WORKING PAPER

Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials

June 2012

John Deke and Lisa Dragoset

MATHEMATICA
Policy Research

ABSTRACT

The regression discontinuity design (RDD) has the potential to yield findings with causal validity approaching that of the randomized controlled trial (RCT). However, Schochet (2008a) estimated that, on average, an RDD study of an education intervention would need to include three to four times as many schools or students as an RCT to produce impacts with the same level of statistical precision. We extend the work of Schochet by empirically assessing the effect on sample size requirements of accounting for selection of an optimal bandwidth and the adjustment for random misspecification error, both of which are needed to estimate consistent RDD impacts and control the Type I error rate. We used data from four previously published education studies covering more than 25,000 students in kindergarten to grade 9 in 24 states and 330 schools to calculate empirical estimates of the RDD design effect taking into account these additional factors. We find that an RDD study needs between 9 and 17 times as many schools or students as an RCT to produce an impact with the same level of statistical precision, and that the need for a large sample is driven primarily by bandwidth selection, not adjusting for random misspecification error.

I. INTRODUCTION

The regression discontinuity design (RDD) can be used to estimate the impact of an intervention in cases in which a randomized controlled trial (RCT) is not feasible. (See Cook [2008] for a thorough account of the multidisciplinary history of RDD.) Under this design, individuals are selected to receive an intervention based on where they fall relative to a cutoff value on a continuous assignment variable. If the relationship between such a variable and the outcome is continuous in the absence of the intervention, then a discontinuity in the outcome-assignment variable relationship at the cutoff value can be interpreted as the impact of the intervention. Because the researcher observes the assignment mechanism, RDD is not as susceptible to omitted variable bias as other nonexperimental designs (for example, propensity score matching). This key advantage of RDD is reflected in the decision by the U.S. Department of Education's What Works Clearinghouse to allow RDD studies to be classified in the same category as RCTs (U.S. Department of Education 2010).

Although many RDD studies are conducted retrospectively using existing data, some use varying combinations of administrative data and data collected specifically for the study at hand. Researchers conducting a prospective RDD must decide how large their study needs to be to detect the smallest effect that would be considered meaningful. This type of statistical power analysis is frequently conducted in the design of RCTs.

Schochet (2008a) provides a detailed illustration of a key source of the difference in the statistical variance of an RDD impact relative to an RCT impact, which we will refer to as the regression discontinuity design effect (RDDE). In short, the RDDE arises because estimation of unbiased RDD impacts requires regression adjustment for the assignment variable. Because the assignment variable is correlated with treatment status, including the assignment variable in the impact regression unavoidably reduces the statistical precision of the impact estimate.

To bring attention to the main source of the RDDE, Schochet focused on the case of an RDD impact analysis that used all available data regardless of distance from the cutoff value and a linear functional form. In planning an RDD evaluation, however, researchers must also account for further erosion of statistical precision in the RDD impact due to the additional analytic steps necessary to estimate a credible impact. In particular, selection of an optimal bandwidth can reduce the sample size of the study.[1] Furthermore, a lack of continuity in the assignment variable can reduce precision because standard errors must be inflated to account for what Lee and Card (2008) describe as random misspecification error, which occurs when multiple units share the same value of the assignment variable. This issue can also be described as a form of clustering—if multiple units share the same value of the assignment variable, then clusters are being assigned to treatment or control status, not individuals. As in a clustered RCT, this clustered assignment must be accounted for when calculating standard errors.

These additional contributors to the RDDE depend on multiple characteristics of, and relationships among, the variables involved in estimating RDD impacts. These relationships are not

[1] Most RDD studies estimate the relationship between the outcome and the assignment variable using either a linear regression within a bandwidth or a polynomial regression using all the data regardless of distance from the cutoff. We focus solely on the optimal bandwidth method because preliminary simulation results indicate that this method performs better than alternative approaches, in terms of having smaller bias and accurately estimated standard errors. We plan to report our simulation findings in future work.

generally known during the planning stages of an evaluation, yet they can greatly influence the statistical precision of RDD impacts and, therefore, the required sample size for an evaluation.

Although this situation can be challenging to researchers designing prospective RDD evaluations, it is not without precedent. Researchers conducting clustered RCTs (for example, studies in which schools, not students, are randomly assigned to treatment and control groups) must contend with a design effect due to clustering that is also not known until data have been collected and analyzed. To help researchers anticipate the potential effect of clustering, the following papers report on the effects of clustering in varying contexts: Hedges and Hedberg (2007); Bloom et al. (2007); Schochet (2008b); and Deke et al. (2010). These papers report two key parameters—the intraclass correlation (ICC) and regression R^2—for different combinations of outcomes and covariates for varying populations of interest. Researchers planning RCTs can use these parameter estimates to calculate the sample size needed to detect meaningful effect sizes.

The purpose of this paper is to contribute to a similar literature for researchers planning prospective RDD studies in education in which an academic pre-test score is used as the assignment variable and a post-test score is used as the outcome. We estimate three quantities using data from past education studies: (1) the proportion of data excluded from RDD impact analyses using the Imbens and Kalyanaraman (2009) (hereafter, I&K) optimal bandwidth selection algorithm, (2) the increase in the variance due to the Lee and Card (2008) (hereafter, L&C) correction, and (3) the full RDDE.

Our empirical strategy draws on data from four past education studies conducted for the Institute of Education Sciences (IES) covering more than 25,000 students in kindergarten to grade 9 in 24 states and 330 schools. We use the pre-test scores from these studies as if they were the assignment variable in an RDD impact analysis and the post-test scores as if they were the outcomes. We conduct this analysis at three pseudo cutoff values of the pseudo assignment variables: the 25th, 50th, and 75th percentiles of the pre-test scores. Because we know that no intervention was actually allocated using these pre-tests (these pre-tests were administered as part of the earlier studies and were not used by the study authors to assign students to any intervention), our analysis is implicitly conducted under the null hypothesis of no RDD impacts. This is appropriate because statistical power analyses are typically conducted under the null hypothesis of no impacts.

II. MINIMUM DETECTABLE EFFECT SIZES IN RDD EVALUATIONS

The statistical power of an evaluation is often expressed in terms of a minimum detectable effect size (MDES). The MDES is the smallest effect a study could detect with high probability, measured relative to the standard deviation of the outcome variable (either for the study's sample or a benchmark population of interest). Here we express the MDES for an RDD study as a function of the RDDE and the variance of an RCT impact that would be estimated using the same outcome and sample size,

$$(1)\quad MDES = \left(T^{-1}(\alpha, df) + T^{-1}(1-\beta, df)\right) * \frac{\sqrt{RDDE * Var(RCT\_impact)}}{\sigma}$$

where *Var(RCT_impact)* is the variance of an RCT impact estimate, σ is the standard deviation of the outcome variable (either in the study sample or a benchmark population of interest), $T^{-1}(.)$ is the inverse of the t-distribution, α is the probability of a type 1 error, β is the probability of a type 2 error (so 1-β is power), and *df* is the number of degrees of freedom.

The RDDE is the ratio of the RDD impact variance to the RCT impact variance, holding sample size constant. Schochet (2008a) shows that in the case of an RDD analysis using a linear functional form and all available data, the RDDE reduces to:

$$(2) \quad RDDE = \frac{1}{\left(1-\rho^2\right)}$$

where ρ is the correlation between treatment status and the continuous assignment variable. We can then calculate the MDES for an RDD study by multiplying the MDES for an RCT of the same size by the square root of the RDDE. For an assignment variable following the uniform distribution and a cutoff at the median, Schochet showed that the RDDE is 4.0. If the assignment variable follows the normal distribution, the RDDE is 2.75. The effect on the MDES is the square root of the RDDE. (For example, in the case of the uniform distribution, the MDES for an RDD study is twice the MDES for an RCT with the same sample size.)

In practice, the RDDE is likely to be even larger than shown in equation 2 due to selection of an optimal bandwidth and due to the L&C adjustment, where applicable. The effect of selecting an optimal bandwidth is straightforward—it reduces the sample size for the impact analysis. The effect of the L&C adjustment depends on the residual intraclass correlation coefficient (ICC), that is, the proportion of variation in the residual remaining after the outcome is regressed on the assignment variable that is between, rather than within, unique values of the assignment variable. The greater the ICC, the greater the RDDE, and hence, the more precision is lost when making the necessary L&C adjustment.

III. DATA

This study uses baseline and follow-up student-level test score data from four large-scale experimental studies that Mathematica conducted for IES. The data from these studies are available as restricted use files from IES. The four studies are as follows:

- Evaluation of Reading Comprehension Interventions (James-Burdumy et al. 2010)
- Impact Evaluation of Teacher Preparation Models (Constantine et al. 2009)
- Evaluation of Mathematics Curricula (Agodini et al. 2009)
- Evaluation of Educational Technology Interventions (Campuzano et al. 2009)

See Table 1 for a brief description of each study, including the grades covered.

Together, these studies yield 25 combinations of baseline and follow-up test scores covering a total of 25,000 students drawn from kindergarten to grade 9 in 24 states and 330 schools. (In accordance with the National Center for Education Statistics [NCES] publication policy, all student sample sizes are rounded to the nearest 10. State, district, and school sample sizes come from the published reports referenced above.) We analyzed one year of data for each of the studies, meaning that the baseline test was conducted at the beginning of the school year and the follow-up test was conducted at the end of the same school year.

Table 1. Description of Data Sources

Study	Purpose of Study	Student Grade(s)	Student Outcome Measures	Unit of Random Assignment	Number of States	Number of Districts	Number of Schools	Number of Students	Response Rate Pre-Test	Response Rate Post-Test
Evaluation of Reading Comprehension Interventions (James-Burdumy et al. 2010)	Evaluates the impact of four interventions on fifth-grade reading achievement.	5	Group Reading Assessment and Diagnostic Evaluation (GRADE); Educational Testing Service (ETS) Science Reading Comprehension Assessment; ETS Social Studies Reading Comprehension Assessment	School	8	10	90	6,350	0.99	0.88
Evaluation of Early Elementary School Mathematics Curricula (Agodini et al. 2009)	Compares the effects of four different elementary math curricula on improving student math achievement.	1, 2	Early Childhood Longitudinal Study Mathematics Assessment	School	4	10	40	1,580	0.96	0.87
Evaluation of Teacher Preparation Models (Constantine et al. 2009)	Examines the effect of different approaches to teacher preparation on teacher practice and student performance.	K-5	Reading comprehension, vocabulary, math concepts and applications, and math computation subtests of the California Achievement Tests, Fifth Edition	Student	7	20	60	2,490	0.97	0.90
Evaluation of the Effectiveness of Reading and Mathematics Software Products (EERMSP) (Campuzano et al. 2009)	Study randomly assigned teachers to a treatment group that used a specified educational technology or a control group that used conventional teaching approaches. The study consisted of four sub-studies of different interventions at different grade levels (see four rows below).	-	-	-	-	-	-	-	-	-
EERMSP Grade 1	-	1	Stanford Achievement Test (Version 10), Reading Test of Word Reading Efficiency	Teacher	12	20	50	4,420	0.97	0.95
EERMSP Grade 4	-	4	Stanford Achievement Test (Version 10), Reading	Teacher	9	10	40	3,110	0.93	0.93
EERMSP Grade 6	-	6	Stanford Achievement Test (Version 10), Math	Teacher	7	10	30	4,260	0.96	0.89
EERMSP Algebra	-	8, 9	ETS End-of-Course Algebra Assessment	Teacher	8	10	20	3,010	0.82	0.81

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: Student, district, and school sample sizes are rounded to the nearest 10 in accordance with NCES publication policy. State sample sizes are taken from the references listed in the first column.

For all studies, the pre- and post-tests were selected and administered by the study to ensure we would understand how these tests were used. In particular, we knew they were not used to assign students to receive an intervention. Thus, we knew that all our RDD impact analyses were being conducted under the null hypothesis of no impact. Table 2 describes each combination of pre- and post-test, providing a letter code for each case. These letter codes are used in subsequent tables.

IV. METHODOLOGY

For each combination of outcome and assignment variable described in Section III, we calculated RDD impacts and standard errors at the 25th, 50th, and 75th percentiles of the assignment variable using the following three approaches:

- A linear functional form using all observations
- A linear functional form using only observations within an optimal bandwidth
- A linear functional form using only observations within an optimal bandwidth and the L&C adjustment

In this section, we describe our approach to calculating RDD impacts, the I&K bandwidth selection algorithm, and the L&C adjustment for random misspecification error.

A. Calculating RDD Impacts

We follow Imbens and Lemieux (2008)[2] and estimate impact regressions described in Equations 3 and 4. The RDD impact is the difference between the constant terms: $\hat{\tau}_{RD} = \hat{\alpha}_L - \hat{\alpha}_R$. The other terms in Equations 3 and 4 are the outcome (Y), the assignment variable (X), an index of students (i), an index of unique values of the assignment variable (j), random misspecification error[3] (u), a student error term (ε), and the cutoff value of the assignment variable (c). We center X at c so that the intercept terms can be interpreted as the predicted average outcomes at the cutoff value of the assignment variable. These two equations are estimated separately. In estimating these equations, we use kernel weights associated with the I&K bandwidth selection algorithm (discussed in Section B).

$$(3) \quad Y_{ij} = \alpha_L + \beta_L X_{ij} + u_j + \varepsilon_{ij};\ i, j \in \left\{ i, j \mid X_{ij} \le c \right\}$$

$$(4) \quad Y_{ij} = \alpha_R + \beta_R X_{ij} + u_j + \varepsilon_{ij};\ i, j \in \left\{ i, j \mid X_{ij} > c \right\}$$

[2] Imbens and Lemieux assume that the treatment group consists of observations above the cutoff value of the assignment variable. We make the opposite assumption because we are focused on cases where low-achieving students are assigned to receive an intervention using a cutoff value on a continuous measure of achievement.

[3] This reflects the idea of a random misspecification error introduced by L&C. Imbens and Lemieux did not include this term.

Table 2. Outcome and Assignment Variable Codes

Code	Grade Administered	Outcome (measured post-intervention)	Assignment Variable (measured pre-intervention)
A	5	Group Reading Assessment and Diagnostic Evaluation	Group Reading Assessment and Diagnostic Evaluation
B	5	ETS Science Reading Comprehension Assessment	Group Reading Assessment and Diagnostic Evaluation
C	5	ETS Science Reading Comprehension Assessment	Test of Silent Contextual Reading Fluency
D	5	ETS Social Studies Reading Comprehension Assessment	Group Reading Assessment and Diagnostic Evaluation
E	5	ETS Social Studies Reading Comprehension Assessment	Test of Silent Contextual Reading Fluency
F	K-5	CAT-5: Vocabulary	CAT-5: Vocabulary
G	K-5	CAT-5: Reading Comprehension	CAT-5: Reading Comprehension
H	K-5	CAT-5: Math Computation	CAT-5: Math Computation
I	K-5	CAT-5: Math Concepts and Applications	CAT-5: Math Concepts and Applications
J	1	ECLS Mathematics Assessment	ECLS Mathematics Assessment
K	9	ETS End-of-Course Algebra Assessment: Skills Score	ETS End-of-Course Algebra Assessment: Skills Score
L	9	ETS End-of-Course Algebra Assessment: Concepts Score	ETS End-of-Course Algebra Assessment: Concepts Score
M	9	ETS End-of-Course Algebra Assessment: Total Score	ETS End-of-Course Algebra Assessment: Total Score
N	9	ETS End-of-Course Algebra Assessment: Processes Score	ETS End-of-Course Algebra Assessment: Processes Score
O	1	SAT-10: Sounds and Letters Score	SAT-10: Sounds and Letters Score
P	1	SAT-10: Sentence Reading Score	SAT-10: Sentence Reading Score
Q	1	SAT-10: Word Reading Score	SAT-10: Word Reading Score
R	1	SAT-10: Total Reading Score	SAT-10: Total Reading Score
S	4	SAT-10: Reading Comprehension Score	SAT-10: Reading Comprehension Score
T	4	SAT-10: Reading Vocabulary Score	SAT-10: Reading Vocabulary Score
U	4	SAT-10: Total Reading Score	SAT-10: Total Reading Score
V	4	SAT-10: Work Study Skills Score	SAT-10: Work Study Skills Score
W	6	SAT-10: Problem Solving Score	SAT-10: Problem Solving Score
X	6	SAT-10: Total Math Score	SAT-10: Total Math Score
Y	6	SAT-10: Procedures Score	SAT-10: Procedures Score

Source: Randomized controlled trials previously completed by Mathematica for IES.

CAT-5= California Achievement Test, Fifth Edition; ECLS = Early Childhood Longitudinal Study; ETS = Educational Testing Service; SAT-10 = Stanford Achievement Test-Version 10.

B. The I&K Bandwidth Selection Algorithm

Equations 3 and 4 are estimated only for observations within an optimal bandwidth selected by the I&K algorithm. This algorithm is a data-driven approach to estimate the optimal bandwidth that minimizes the mean squared error (MSE) of the impact estimate. This optimal bandwidth balances the tradeoff between bias and variance. To be specific, let τ_{RD} denote the average effect of the treatment for students with assignment variable values equal to the cutoff value, and let $\hat{\tau}_{RD}(h)$ denote the estimate of τ_{RD} using a bandwidth of h. From I&K, the MSE of the impact estimate is:

(5) $$MSE(h) = E[(\hat{\tau}_{RD}(h) - \tau_{RD})^2].$$

Let h^* be the optimal bandwidth that minimizes this criterion:

(6) $$h^* = \arg\min MSE(h).$$

Now define the asymptotic mean squared error (AMSE) as a function of the bandwidth:

(7) $$AMSE(h) = C_1 * h^4 * (m_+^{(2)}(c) - m_-^{(2)}(c))^2 + \frac{C_2}{N * h} * \left(\frac{\sigma_+^2(c)}{f_+(c)} + \frac{\sigma_-^2(c)}{f_-(c)} \right),$$

where C_1 and C_2 are constants, c is the cutoff value, N is the sample size of the original data set, $m_+^{(2)}(c)$ and $m_-^{(2)}(c)$ are the right and left limits of the second derivative of the relationship between the outcome and the assignment variable at the cutoff, $\sigma_+^2(c)$ and $\sigma_-^2(c)$ are the right and left limits of the conditional variance of the outcome variable given the assignment variable at the cutoff, and $f_+(c)$ and $f_-(c)$ are the right and left limits of the density of the assignment variable at the cutoff. The first term in Equation 7 corresponds to the bias of the $\hat{\tau}_{RD}(h)$ estimator, and the second term corresponds to the variance. The bias of the $\hat{\tau}_{RD}(h)$ estimator is affected by the curvature (that is, the second derivative) of the relationship between the outcome and the assignment variable at the cutoff because the more curvature that exists in the data, the more bias will exist in a linear regression estimate at the cutoff.

The I&K bandwidth, $\hat{h}_{opt}$, is an estimate of h^*, and equals:

(8) $$h_{opt} = C_K * \left(\frac{2 * \hat{\sigma}^2(c) / \hat{f}(c)}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c)\right)^2 + (\hat{r}_+ + \hat{r}_-)} \right)^{1/5} * N^{-1/5},$$

where C_K is a constant that depends on the kernel used (we follow I&K and use C_K = 3.4375, which corresponds to the edge kernel), $\hat{\sigma}^2(c)$ is an estimate at the cutoff value of the conditional variance of the outcome variable given the assignment variable, $\hat{f}(c)$ is an estimate of the density of the assignment variable at the cutoff, $\hat{m}_+^{(2)}(c)$ and $\hat{m}_-^{(2)}(c)$ are estimates of the limits of the second derivatives at the cutoff value from the right and left, respectively (that is, $\hat{m}_+^{(2)}(c)$ and $\hat{m}_-^{(2)}(c)$

estimate the curvature of the data at the cutoff value), and $(\hat{r}_+ + \hat{r}_-)$ is what I&K call a "regularization term" that is a function of the previous four components. (All of these estimates are calculated as in I&K.) The regularization term addresses the problem that the curvature of the data could be spuriously underestimated when the sample size is low. It does this by imposing the conservative assumption that at least some curvature exists in the data. The size of the regularization term decreases with sample size, and nearly vanishes when the sample size is large.

C. The L&C Adjustment for Random Misspecification Error

The RDD is predicated on assignment to treatment and comparison status using a continuous variable. However, in practice, many variables are not truly continuous. For example, there are a finite number of unique values of a test score because there are a finite number of questions on a test. A more extreme example would be to assign students to treatment and comparison groups using letter grades (A through F)—a variable that is clearly not continuous, though it is ordinal.

One way to interpret the contribution of L&C is as a solution to this problem of ambiguity in the continuity of assignment variables. L&C observe that a lack of continuity in an assignment variable can lead to random misspecification error and they suggest using clustered standard errors to protect against false inferences that might arise from this error. In this case, students are clustered within unique values of the assignment variable. This key insight by L&C reduces the need for subjective or arbitrary judgments about which assignment variables are continuous enough because the standard errors are simply inflated to protect against the consequences of a lack of continuity.

Random misspecification error is reflected in Equations 3 and 4 by the term *u*. To account for the effects of this random error, we use two approaches suggested by L&C. The first approach assumes that, at the cutoff value of the assignment, variable *u* has the same sign and magnitude in Equation 3 as it does in Equation 4 (meaning the specification error is unrelated to treatment status). L&C refer to this case as identical specification errors. Following L&C, we adjust standard errors in this case using clustered standard errors. Specifically, we estimate Equations 3 and 4 using generalized estimating equations in which the cluster identification variable is the assignment variable itself.

The second approach assumes that the values of *u* at the cutoff value from Equations 3 and 4 can be different. L&C refer to these differences as independent specification errors. Assuming independent specification errors implies a more severe adjustment of standard errors than assuming identical specification errors. Following L&C, we adjust standard errors in this case by first estimating the same clustered impact variance as in the case of identical specification errors. We then add to that impact variance an additional term, which is described in Equations 12 and 13 of L&C. Specifically, the term is $2\hat{\sigma}_u^2$, where $\hat{\sigma}_u^2$ is a consistent estimator for the variance of u_j, the misspecification error term in Equations 3 and 4.[4] The square root of the sum of the clustered

[4] $\hat{\sigma}_u^2 \equiv \frac{1}{N}\sum_{j=1}^{J} n_j(\bar{Y}_j - W_j\hat{\theta})^2 - \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j - 1}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2$ where N is the total sample size, n_j is the sample size within each of the *j* unique values of the assignment variable, Y is the outcome, W is the vector of covariates from Equations 3 and 4, and $\hat{\theta}$ are the estimated coefficients from Equations 3 and 4.

impact variance (assuming identical specification errors) and the additional term $2\hat{\sigma}_u^2$ equals the impact standard error adjusted for independent specification errors.

V. FINDINGS

We begin our presentation of findings by examining the two underlying contributors to the ultimate RDDE: sample size reduction from the I&K bandwidth and clustering effects from the L&C adjustment. Next, we calculate a progression of standard errors for each data set, starting with RCT standard errors and ending with RDD standard errors that account for both I&K and L&C. This culminates in the RDDE for each data set. Finally, we consider the difference in sample sizes needed to detect desired effect sizes.

A. Contributors to the RDDE

In Tables 3 and 4, we examine the contributors to the RDDE that are the focus of this paper, namely the effect on sample size of the I&K bandwidth selection algorithm (Table 3) and the extent to which observations are clustered within unique values of the assignment value along with the ICC (Table 4).

For the median outcome-assignment case in Table 3 (all discussion in this paragraph is focused on the median case), approximately half of all students are included in the I&K bandwidth regardless of the cutoff value. To see this, note that when the cutoff is at the 25th percentile of the assignment variable, 96 percent of observations below the cutoff and 31 percent of observations above the cutoff are included in the I&K bandwidth, which implies that 47 percent of all observations are included: 0.25*0.96 + 0.75*0.31 = 0.47. A similar calculation shows that 66 percent and 56 percent of all observations are included in the bandwidth when the cutoff is at the 50th and 75th percentiles of the assignment variable, respectively.

Because the I&K algorithm selects a bandwidth equal in width on both sides of the cutoff, the proportion of observations above or below the cutoff value varies across cutoff values. Specifically, 96 percent of observations below the cutoff are in the bandwidth when the cutoff is at the 25th percentile of the assignment variable, while 31 percent of observations above the cutoff are included in the bandwidth. When the cutoff value is at the 75th percentile of the assignment variable, 94 percent of observations above the cutoff are in the bandwidth, while 43 percent of observations below the cutoff are in the bandwidth. When the cutoff value is at the median of the assignment variable, the proportion of observations above the cutoff that are in the bandwidth (72 percent) is closer to the proportion of observations below the cutoff that are in the bandwidth (59 percent). A consequence of this pattern is that, even when the cutoff value is located away from the median of the assignment variable, the bandwidth selection algorithm will tend to yield equally sized treatment and comparison groups.

The L&C adjustment to standard errors will be more severe when there are fewer unique values of the assignment variable and when the residual ICC (after controlling for the assignment variable) is high. For all but one of the cases examined in Table 4, the number of unique values of the assignment variable is substantially lower than the total number of observations. For the median case, the ratio of total observations to unique values of the assignment variable is 82 (not shown in table). This suggests the potential for a severe clustering adjustment. However, the residual ICC is typically very low, particularly within the I&K bandwidth. Specifically, in the median case at the

Table 3. Proportion of Observations Included in the Optimal Bandwidth, by Cutoff Value

Outcome-Assignment Code	Cutoff: 25th Percentile				Cutoff: 50th Percentile				Cutoff: 75th Percentile			
	Observations Below		Observations Above		Observations Below		Observations Above		Observations Below		Observations Above	
	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique
A	0.97	0.63	0.25	0.22	0.52	0.27	0.48	0.31	0.36	0.26	0.93	0.75
B	0.99	0.75	0.30	0.26	0.62	0.33	0.57	0.38	0.36	0.26	0.94	0.75
C	0.90	0.75	0.46	0.27	0.56	0.28	0.72	0.28	0.35	0.33	0.98	0.94
D	0.97	0.63	0.25	0.22	0.53	0.27	0.48	0.31	0.64	0.43	1.00	1.00
E	0.85	0.65	0.33	0.22	0.62	0.30	0.80	0.32	0.43	0.36	0.97	0.88
F	1.00	1.00	0.60	0.39	0.56	0.44	0.77	0.44	0.32	0.32	0.88	0.88
G	1.00	1.00	0.78	0.50	0.54	0.40	0.78	0.41	0.30	0.29	0.75	0.78
H	1.00	1.00	0.37	0.38	0.81	0.58	0.65	0.58	0.49	0.33	1.00	1.00
I	0.75	0.88	0.46	0.30	0.42	0.27	0.56	0.28	0.44	0.41	1.00	1.00
J	0.87	0.87	0.37	0.36	0.87	0.87	0.78	0.78	0.49	0.49	0.74	0.74
K	1.00	1.00	0.57	0.35	0.64	0.55	0.85	0.58	0.48	0.47	1.00	0.83
L	0.81	0.83	0.47	0.29	0.48	0.36	0.79	0.33	0.25	0.24	0.88	0.67
M	1.00	1.00	0.84	0.62	0.47	0.29	0.48	0.29	0.52	0.43	0.98	0.71
N	0.42	0.5	0.37	0.21	0.21	0.25	0.55	0.31	0.44	0.50	0.98	0.71
O	0.72	0.88	0.34	0.64	0.93	0.75	0.60	0.76	0.33	0.13	0.32	0.44
P	0.93	0.88	0.27	0.30	0.85	0.47	0.36	0.50	0.34	0.26	0.57	0.75
Q	0.83	0.89	0.12	0.31	0.89	0.65	0.47	0.67	0.52	0.27	0.63	0.78
R	0.98	0.91	0.26	0.44	0.77	0.42	0.31	0.43	0.37	0.19	0.43	0.57
S	1.00	1.00	0.29	0.36	0.90	0.64	0.75	0.67	0.48	0.29	0.96	0.88
T	0.96	0.60	0.13	0.25	0.87	0.50	0.58	0.55	0.65	0.33	1.00	1.00
U	1.00	0.93	0.31	0.33	0.79	0.52	0.65	0.52	0.36	0.23	0.90	0.67
V	0.90	0.60	0.09	0.20	0.92	0.40	0.57	0.40	0.54	0.27	0.81	0.60
W	0.80	0.43	0.17	0.18	0.72	0.43	0.52	0.47	0.41	0.29	0.96	0.88
X	0.97	0.75	0.29	0.26	0.73	0.48	0.59	0.48	0.31	0.24	0.87	0.67
Y	0.95	0.60	0.26	0.25	0.90	0.60	0.78	0.64	0.90	0.67	1.00	1.00
Median	**0.96**	**0.87**	**0.31**	**0.30**	**0.72**	**0.43**	**0.59**	**0.44**	**0.43**	**0.29**	**0.94**	**0.78**

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: Three assignment variable cutoff values are considered: the 25th, 50th, and 75th percentiles. For each cutoff, we report the proportion of all observations above (and below) the cutoff that are included in the I&K bandwidth. We also report the proportion of unique values of the assignment variable included in the bandwidth.

Table 4. Clustering Within Unique Values of the Assignment Variable

Outcome-Assignment Code	Student Sample Size	Number of Unique Assignment Variable Values	Cutoff: 25th Percentile		Cutoff: 50th Percentile		Cutoff: 75th Percentile	
			ICC All Observations	ICC Observations in Bandwidth	ICC All Observations	ICC Observations in Bandwidth	ICC All Observations	ICC Observations in Bandwidth
A	5,170	31	0.0000	0.0000	0.0031	0.0000	0.0069	0.0000
B	2,550	31	0.0055	0.0000	0.0071	0.0047	0.0068	0.0089
C	2,540	64	0.0020	0.0000	0.0049	0.0003	0.0038	0.0018
D	2,550	31	0.0000	0.0073	0.0012	0.0009	0.0020	0.0000
E	2,550	67	0.0069	0.0048	0.0061	0.0033	0.0049	0.0049
F	2,590	96	0.0000	0.0000	0.0000	0.0020	0.0000	0.0049
G	2,560	91	0.0162	0.0103	0.0191	0.0089	0.0211	0.0123
H	1,210	96	0.0086	0.0000	0.0000	0.0052	0.0000	0.0070
I	2,520	99	0.0044	0.0016	0.0043	0.0061	0.0045	0.0091
J	1,310	1,283	0.0000	0.0034	0.0000	0.0000	0.0000	0.0000
K	1,930	23	0.0105	0.0103	0.0126	0.0087	0.0182	0.0019
L	1,930	23	0.0130	0.0000	0.0115	0.0013	0.0123	0.0000
M	1,930	28	0.0154	0.0022	0.0212	0.0025	0.0299	0.0000
N	1,930	25	0.0182	0.0049	0.0155	0.0490	0.0168	0.0123
O	3,490	33	0.0239	0.0000	0.0159	0.0000	0.0092	0.0000
P	3,490	31	0.0026	0.0000	0.0054	0.0000	0.0080	0.0097
Q	3,530	35	0.0267	0.0000	0.0378	0.0000	0.0440	0.0049
R	3,340	91	0.0226	0.0000	0.0302	0.0004	0.0243	0.0025
S	2,520	29	0.0173	0.0000	0.0013	0.0033	0.0024	0.0000
T	2,430	21	0.0087	0.0000	0.0020	0.0000	0.0060	0.0004
U	2,400	58	0.0082	0.0000	0.0000	0.0038	0.0074	0.0000
V	2,490	20	0.0257	0.0000	0.0085	0.0000	0.0293	0.0037
W	3,610	29	0.0000	0.0000	0.0017	0.0012	0.0044	0.0000
X	3,560	46	0.0014	0.0000	0.0065	0.0008	0.0065	0.0000
Y	3,550	21	0.0039	0.0000	0.0000	0.0000	0.0026	0.0000
Median			**0.0082**	**0.0000**	**0.0054**	**0.0012**	**0.0068**	**0.0018**

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: Sample sizes are rounded to the nearest 10 in accordance with NCES publication policy. The ICCs shown in this table are calculated conditional on the assignment variable and the treatment indicator variable.

ICC = intraclass correlation coefficient.

median cutoff value, the overall ICC is 0.0054 and the ICC within the bandwidth is even lower: 0.0012 (a similar pattern can be seen for the other two cutoff values). The lower ICC within the bandwidth is consistent with the goal of the I&K algorithm, which is to reduce misspecification error. By way of comparison, the (unadjusted) ICCs for these test score outcomes due to students being clustered within schools range from 0.07 to 0.30 (not shown in table).

Taken together, Tables 3 and 4 suggest that the primary source of the RDDE (beyond that identified in Schochet 2008a) will be the bandwidth selection algorithm, not the L&C adjustment.

B. Standard Errors and the RDDE

In Tables 5a, 5b, and 5c, we illustrate how standard errors increase as we move incrementally from an RCT to an RDD that both uses the I&K bandwidth selection algorithm and corrects standard errors for random misspecification error using the L&C adjustment. The three tables correspond to the three pseudo-cutoff values of the assignment variables.

Consistent with what we saw in Tables 3 and 4, these tables show that the bandwidth selection algorithm, not the L&C adjustment, is the more common source of the RDDE. Of the 25 cases examined in Tables 5a, 5b, and 5c, the standard error increases in 24, 25, and 25 cases as we move from the column "RDD, All Observations, No Clustering" to the column "RDD, Observations in Bandwidth, No Clustering." Meanwhile, the number of cases in which the standard error increases further due to the L&C adjustment is 7, 13, and 12 (moving from the column "RDD, Observations in Bandwidth, No Clustering," to the column "RDD, Observations in Bandwidth, Clustering Independent"). Not surprisingly, the cases in which the standard error increases due to the L&C adjustment tend to be the cases with higher ICC values.[5]

In Table 6, we present the full RDDE for every outcome-assignment case, three cutoff values, and both L&C adjustments (identical and independent). At the bottom of the table, we present the median RDDE across cases. (Recall throughout this discussion that the RDDE is defined in terms of the *variance* of the impact, whereas in Table 5, we focus on the *standard error*.) Whereas Schochet (2008a) found RDDE values no greater than 5.37 (Tables 4.2 and 4.3 of Schochet 2008a), we found RDDE median values ranging from 9.39 to 17.16, and in some individual cases they can be substantially larger than that.

C. Sample Sizes Needed to Detect Effects

In Table 7, we report the sample sizes needed to support an MDES of 0.25, 0.20, or 0.10 in RCT-based studies or in RDD-based studies with three different RDDE values (9, 14, and 17, to capture the range of median RDDE values shown in Table 6). Schochet (2008a) found that the sample size requirement to detect a given effect size is approximately three times larger for an RDD-based study than for an RCT-based study. We found that accounting for the I&K bandwidth selection algorithm and L&C adjustment for random misspecification error results in a sample size requirement in the range of 9 to 17 times the size of an RCT (using Schochet's assumption of an impact regression R^2 of 0.50 and balanced treatment and comparison groups).

[5] A noticeable exception is Case J, which has nearly as many unique values of the assignment variable as total observations and the ICC is nearly zero. Yet the L&C adjustment is very large in this case. We suspect that this is a degenerative case and that in this situation it would be best to forego the L&C adjustment.

Table 5a. Impact Standard Errors, RDD Cutoff at 25th Percentile

Outcome-Assignment Code	RCT, All Observations, No Clustering	RDD, All Observations, No Clustering	RDD, Observations in Bandwidth, No Clustering	RDD, Observations in Bandwidth, Clustering Identical	RDD, Observations in Bandwidth, Clustering Independent
A	0.022	0.062	0.087	0.087	0.087
B	0.036	0.114	0.151	0.151	0.151
C	0.042	0.122	0.175	0.175	0.175
D	0.036	0.120	0.169	0.169	0.174
E	0.043	0.131	0.165	0.194	0.198
F	0.032	0.064	0.075	0.075	0.075
G	0.035	0.072	0.083	0.096	0.135
H	0.058	0.146	0.204	0.204	0.204
I	0.033	0.061	0.085	0.103	0.121
J	0.044	0.085	0.124	0.124	1.019
K	0.049	0.084	0.105	0.148	0.203
L	0.052	0.082	0.110	0.110	0.110
M	0.046	0.114	0.108	0.108	0.117
N	0.051	0.104	0.168	0.168	0.168
O	0.033	0.241	0.281	0.281	0.281
P	0.031	0.111	0.150	0.150	0.150
Q	0.03	0.116	0.194	0.194	0.194
R	0.027	0.121	0.150	0.150	0.150
S	0.033	0.093	0.100	0.100	0.100
T	0.034	0.188	0.255	0.255	0.255
U	0.027	0.057	0.081	0.081	0.081
V	0.035	0.253	0.274	0.274	0.274
W	0.025	0.078	0.124	0.124	0.124
X	0.024	0.045	0.063	0.063	0.063
Y	0.029	0.084	0.112	0.112	0.112

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: In all cases, the outcomes were standardized to have mean 0 and standard deviation 1. The RCT impact standard errors are calculated using the formula $\sqrt{\frac{1}{N_T}+\frac{1}{N_C}}$, where N_T (the treatment group sample size) and N_C (the comparison group sample size) sum to the total sample sizes reported in Table 4, and N_T constitutes 25 percent of the total sample (to make the balance between the treatment and comparison groups the same for the RCT calculations as it is for the RDD calculations).

RCT = randomized controlled trial; RDD = regression discontinuity design.

Table 5b. Impact Standard Errors, RDD Cutoff at 50th Percentile

Outcome-Assignment Code	RCT, All Observations, No Clustering	RDD, All Observations, No Clustering	RDD, Observations in Bandwidth, No Clustering	RDD, Observations in Bandwidth, Clustering Identical	RDD, Observations in Bandwidth, Clustering Independent
A	0.019	0.036	0.066	0.066	0.066
B	0.031	0.062	0.082	0.104	0.124
C	0.036	0.057	0.089	0.089	0.089
D	0.031	0.061	0.110	0.110	0.110
E	0.037	0.060	0.090	0.090	0.098
F	0.028	0.049	0.071	0.071	0.071
G	0.030	0.049	0.076	0.083	0.126
H	0.050	0.086	0.127	0.127	0.131
I	0.029	0.049	0.092	0.113	0.138
J	0.038	0.068	0.084	0.084	0.936
K	0.043	0.078	0.128	0.169	0.217
L	0.045	0.079	0.151	0.191	0.204
M	0.040	0.072	0.132	0.132	0.151
N	0.044	0.078	0.252	0.897	0.956
O	0.029	0.086	0.109	0.109	0.109
P	0.027	0.063	0.091	0.091	0.091
Q	0.026	0.075	0.095	0.100	0.100
R	0.023	0.062	0.093	0.093	0.093
S	0.028	0.056	0.075	0.118	0.132
T	0.029	0.072	0.097	0.097	0.097
U	0.023	0.043	0.063	0.080	0.094
V	0.030	0.057	0.076	0.076	0.076
W	0.021	0.044	0.069	0.069	0.069
X	0.020	0.039	0.057	0.057	0.057
Y	0.025	0.048	0.061	0.061	0.061

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: In all cases the outcomes were standardized to have mean 0 and standard deviation 1. The RCT impact standard errors are calculated using the formula $\sqrt{\frac{1}{N_T}+\frac{1}{N_C}}$, where N_T (the treatment group sample size) and N_C (the comparison group sample size) sum to the total sample sizes reported in Table 4, and N_T constitutes 50 percent of the total sample (to make the balance between the treatment and comparison groups the same for the RCT calculations as it is for the RDD calculations).

RCT = randomized controlled trial; RDD = regression discontinuity design.

Table 5c. Impact Standard Errors, RDD Cutoff at 75th Percentile

Outcome-Assignment Code	RCT, All Observations, No Clustering	RDD, All Observations, No Clustering	RDD, Observations in Bandwidth, No Clustering	RDD, Observations in Bandwidth, Clustering Identical	RDD, Observations in Bandwidth, Clustering Independent
A	0.022	0.039	0.052	0.052	0.052
B	0.036	0.071	0.091	0.137	0.162
C	0.042	0.171	0.183	0.183	0.183
D	0.036	0.067	0.075	0.075	0.075
E	0.043	0.189	0.237	0.282	0.290
F	0.032	0.107	0.147	0.159	0.177
G	0.035	0.101	0.157	0.177	0.202
H	0.058	0.109	0.152	0.152	0.185
I	0.033	0.089	0.117	0.117	0.167
J	0.044	0.068	0.101	0.101	0.872
K	0.049	0.125	0.160	0.160	0.160
L	0.052	0.191	0.382	0.382	0.382
M	0.046	0.143	0.177	0.177	0.177
N	0.051	0.130	0.165	0.165	0.213
O	0.033	0.054	0.137	0.137	0.137
P	0.031	0.058	0.095	0.095	0.120
Q	0.030	0.049	0.083	0.083	0.096
R	0.027	0.045	0.078	0.078	0.086
S	0.033	0.051	0.081	0.081	0.081
T	0.034	0.051	0.074	0.074	0.074
U	0.027	0.046	0.067	0.067	0.067
V	0.035	0.067	0.106	0.137	0.150
W	0.025	0.040	0.062	0.062	0.062
X	0.024	0.042	0.066	0.066	0.066
Y	0.029	0.050	0.058	0.058	0.058

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: In all cases the outcomes were standardized to have mean 0 and standard deviation 1. The RCT impact standard errors are calculated using the formula $\sqrt{\frac{1}{N_T}+\frac{1}{N_C}}$, where N_T (the treatment group sample size) and N_C (the comparison group sample size) sum to the total sample sizes reported in Table 4, and N_T constitutes 75 percent of the total sample (to make the balance between the treatment and comparison groups the same for the RCT calculations as it is for the RDD calculations).

RCT = randomized controlled trial; RDD = regression discontinuity design.

Table 6. Total RDD Design Effect Relative to RCT, by Cutoff Value

Outcome-Assignment Code	25th Percentile		50th Percentile		75th Percentile	
	Identical	Independent	Identical	Independent	Identical	Independent
A	15.64	15.64	12.07	12.07	5.59	5.59
B	17.59	17.59	11.25	16.00	14.48	20.25
C	17.36	17.36	6.11	6.11	18.98	18.98
D	22.04	23.36	12.59	12.59	4.34	4.34
E	20.35	21.20	5.92	7.02	43.01	45.48
F	5.49	5.49	6.43	6.43	24.69	30.59
G	7.52	14.88	7.65	17.64	25.57	33.31
H	12.37	12.37	6.45	6.86	6.87	10.17
I	9.74	13.44	15.18	22.64	12.57	25.61
J	7.94	536.34	4.89	606.71	5.27	392.76
K	9.12	17.16	15.45	25.47	10.66	10.66
L	4.47	4.47	18.02	20.55	53.97	53.97
M	5.51	6.47	10.89	14.25	14.81	14.81
N	10.85	10.85	415.60	472.07	10.47	17.44
O	72.51	72.51	14.13	14.13	17.24	17.24
P	23.41	23.41	11.36	11.36	9.39	14.98
Q	41.82	41.82	14.79	14.79	7.65	10.24
R	30.86	30.86	16.35	16.35	8.35	10.15
S	9.18	9.18	17.76	22.22	6.02	6.02
T	56.25	56.25	11.19	11.19	4.74	4.74
U	9.00	9.00	12.10	16.70	6.16	6.16
V	61.29	61.29	6.42	6.42	15.32	18.37
W	24.60	24.60	10.80	10.80	6.15	6.15
X	6.89	6.89	8.12	8.12	7.56	7.56
Y	14.92	14.92	5.95	5.95	4.00	4.00
Median	**14.92**	**17.16**	**11.25**	**14.13**	**9.39**	**14.81**

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: The design effects are the ratios of the variance of RDD impacts to RCT impacts. RCT impacts are calculated using all available data and no clustering, whereas the RDD impacts are calculated using only students in the optimal bandwidth and a clustering adjustment to account for random misspecification error. Design effects are reported for two different clustering adjustments—one that assumes the specification error is unrelated to treatment status (identical) and one that assumes the opposite (independent). The last row reports the median of the design effects shown in each column.

RCT = randomized controlled trial; RDD = regression discontinuity design.

Table 7. Sample Sizes Needed for RCT and RDD Studies

Target MDES	RCT Sample Size	Assumed RDDE	RDD Sample Size
		9	2,250
0.25	250	14	3,500
		17	4,250
		9	3,530
0.20	390	14	5,490
		17	6,660
		9	14,110
0.10	1,570	14	21,950
		17	26,660

Source: Randomized controlled trials previously completed by Mathematica for IES.

Note: Sample sizes are rounded to the nearest 10 in accordance with NCES publication policy. Sample sizes were selected to minimize the distance between the target MDES and the quantity $fct(\alpha,\beta,df) * \sqrt{RDDE * \left(\left(1-R^2\right)\left(\frac{1}{N^T}+\frac{1}{N^C}\right)\right)}$, where *fct* is the sum of two critical values (corresponding to α =0.05 and β =0.20) from the T-distribution with *df* degrees of freedom; *RDDE* is the regression discontinuity design effect, R^2 is the proportion of variance in the outcome explained by covariates (assumed to be 0.50), and N_T and N_C are the numbers of students in the treatment and comparison groups (assumed to be equal).

MDES = minimum detectable effect size; RCT = randomized controlled trial; RDD = regression discontinuity design; RDDE = regression discontinuity design effect.

VI. CONCLUSION

Accounting for the correlation between the RDD assignment variable and treatment status, as shown in Schochet (2008a), results in an RDD impact variance between three and five times the size of an RCT impact variance with the same sample size. However, that design effect did not account for the sample loss associated with selecting an optimal bandwidth (as in I&K) or the effects of the L&C adjustment to standard errors that accounts for random misspecification bias. Using pre- and post-test score data from past education studies and supposing that the pre-test would be used as an RDD assignment variable with the post-test used as an outcome, we found that accounting for these necessary components of RDD impact analysis further increases the RDD impact variance to be 9 to 17 times higher than an RCT impact variance in a study with the same sample size. In the context of retrospective RDD studies using large administrative data sets, this may not necessarily be a concern. But in the context of prospective studies that must collect student data, the cost difference between these two designs could be substantial.

We emphasize that these findings may not apply to studies that do not use academic test scores as both the outcome and the assignment variable. Both the I&K bandwidth and the L&C adjustment depend on the relationship between the outcome and the assignment variable. In situations in which the relationship between the outcome and assignment variable is less linear, the bandwidth will be more narrow. In situations in which the relationship between the outcome and assignment variable is less smooth, the L&C adjustment will be more severe.

REFERENCES

Agodini, Roberto, Barbara Harris, Sally Atkins-Burnett, Sheila Heaviside, Timothy Novak, and Robert Murphy. "Achievement Effects of Four Early Elementary School Math Curricula: Findings from First Graders in 39 Schools (NCEE 2009-4052)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Bloom, Howard, Lashawn Richburg-Hayes, and Alison Black. "Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions." *Educational Evaluation and Policy Analysis*, vol. 29, no. 1, 2007, pp. 30-59.

Campuzano, Larissa, Mark Dynarski, Roberto Agodini, and Kristina Rall. "Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts." Report prepared for the U.S. Department of Education, Institute of Education Sciences. Princeton, NJ: Mathematica Policy Research, February 2009.

Constantine, J., D. Player, T. Silva, K. Hallgren, M. Grider, and J. Deke. "An Evaluation of Teachers Trained Through Different Routes to Certification, Final Report (NCEE 2009-4043)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Cook, T. "Waiting for Life to Arrive: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics." *Journal of Econometrics*, vol. 142, no. 4, 2008, pp. 636-654.

Deke, John, Lisa Dragoset, and Ravaris Moore. "Precision Gains from Publically Available School Proficiency Measures Compared to Study-Collected Test Scores in Education Cluster-Randomized Trials (NCEE 2010-4003)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010.

Hedges, L., and E. Hedberg. "Intraclass Correlation Values for Planning Group-Randomized Trials in Education." *Educational Evaluation and Policy Analysis,* vol. 29, no. 1, 2007, pp. 60-87.

Imbens, G., and K. Kalyanaraman. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." NBER Technical Working Paper Series: Technical Working Paper 14726. Cambridge, MA: National Bureau of Economic Research, 2009.

Imbens, G., and T. Lemieux. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, vol. 142, no. 2, 2008, pp. 615-635.

James-Burdumy, S., J. Deke, J. Lugo-Gil, N. Carey, A. Hershey, R. Gersten, R. Newman-Gonchar, J. Dimino, K. Haymond, and B. Faddis. "Effectiveness of Selected Supplemental Reading Comprehension Interventions: Findings from Two Student Cohorts (NCEE 2010-4015)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010.

Lee, D., and D. Card. "Regression Discontinuity Inference with Specification Error. *Journal of Econometrics*, vol. 142, no. 2, 2008, pp. 655-674.

Schochet, Peter Z. "Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Education Evaluations (NCEE 2008-4026)." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2008a.

Schochet, Peter Z. "Statistical Power for Random Assignment Evaluations of Education Programs." *Journal of Educational and Behavioral Statistics,* vol. 33, no. 1, 2008b, pp. 62-87.

U.S. Department of Education, What Works Clearinghouse. "Standards for Regression Discontinuity Designs (Version 1)." Washington, DC: ED, WWC, June, 2010.

ABOUT THE SERIES

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact John Deke, senior researcher, at jdeke@mathematica-mpr.com.

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ • Ann Arbor, MI • Cambridge, MA • Chicago, IL • Oakland, CA • Washington, DC

MATHEMATICA
Policy Research

www.mathematica-mpr.com