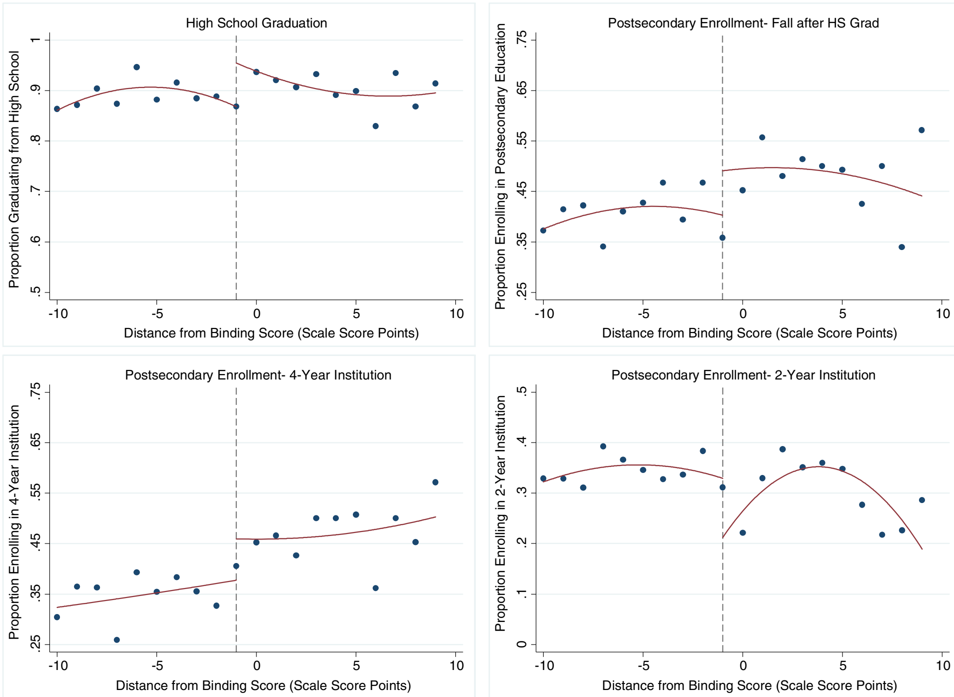


Figure A1. *Continued.*



Notes: The markers in each panel of the figure represent the mean of the outcome variable at each scale score within 10 points of the automatic reclassification threshold on the ACCESS exam. Each panel also contains a quadratic function fitted separately on each side of the automatic reclassification cutoff.

Figure A2. Mean Proportion of Students Taking ACT and Mean ACT Scale Score, by Distance from Automatic Reclassification Threshold.



Notes: The markers in each panel of the figure represent the mean of the outcome variable at each scale score within 10 points of the automatic reclassification threshold on the ACCESS exam. Each panel also contains a quadratic function fitted separately on each side of the automatic reclassification cutoff.

Figure A3. Mean Proportion of Students Graduating High School and Mean Proportion of Students Enrolling in Postsecondary Education, by Distance from Automatic Reclassification Threshold.

APPENDIX B: Additional Specifications and Robustness Checks

Based on a regression discontinuity design, our primary analysis indicates that being reclassified as fully English proficient during 10th grade has a positive effect on multiple outcomes related to students' postsecondary attainments, including ACT scores and postsecondary enrollment. In this section, we assess the robustness of these findings to alternative analytic choices.

Alternative Sample Bandwidths, Different Estimators, and Placebo Cut Scores

To ensure that our results are not driven by the choice of sample bandwidth, we estimated the two specifications of the reduced-form models presented in equation (3) over samples of several different bandwidths. Specifically, we estimated the models over a sample of students scoring within 4 points of the reclassification threshold—the minimum bandwidth suggested by Schochet et al. (2010)—as well as over samples of students scoring within 6, 8, and 12 points of the cutoff. Tables B1 and B2 present the results of this sensitivity analysis and, across all bandwidths and model specifications, these additional results are remarkably consistent with our primary results presented in Tables 3 and 4.

In addition to assessing the sensitivity of our primary results to different sample bandwidths, we also gauge their robustness to an alternative estimator. In

Table B1. Coefficients and standard errors on measure of scoring above automatic reclassification threshold, by ACT subject score, sample bandwidth, and model specification.

	Running variable: Linear on both sides of cutoff					Running variable: Quadratic on both sides of cutoff				
	Composite	English	Reading	Math	Science	Composite	English	Reading	Math	Science
	Bandwidth of ±4									
Scoring above reclassification threshold	0.883*	1.295**	1.373**	0.355	0.592	0.871	0.835	0.853	1.667*	0.230
	(0.457)	(0.628)	(0.660)	(0.559)	(0.504)	(0.789)	(1.172)	(1.118)	(0.996)	(0.842)
N	583	583	583	583	583	583	583	583	583	583
	Bandwidth of ±6									
Scoring above reclassification threshold	0.712*	0.958*	1.055*	0.395	0.318	0.832	1.384*	1.386	0.444	0.414
	(0.392)	(0.518)	(0.569)	(0.455)	(0.426)	(0.607)	(0.804)	(0.846)	(0.782)	(0.595)
N	876	876	876	876	876	876	876	876	876	876
	Bandwidth of ±8									
Scoring above reclassification threshold	0.703**	0.835**	1.121***	0.462	0.228	0.769	1.250*	1.187	0.380	0.323
	(0.344)	(0.418)	(0.535)	(0.415)	(0.353)	(0.519)	(0.697)	(0.721)	(0.596)	(0.586)
N	1,162	1,162	1,162	1,162	1,162	1,162	1,162	1,162	1,162	1,162
	Bandwidth of ±12									
Scoring above reclassification threshold	0.533**	0.677*	0.547	0.491	0.263	0.800*	1.037*	1.342**	0.494	0.256
	(0.259)	(0.348)	(0.397)	(0.319)	(0.267)	(0.443)	(0.537)	(0.632)	(0.517)	(0.459)
N	1,685	1,685	1,685	1,685	1,685	1,685	1,685	1,685	1,685	1,685

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within the bandwidth specified in the table. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the left panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the right panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B2. Coefficients and standard errors on measure of scoring above automatic reclassification threshold, by postsecondary-related outcome, sample bandwidth, and model specification.

	Running variable: Linear on both sides of cutoff				Running variable: Quadratic on both sides of cutoff			
	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution
	Bandwidth of ±4							
Scoring above reclassification threshold	0.054 (0.039)	0.104 (0.075)	0.038 (0.077)	−0.106 (0.069)	0.034 (0.084)	0.084 (0.154)	−0.130 (0.157)	−0.066 (0.122)
N	756	756	756	756	756	756	756	756
	Bandwidth of ±6							
Scoring above reclassification threshold	0.068** (0.030)	0.090 (0.056)	0.071 (0.062)	−0.052 (0.053)	0.037 (0.050)	0.113 (0.097)	−0.027 (0.104)	−0.127 (0.090)
N	1,111	1,111	1,111	1,111	1,111	1,111	1,111	1,111
	Bandwidth of ±8							
Scoring above reclassification threshold	0.050* (0.027)	0.084* (0.049)	0.080 (0.051)	−0.028 (0.041)	0.067* (0.039)	0.084 (0.083)	0.021 (0.082)	−0.087 (0.075)
N	1,474	1,474	1,474	1,474	1,474	1,474	1,474	1,474
	Bandwidth of ±12							
Scoring above reclassification threshold	0.033 (0.026)	0.052 (0.041)	0.059 (0.044)	−0.023 (0.037)	0.052* (0.030)	0.101* (0.060)	0.077 (0.065)	−0.063 (0.053)
N	2,204	2,204	2,204	2,204	2,204	2,204	2,204	2,204

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within the bandwidth specified in the table. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the left panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the right panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B3. Coefficients and standard errors for indicator of scoring above automatic reclassification threshold from nonparametric RD estimator, by outcome measure.

Outcome	Coefficient	Standard error	*N* left of cutoff	*N* right of cutoff	Bandwidth
ACT composite score	0.796*	0.445	687	475	13.8
ACT English score	1.222**	0.596	581	427	12.7
ACT reading score	1.389**	0.649	581	427	11.5
ACT math score	0.325	0.552	788	517	14.2
ACT science score	0.308	0.432	986	574	17.6
Graduate from high school	0.080**	0.041	781	512	12.0
Postsecondary enrollment—fall	0.094	0.069	916	558	12.0
Postsecondary enrollment—four years	0.066	0.069	781	512	11.8
Postsecondary enrollment—two years	−0.076	0.059	916	558	14.8

Notes: Coefficients and standard errors from nonparametric RD estimator proposed by Calonico, Cattaneo, and Titiunik (2014) presented in first and second columns of table. Estimates based on analytic samples containing students within bandwidth selected by method proposed in Calonico, Cattaneo, and Titiunik (2014). Selected bandwidth reported in the fifth column of the table. Third and fourth columns of the table report the number of cases to the left and right of the cutoff, respectively, that inform the estimates. Estimates generated using local linear regression with a triangular kernel and a local quadratic regression used to construct the bias correction. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

particular, we employ the nonparametric local polynomial estimator proposed by Calonico, Cattaneo, and Titiunik (2014) to generate reduced-form estimates of the effects of 10th-grade reclassification on ACT scores, high school graduation, and postsecondary enrollment. Using a bias-corrected regression discontinuity estimator along with a new approach to calculating standard errors, Calonico, Cattaneo, and Titiunik's (2014) estimator is designed to provide confidence intervals robust to the selection of large bandwidths commonly produced by standard bandwidth selection techniques.[34] Table B3 presents the estimated effects of reclassification from the nonparametric framework described above. The results in Table B3 are again remarkably similar to our primary results presented in Tables 3 and 4 in sign, significance, and magnitude.

Finally, if our regression discontinuity design is valid, we should not observe systematic changes in the probability of reclassification at other values of the running variable. In order to test this, we first created six placebo cut scores—3, 5, and 7 points below the true threshold and 3, 5, and 7 points above the true reclassification cutoff. We then estimated the 2SLS models presented in equations (4) and (5) to assess the probability of reclassification at other values of the running variable.[35]

[34] In our application, we estimate the effects of reclassification using local linear regression with a local quadratic regression used to construct the bias correction. We use the bandwidth selection procedure proposed by Calonico, Cattaneo, and Titiunik (2014).

[35] We estimated these models over samples containing students within 10 points of the placebo cutoff on the side unaffected by the true reclassification threshold. On the side of the placebo cutoff that runs up against the true cutoff, we include all students up to the value of the placebo running variable where the true reclassification threshold sits. For example, in our analysis of the placebo cutoff 5 points below the true cutoff, we include all students whose scores are within 10 points of the threshold on the side below the placebo cutoff. On the side above the placebo cutoff, however, we are only able to include students whose scores are within 5 points of the placebo threshold. Including students with higher scores would result in the inclusion of students who scored above the true reclassification cutoff, and would thus invalidate the results of this analysis. In all models, the flexible function of the placebo running variable is specified as linear and quadratic terms that are each allowed to have different slopes on each side of the placebo cutoff.

Table B4. Coefficients and standard errors from 2SLS model using placebo cutoffs, by outcome measure and value of placebo cutoff.

	3 points below true cutoff	5 points below true cutoff	7 points below true cutoff	3 points above true cutoff	5 points above true cutoff	7 points above true cutoff
	Composite ACT score					
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	−0.047 (0.058)	0.014 (0.058)	−0.026 (0.058)	−0.096 (0.097)	0.084* (0.049)	0.030 (0.042)
2SLS—2nd stage—effect of reclassification on standardized achievement	1.260 (8.151)	−21.548 (123.708)	−12.408 (29.537)	17.955 (24.102)	−11.491 (10.428)	−31.314 (47.769)
N	1,168	1,344	1,559	627	671	714
	Postsecondary enrollment—fall after high school graduation					
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	−0.004 (0.045)	−0.024 (0.041)	−0.051 (0.039)	−0.100 (0.082)	0.008 (0.042)	−0.039 (0.047)
2SLS—2nd stage—effect of reclassification on standardized achievement	16.737 (209.703)	−1.178 (2.824)	1.497 (1.737)	−2.415 (1.978)	2.127 (15.811)	−1.191 (2.511)
N	2,405	2,889	3,393	1,048	1,115	1,172

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within 10 points of the placebo cutoff on the side unaffected by the true reclassification threshold. On the side of the placebo cutoff that runs up against the true cutoff, we include all students up to the value of the placebo running variable where the true reclassification threshold sits. The first stage of the 2SLS models predicts reclassification using a model containing an indicator for scoring above the automatic reclassification threshold as well as measures of sex, disability status, eligibility for receipt of subsidized lunch, grade, and school year. The models control for distance from the reclassification threshold using a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. In the second stage, scoring above the automatic reclassification threshold is used as an instrument for being reclassified. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Results from estimation of the 2SLS models are presented in Table B4 and they demonstrate that, consistent with expectations, scoring above the placebo cutoffs had no systematic effect on the likelihood of reclassification. These null results are quite consistent across each of the six placebo cutoffs and each of the three sets of outcomes we analyze.

APPENDIX C: Effects of Reclassification on Content-Area Achievement

Although content-area achievement is not the primary focus of this paper, our data permit us to estimate the effects of reclassification on content-area achievement and we do so using the following reduced-form model:

$$Y_{it+1} = f(A_{it}) + \delta H_{it} + \boldsymbol{X_{it}\beta} + \varepsilon_{it} \tag{C.1}$$

where Y represents our outcome of interest—reading or math achievement—and i and t index students and time, respectively. In this model, $f(A_{it})$ is a flexible function of the distance in ACCESS scale score points from the automatic reclassification

Table C1. Sample means for observations within 10 scale score points of reclassification threshold, 2006–2007 to 2011–2012.

Variable	All observations	Observations below reclassification threshold	Observations above reclassification threshold
		All grades	
Female	0.528	0.531	0.523
Free lunch	0.586	0.602	0.554
Reduced-price lunch	0.135	0.136	0.132
No subsidized lunch	0.279	0.262	0.313
Asian	0.333	0.324	0.352
Black	0.014	0.014	0.014
Hispanic	0.558	0.574	0.527
White	0.091	0.084	0.105
Other race	0.004	0.004	0.002
Disability	0.025	0.026	0.024
Reclassified	0.376	0.111	0.912
Standardized reading score	0.850	0.765	1.030
Standardized math score	0.817	0.725	1.013
N	17,027	11,393	5,634

Notes: Because the WKCE is not administered in every grade, the *N*s for the standardized reading score and standardized math score variables are lower than the *N*s for the other measures in the table. Specifically, the respective *N*s for the standardized reading and math scores are 12,037 and 12,046 for the full sample, 8,166 and 8,178 for observations below the reclassification threshold, and 3,871 and 3,868 for observations above the reclassification threshold.

threshold, H is an indicator for scoring above the threshold, $\boldsymbol{X}$ is a vector of student background characteristics, and ε is the error term.[36] Below we present results from models where $f(A_{it})$ is specified as a linear term with different slopes below and above the reclassification threshold, as well as from models where $f(A_{it})$ is specified as linear and quadratic terms that are both allowed to have different slopes on each side of the cutoff. We cluster standard errors by school district. As noted in the main body of the paper, the ACCESS exam is administered in November or December of each school year, and the corresponding reclassification decision occurs at the end of the school year. The WKCE is administered in November of each school year. Thus, we estimate the effect of being reclassified on the basis of the ACCESS exam taken in school year t on WKCE performance in school year $t + 1$. Because of the grades in which the WKCE is administered—grades 3 through 8 and 10—we are unable to estimate the effects of 10th-grade reclassification. Instead, we estimate the effects of reclassification for students in three different grade ranges: (1) all grades; (2) grades 2 through 5; and (3) grades 6 through 9. Our decision to estimate the model separately across grade ranges is informed by the evidence that Robinson (2011) presents regarding heterogeneity across grades in the effects of reclassification on content-area achievement. For each of the three grade ranges we analyze, we estimate equation (C.1) over all cases within 10 scale score points of the reclassification threshold—this provides us with an analytic sample of over 17,000 students for the "all grades" sample. Table C1 presents descriptive statistics for the

[36] The vector of background characteristics contains measures of sex, subsidized lunch eligibility, disability status, grade, and school year.

"all grades" sample, both as a whole and separately for students that score above and below the reclassification threshold.[37]

The reduced-form estimates of δ resulting from estimation of equation (C.1) can be interpreted as the effect of increasing the probability that a student is reclassified, at least for students near the reclassification threshold. To obtain an estimate that can be interpreted as the effect of reclassification per se we use an instrumental variables approach—estimated in a two-stage least squares (2SLS) framework—very similar to the one we employed in the main body of the paper to estimate the effect of 10th-grade reclassification on ACT scores, high school graduation, and postsecondary enrollment. The first stage of this approach predicts reclassification using the following model:

$$R_{it} = f(A_{it}) + \psi H_{it} + \boldsymbol{X_{it}\pi} + \omega_{it} \tag{C.2}$$

where R is an indicator for being reclassified as fully English proficient, $f(A_{it})$ is the flexible function of distance from the reclassification threshold described earlier, H is an indicator for scoring above the reclassification threshold, $\boldsymbol{X}$ is the vector of student background characteristics listed above, and ω is the error term. The predicted values of R resulting from estimation of equation (C.2)—denoted as $\hat{R}$ below—are then inserted into the second-stage equation, taking the place of the indicator for scoring above the reclassification threshold from the reduced-form model above. The second-stage model can be written as:

$$Y_{it+1} = f(A_{it}) + \lambda \hat{R}_{it} + \boldsymbol{X_{it}\beta} + \varepsilon_{it}. \tag{C.3}$$

Because $\hat{R}$ contains only the variation in reclassification attributable to scoring above the specified cutoff, it is uncorrelated with ε and the resulting estimate of λ thus represents the local average treatment effect (LATE) of reclassification on content-area achievement.

Table C2 presents estimates of the effect of reclassification in a given year on students' content-area achievement—reading and math—in the following school year. The table presents results from two sets of models, one where the flexible function of the running variable is specified as a linear term with different slopes on each side of the reclassification cutoff and a second where it is specified as a linear and quadratic term, each of which are allowed to have different slopes on each side of the reclassification threshold. The top panel of the table presents the results for students across all grades and they demonstrate that, on average, reclassification has no effect on either reading or math achievement. Point estimates from both specifications of the reduced-form model are close to zero—within 0.03 standard deviations across both specifications and subjects.

As noted above, the reduced-form estimates are not the effect of reclassification per se, but rather the effects of increasing the probability of reclassification. Estimates of the LATE are provided by the 2SLS results. The first row of the 2SLS results in the top panel of Table C2 present the estimated coefficients on the measure of scoring above the reclassification threshold from the first-stage equation predicting reclassification. The second row of the 2SLS results presents the estimated coefficients for the measure of reclassification, which has been instrumented with

[37] Summary statistics for the lower and upper grade samples are substantively similar to the "all grades" sample and are available from the authors upon request. In addition, the results of all validity checks for the design provide confidence in the ability of our regression discontinuity approach to return valid causal estimates in this context. The results of the validity checks are available from the authors upon request.

Table C2. Coefficients and standard errors on measures of reclassification, by subject, grade range, and model specification.

	Running variable: Linear on both sides of cutoff		Running variable: Quadratic on both sides of cutoff	
Model	Reading	Math	Reading	Math
		All grades		
Reduced form—scoring above reclassification threshold	−0.024*	−0.029	0.003	0.012
	(0.014)	(0.020)	(0.032)	(0.031)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.773***	0.772***	0.802***	0.802***
	(0.010)	(0.010)	(0.014)	(0.014)
2SLS—2nd stage—effect of reclassification on standardized achievement	−0.032*	−0.038	0.004	0.015
	(0.018)	(0.026)	(0.038)	(0.040)
N	17,027	17,004	17,027	17,004
		Grades K-5		
Reduced form—scoring above reclassification threshold	−0.016	−0.048*	0.025	−0.014
	(0.020)	(0.027)	(0.040)	(0.044)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.790***	0.789***	0.812***	0.811***
	(0.013)	(0.013)	(0.018)	(0.018)
2SLS—2nd stage—effect of reclassification on standardized achievement	−0.021	−0.061*	0.031	−0.017
	(0.025)	(0.034)	(0.050)	(0.054)
N	10,939	10,920	10,939	10,920
		Grades 6 to 12		
Reduced form—scoring above reclassification threshold	−0.040	0.003	−0.039	0.064
	(0.026)	(0.030)	(0.043)	(0.045)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.742***	0.742***	0.786***	0.787***
	(0.017)	(0.017)	(0.025)	(0.025)
2SLS—2nd stage—effect of reclassification on standardized achievement	−0.054	0.004	−0.050	0.081
	(0.034)	(0.040)	(0.054)	(0.058)
N	6,088	6,084	6,088	6,084

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students in the relevant grades within 10 scale score points of automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification controls for it using a linear term with different slopes below and above the cutoff. The second specification controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. The first stage of the 2SLS model predicts reclassification using a model identical in structure to the reduced-form model. Other than substitution of the predicted value of reclassification for the indicator of scoring above the reclassification threshold, the second stage of the 2SLS model is identical in structure to the reduced-form model. $p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

the indicator for scoring above the reclassification cutoff and thus represents the estimated effect of reclassification. Mirroring the reduced-form results, the 2SLS results provide no evidence of any systematic effect of reclassification on content-area achievement, at least across all grades.

As noted earlier, we estimate both the reduced-form and 2SLS models separately for students in grades 5 and below and students in grades 6 and above. The respective results are presented in the middle and bottom panels of Table C2. Among students reclassified in elementary grades, the results indicate a small negative effect of reclassification on math achievement—about −0.05 standard deviations—when the running variable is specified as a linear term with different slopes on each side of the cutoff. The estimated negative effect is smaller in magnitude and statistically insignificant, however, in the model with the second specification of the running variable. Neither specification returns evidence of any effect of reclassification on reading achievement for students in elementary grades.

For students reclassified in grades 6 and above, the results in Table C2 again provide no evidence of a significant effect on content-area achievement. Both specifications of the reduced-form model return point estimates of the effect of reclassification on reading achievement of −0.04 standard deviations, but neither estimate is statistically significant. In math, the point estimates from the two models are noticeably different—0.003 in the model where the flexible function of the running variable is specified as linear on both sides of the cutoff and 0.064 when it is specified as both first- and second-order terms on each side—but both are positive and statistically insignificant. Considered together, the results suggest that being reclassified as fully English proficient has little effect on content-area achievement. There is suggestive evidence of a small negative effect of reclassification on reading achievement of middle school and high school students—the point estimates are consistent across specifications, but the estimated effects are not statistically different from zero. For the other grade ranges and subjects, the point estimates are either very close to zero or inconsistent across specifications.

ELL status has implications for the manner in which the Wisconsin Knowledge and Concepts Examination (WKCE)—the test Wisconsin uses to meet federal accountability requirements—is administered to students. In particular, for students who have been reclassified as fully proficient, the test is administered just as it is to native English speakers—in English and with no accommodations unless otherwise dictated by a student's Individualized Education Plan. Students classified as an ELL, on the other hand, are eligible for a wide range of accommodations when taking the WKCE. Indeed, in any given year approximately 30 percent of students classified as an ELL take the WKCE with at least one accommodation. In contrast, less than 5 percent of students who were formerly classified as an ELL but have subsequently been reclassified as fully proficient take the test with an accommodation. To assess the extent to which differential eligibility criteria for accommodations influence the results we estimate a variant of the reduced-form model in equation (C.1) where we include an indicator of whether students received an accommodation on the WKCE in school year $t + 1$. More specifically, for the model in which a student's reading score is specified as the outcome, we include a measure indicating whether the student received an accommodation on the reading portion of the test. Similarly, we include an indicator for receiving an accommodation on the math portion of the test when math scores are specified as the outcome. We estimate this variant of equation (C.1) separately for all grades, students in grades 5 and below, and for students in grades 6 and above. In performing this analysis, we note that the discontinuity in accommodation receipt at the automatic reclassification threshold does not render our estimated effects of reclassification invalid. Indeed, accommodations are a potential mechanism through which reclassification may affect student outcomes—they can be considered part of the reclassification treatment. However, the extent

Table C3. Coefficients and standard errors on measure of scoring above automatic reclassification threshold from models predicting WKCE score, by model specification and inclusion of covariate indicating use of accommodation.

Model	Running variable: Linear on both sides of cutoff		Running variable: Quadratic on both sides of cutoff	
	Reading	Math	Reading	Math
		All grades		
Reduced form—scoring above reclassification threshold	−0.024*	−0.029	0.003	0.012
	(0.014)	(0.020)	(0.032)	(0.031)
Accommodation indicator not included as covariate				
Reduced form—scoring above reclassification threshold	−0.037**	−0.030	−0.010	0.011
	(0.014)	(0.021)	(0.030)	(0.032)
Accommodation indicator included as covariate				
N	17,027	17,004	17,027	17,004
		Grades K-5		
Reduced form—scoring above reclassification threshold	−0.016	−0.048*	0.025	−0.014
	(0.020)	(0.027)	(0.040)	(0.044)
Accommodation indicator not included as covariate				
Reduced form—scoring above reclassification threshold	−0.032	−0.046	0.007	−0.011
	(0.020)	(0.028)	(0.039)	(0.044)
Accommodation indicator included as covariate				
N	10,939	10,920	10,939	10,920
		Grades 6 to 12		
Reduced form—scoring above reclassification threshold	−0.040	0.003	−0.039	0.064
	(0.026)	(0.030)	(0.043)	(0.045)
Accommodation indicator not included as covariate				
Reduced form—scoring above reclassification threshold	−0.045*	−0.001	−0.044	0.058
	(0.026)	(0.030)	(0.043)	(0.045)
Accommodation indicator included as covariate				
N	6,088	6,084	6,088	6,084

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within 10 points of the automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the left panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the right panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. $p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

to which accounting for accommodation use changes the estimated effects from the reduced-form models has important implications for interpretation of the results, as well as for potential actions that policymakers may take in response to them.

The results from the models containing an indicator for accommodation use are presented in Table C3. For purposes of comparison, Table C3 also presents the results from the models that do not include accommodation use as a covariate (i.e., the primary reduced-form results presented in Table C2). Across both specifications and each grade range, the reading results demonstrate that the coefficient estimate for the indicator of scoring above the automatic reclassification threshold is smaller, if only slightly, when the model contains an indicator of accommodation use, relative to the coefficient from the model that contains no indicator for accommodation use. This provides evidence that any negative effect of scoring above the reclassification threshold—slight as it may be—is unlikely to be driven by differential accommodation use above and below the threshold. Rather, it suggests that other aspects of the reclassification treatment, such as instructional changes or differences in peer composition, are likely responsible for any observed effects. In math, the coefficient on the indicator for scoring above the automatic reclassification threshold is not systematically different in the models that do and do not include an indicator for accommodation use. In both subjects, the results are substantively similar if we allow the coefficient on the accommodation indicator to vary on each side of the cutoff.

APPENDIX D: Additional Tables for Analysis of Potential Mechanisms

Table D1. Coefficients and standard errors on measures of reclassification, by ACT-related outcome measure and model specification for students reclassified in grade 9 or 10.

		ACT scores				
Model	Take ACT	Composite	English	Reading	Math	Science
		Running variable: Linear on both sides of cutoff				
Reduced form—scoring above reclassification threshold	0.045**	0.522**	0.607**	0.554*	0.540**	0.294
	(0.019)	(0.206)	(0.303)	(0.295)	(0.264)	(0.232)
2SLS—1st stage—scoring above reclassification threshold	0.649***	0.767***	0.767***	0.767***	0.767***	0.767***
predicting reclassification	(0.018)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
2SLS—2nd stage—effect of reclassification on	0.070**	0.680***	0.791**	0.722*	0.704**	0.384
standardized achievement	(0.028)	(0.256)	(0.379)	(0.370)	(0.334)	(0.300)
N	6,397	2,569	2,569	2,569	2,569	2,569
		Running variable: Quadratic on both sides of cutoff				
Reduced form—scoring above reclassification threshold	0.038	0.837***	1.077***	0.892**	0.955**	0.392
	(0.032)	(0.292)	(0.376)	(0.424)	(0.428)	(0.437)
2SLS—1st stage—scoring above reclassification threshold	0.661***	0.768***	0.768***	0.768***	0.768***	0.768***
predicting reclassification	(0.027)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
2SLS—2nd stage—effect of reclassification on	0.079*	1.086***	1.372***	1.122**	1.285**	0.528
standardized achievement	(0.045)	(0.368)	(0.493)	(0.536)	(0.535)	(0.572)
N	6,397	2,569	2,569	2,569	2,569	2,569

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing observations from ninth or 10th grade students with an ACCESS score within 10 scale score points of automatic reclassification threshold and was the student's maximum ACCESS score in those years. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification controls for it using a linear term with different slopes below and above the cutoff. The second specification controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. The first stage of the 2SLS model predicts reclassification using a model identical in structure to the reduced-form model. Other than substitution of the predicted value of reclassification for the indicator of scoring above the reclassification threshold, the second stage of the 2SLS model is identical in structure to the reduced-form model. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table D2. Coefficients and standard errors on measures of reclassification, by postsecondary-related outcome measure and model specification for students reclassified in grade 9 or 10.

Model	Graduate from high school	Postsecondary enrollment—fall after HS graduation	Attend four-year postsecondary institution	Attend two-year postsecondary institution
	Running variable: Linear on both sides of cutoff			
Reduced form—scoring above reclassification threshold	0.018	0.084**	0.084**	−0.037
	(0.019)	(0.035)	(0.034)	(0.031)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.778***	0.778***	0.778***	0.778***
	(0.023)	(0.023)	(0.023)	(0.023)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.023	0.108**	0.108**	−0.048
	(0.023)	(0.045)	(0.044)	(0.039)
N	3,180	3,180	3,180	3,180
	Running variable: Quadratic on both sides of cutoff			
Reduced form—scoring above reclassification threshold	0.065**	0.106**	0.063	−0.050
	(0.027)	(0.053)	(0.045)	(0.045)
2SLS—1st stage—scoring above reclassification threshold predicting reclassification	0.773***	0.773***	0.773***	0.773***
	(0.034)	(0.034)	(0.034)	(0.034)
2SLS—2nd stage—effect of reclassification on standardized achievement	0.086**	0.135**	0.085	−0.066
	(0.035)	(0.064)	(0.057)	(0.058)
N	3,180	3,180	3,180	3,180

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing observations from ninth or 10th grade students with an ACCESS score within 10 scale score points of automatic reclassification threshold and was the student's maximum ACCESS score in those years. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the top panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the bottom panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff. The first stage of the 2SLS model predicts reclassification using a model identical in structure to the reduced-form model. Other than substitution of the predicted value of reclassification for the indicator of scoring above the reclassification threshold, the second stage of the 2SLS model is identical in structure to the reduced-form model. $p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table D3. Coefficients and standard errors on measure of scoring above automatic reclassification threshold from models predicting postsecondary enrollment the fall after high school graduation, by model specification and inclusion of covariate measuring ACT composite score.

Model	Running variable: Linear on both sides of cutoff	Running variable: Quadratic on both sides of cutoff
Reduced form—scoring above reclassification threshold	0.077*	0.066
ACT score not included as covariate	(0.045)	(0.082)
N	1,127	1,127
Reduced form—scoring above reclassification threshold	0.055	0.035
ACT score included as covariate	(0.045)	(0.080)
N	1,127	1,127

Notes: Robust standard errors clustered by district in parentheses below coefficient estimates. All estimates based on analytic samples containing students within 10 points of the automatic reclassification threshold. Reduced-form models contain measures of sex, disability status, eligibility for subsidized lunch, grade, and school year. Distance from the reclassification threshold is controlled for using two different specifications of the flexible function of the running variable. The first specification—presented in the left panel of the table—controls for it using a linear term with different slopes below and above the cutoff. The second specification—presented in the right panel of the table—controls for it using both a linear and quadratic term, each of which is allowed to have different slopes below and above the cutoff.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

APPENDIX REFERENCES

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82, 2295–2326.

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. Educational Evaluation and Policy Analysis, 33, 267–292.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs. Washington, DC: U.S. Department of Education. Retrieved March 13, 2014, from http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.