Gifted Child Quarterly
56(2) 105–112

DOI: 10.1177/0016986212444845
http://gcq.sagepub.com
SAGE

Regression Discontinuity Design in Gifted and Talented Education Research

Michael S. Matthews[1], Scott J. Peters[2], and Angela M. Housand[3]

Abstract

This Methodological Brief introduces the reader to the regression discontinuity design (RDD), which is a method that when used correctly can yield estimates of research treatment effects that are equivalent to those obtained through randomized control trials and can therefore be used to infer causality. However, RDD does not require the random assignment of individuals to treatment and control groups, making it very attractive for applied researchers in educational settings. This Brief introduces the method, discusses applications and limitations, and illustrates an idealized example as well as some potential pitfalls and their relevance to the context of gifted education research.

Keywords

regression discontinuity, gifted education, causal inference, permissive assignment, misallocation of treatment

A common limitation plagues all areas of research in education. Researchers desire to draw causal inferences from studies in order to evaluate the potential influence of an intervention. However, simple correlation studies (used in approximately two thirds of quantitative gifted education research, according to Matthews et al., 2008) cannot demonstrate causality. Random assignment of individuals to treatment or control groups is considered among the strongest of research designs, precisely because it does allow one to make causal inferences; unfortunately, random assignment can be inappropriate and sometimes even unethical in educational research. For example, it would be inappropriate to assign highly proficient readers to a remedial Reading Recovery program solely for the purpose of research.

When assessing the effects of gifted programs and services, researchers cannot randomly assign students to receive special services, because the assignment to gifted programs is typically based on the demonstrated advanced capabilities or special needs of the students. In other words, the group selection process tends to homogenize groups, whereas the goal in random assignment is to create groups of equivalent heterogeneity. This may explain the use of alternatives to experimental studies, such as using matching strategies for assignment to experimental and control groups in gifted education research (Fan & Nowell, 2011; Goldring, 1990, Marsh, Chessor, Craven, & Roche, 1995). Unfortunately, matching strategies often produce biased results (Marsh, 1998). This may be due, in part, to the fact that researchers often match students based on variables that are easy to measure, such as standardized test scores, thereby precluding the inclusion of more complex factors (e.g., motivation, perseverance, goal orientation, mindset, etc.) in the matching process. Matching strategies also suffer from inherent statistical problems, such as differential regression to the mean and effects from bias, which become larger when assignment scores are more extreme (Goldring, 1990; Marsh, 1998).

Even when researchers have a good match between participants and a program, they may find their work constrained by institutional review board policies, schools, and even their own ethical principles in ways that preclude randomly assigning students to treatment or control groups. One common way to deal with this problem is to assign students to treatment or control groups for a period of time, but to then make the treatment available to all students either during a subsequent year of the project or on its conclusion. Although this approach can address a portion of the ethical problem, it may not satisfy school boards or school administrators, who are understandably reluctant to prioritize the treatment fidelity of a research study over the quality of students' education.

Random assignment is done to provide counterfactual evidence. Counterfactual evidence is "knowledge of what would have happened to those same people [study participants] if they simultaneously had not received treatment" (Shadish, Cook, & Campbell, 2002, p. 5). Random assignment creates two groups that are probabilistically the same; in

[1]University of North Carolina at Charlotte, Charlotte, NC, USA
[2]University of Wisconsin–Whitewater, Whitewater, WI, USA
[3]University of North Carolina–Wilmington, Wilmington, NC, USA

Corresponding Author:
Michael S. Matthews, Department of Special Education and Child Development, University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, USA
Email: michael.matthews@uncc.edu

other words, they yield the strongest counterfactual evidence (e.g., they tell us what would have happened) by eliminating the vast majority of alternate causes to explain an observed effect.

Unfortunately, as already mentioned, random assignment is not always possible. Luckily, developments over the past 20 years in the understanding and application of regression discontinuity research designs (RDD) can yield some of the same benefits, and studies have shown RDD methods are able to yield unbiased estimates of treatment effects much like randomized controlled trials (Cappelleri, Trochim, Stanley, & Reichardt, 1991; Cook, 2008; Shadish, Galindo, Wong, Steiner, & Cook, 2011; Robinson & Stanley, 1989). To determine the treatment effect statistically, a regression equation of the following basic form is used in RDD:

$$Y = b_0 + b_1 Xc + b_2 T + e, \tag{1}$$

where Y is the outcome variable; b_0 is the intercept of the nontreated (control) group; $b_0 + b_2$ is the intercept of the treated group; b_1 is the marginal effect of scoring one point higher on the assignment variable, Xc, on the outcome variable, Y (sometimes referred to as the parallel, or intercept, shifter); b_2 is the estimated treatment effect; Xc is the quantitative assignment variable when centered at the cutoff; T denotes whether or not an individual receives the treatment; and e is the error in prediction (Trochim, 1984; West, Biesanz, & Pitts, 2000).

In practice, the equation often becomes more complex as researchers introduce polynomial terms or more elaborate models, but this basic form provides a general framework on which to build an understanding of the RDD approach. We urge the reader to investigate the many excellent resources that delve into the more complex aspects of RDD models, which space considerations preclude us from addressing in this article.

In an RDD, the researcher assigns participants to groups based on their score on a certain pre-intervention assignment variable (denoted as X in Equation 1). This variable can be anything that is fully known, is determined prior to treatment, and is a quantifiable variable in which individuals fall on one side or another of a cutoff score. It may be a composite score derived from several assignment variables, as we discuss later.

When the assignment variable is on an interval or ratio scale, the actual source of the measure and even its internal validity are irrelevant. For example, achievement, IQ, or aptitude tests would be logical choices, but even hair length or height could be used *because the variable is being used solely for assignment purposes*. A key point to understand is that the assignment variable does *not* need to be any kind of predictor or covariate; in fact, it may be completely unrelated to the outcome of interest, so long as it meets the criteria of being fully known and being determined prior to treatment. In RDD, *the assignment variable is conceptualized as having no error at all* because it is being used solely for assignment, just as if placement into treatment or control conditions were done randomly.

RDD in Gifted Education Settings

RDD and Identification

When Tom Cook spoke at the Business Meeting of the AERA SIG, Research on Giftedness, Creativity, and Talent in Chicago in 2007, some attendees expressed concern that the apparent hard-line reliance on a single assignment variable violated two inherent beliefs among gifted education researchers–that identification decisions must not be made based on a single cut score and that multiple measures should be used in the identification process. In fact, nothing about RDD requires that the assignment variable cut score be limited to a single test or measure. For example, Lohman and Renzulli (2007) proposed an approach to identification combining achievement test scores, aptitude test scores, and teacher ratings to place individuals on one side or another of a composite cut score. Such a combination of multiple measures into a single "score" to be used as an assignment variable in RDD is perfectly acceptable.

An important additional requirement in RDD is that the assignment variable should include sufficient variability to detect the relationships between treatment and control groups. Of particular relevance to gifted education research is that the model can be harmed if the assignment variable cutoff is an extreme score (e.g., the 98th percentile on an achievement test), simply because there are so few observations above that point. This is the same problem as is the case in any regression analysis where there are too few observations (i.e., the standard error increases). However, this problem can be mitigated in gifted education research settings such as talent search programs (Matthews, 2008), where large populations of students having extreme scores are concentrated (Lee, Matthews, & Olszewski-Kubilius, 2008). The use of an above-level or "off-level" test that spreads out the score distribution of the top few percentile ranks, as done in talent search testing, would also offer an effective means of distributing observations of high-ability learners across both sides of a cutoff point.

An Illustrative Example

Regression discontinuity design has the potential to be effective in nearly any situation in gifted education research so long as its few premises are satisfied (Stanley & Robinson, 1986). For example, imagine that we want to test whether or not Total School Cluster Grouping (Gentry & Mann, 2009) increased the achievement test scores of students as measured by the Northwest Evaluation Association's Measure of Academic Progress (MAP) test for those individuals in the "gifted" cluster. Let us also assume that students are only

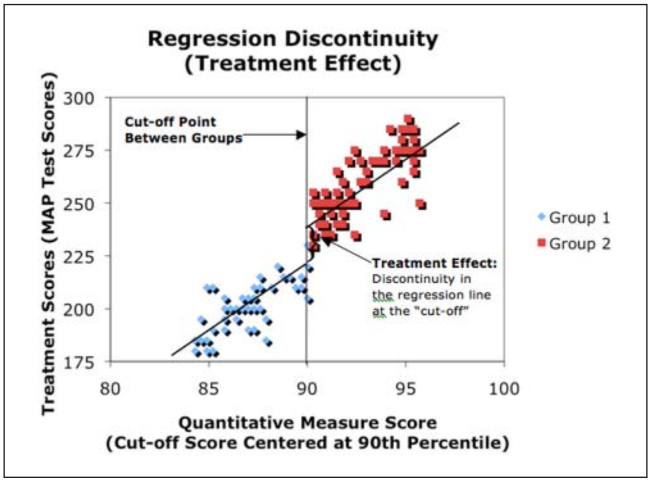

Figure 1. Hypothetical RDD plot illustrating a simple treatment effect

going to be entered into this particular cluster (independent variable) if they score above the 90th percentile on the MAP as a pre-intervention test. Those students who meet the pre-intervention cut score are placed in the program and those who do not are considered the control group. The intervention would then take place for a period of time and then the dependent variable (MAP) would be administered again. Remember, this hypothetical cut score at the 90th percentile is for discussion purposes only. Because this arbitrary cut score criterion of the 90th percentile is not too extreme, there are likely to be sufficient students on both sides of this cut score (depending on the sample, of course). In addition, because the MAP is a computer adaptive test, ceiling effects would be minimal in theory. In other words, the 90th percentile for students in fourth grade would still leave a range of test takers scoring above and below this cut score. Higher cut scores could be used, but parameter estimates become less and less accurate as the cut score became more extreme, due to correspondingly fewer observations falling above the cutoff score.

The reason that RDD works as a way to determine causal inference is that it draws on the full range of data collected to evaluate a discontinuity (or parallel shift) in the area around the cut score. This allows for a comparison of individuals who are similar, but who received different treatments–just as is the case with randomized control/treatment (RCT) designs. There can be cases that are far more complex, of course, but this is the most straightforward scenario (see Figure 1).

Figure 1 presents a hypothetical general plot (output) of such an RDD analysis. Individuals who were in the treatment group appear on the right of the vertical line at the cut score; control group individuals are plotted to the line's left. Here we have depicted only a few cases for clarity; in an actual data set, there likely would be hundreds or thousands of data points in each group. In this simple hypothetical example, the slope of the regression lines on either side of the cut score does not differ; only the intercept at the cut score differs. The graph depicts a clear effect of the treatment, which is visible as the difference in intercept between the parallel regression lines on either side of the cut score.

Effects in such RDD plots can be seen as differences in intercepts, shifts, or both. This difference is quantifiable and can be tested for statistical significance using traditional ordinary least squares regression. In a case such as Figure 1 depicts, the researcher would now have nearly identical counterfactual information as if a randomized control trial had been conducted. In addition, in the RDD approach, far more confounding variables are controlled for than in other quasi-experimental or observational techniques. This allows for a much stronger argument for causal inference; however, the RDD approach comes at a cost in terms of sample size, as we describe in the section that follows.

Avoiding Violations of RDD Assumptions

Although RDD may seem like the perfect solution for research in gifted education, there remain a few challenges associated with the nature of research in the social sciences in general and field research in education specifically. In the identification of academically gifted learners, standardized testing is still widely used. This is likely due to the fact that it is more efficient than other identification methods available, at least in terms of the combined considerations of validity, reliability, and cost effectiveness. Fortunately, as Shadish et al. (2002) point out, "The assignment variable can be a pretest on the dependent variable . . . RDD works whether the assignment variable is or is not related to the outcome" (pp. 216-217).

RDD requires larger samples. Of course, there are a few additional complexities. As a general statement, RDD is less statistically powerful than RCT (Trochim, 1984; West et al., 2000). The underlying cause of this decrease in power is the high correlation that exists between the assignment variable (X) and treatment (T) and the treatment's interaction with its respective regression coefficient (b_2; Schochet, 2009). Because whether or not an individual receives the treatment in RDD is due in large part, if not exclusively, to the assignment variable, a relationship now exists between these two factors that would not exist in an RCT design (where treatment status is decided at random). This correlation enters into the overall standard error estimation for the RDD model in such a way that as the correlation increases, so does the standard error; this necessitates larger sample sizes to obtain similar levels of significance. Estimates suggest that RDD requires from 2.75 (Trochim, 1984; West, Biesanz, & Pitts, 2000) to as many as four times (Schochet, 2009) the sample size needed for an equivalent RCT design, in order to yield equivalent statistical power. However, one potential approach to address this issue is to decrease artificially the relationship between the assignment variable and treatment status by making the assignment variable a composite of multiple measures. In addition to the benefit of increasing power via decreasing the key correlation, this approach also has the benefit of aligning with best practice in gifted education identification regarding the use of multiple measures. When working with large secondary data sets, sample sizes are usually well over the minimum needed to ensure adequate power, so this potential limitation of RDD may be less problematic in practice than the other caveats we describe.

Misallocation of treatment. Perhaps the most difficult RDD assumption to satisfy in applied research in gifted education settings, regardless of the measure or measures used to establish a cut score, is the requirement that the assignment process be fully known. The makeup of the X variable from Equation 1 must be clearly understood and well established, with few or no violations of these criteria. Unfortunately, in practice the rules for assignment to gifted programming (i.e., identification) are bent more often than school districts would like to admit (Davis, Engberg, Epple, Sieg, & Zimmer, 2010). This may be done with the best of intentions, in order to avoid denying services to students who might benefit from treatment even though their performance fell slightly short of the cut score requirement. This type of overly inclusive identification in a program or treatment in gifted education has been referred to as *permissive assignment*, and it often may be observed in practice as a lower than expected number of students scoring just below the cutoff. What typically has happened in such cases is that identification committees, program administrators, or even school psychologists have subjectively added points or simply "rounded up" an individual's score in order for that individual to meet the cut score criterion (Matthews, Kirsch, Shaunessy, & Fahmy, 2007).

Though at first it may seem that the problem lies in the relationship between the identification variable and the outcome measure, in RDD this relationship does not matter. Actually, such rule-bending violates the assumption that the assignment process be fully known and based solely on the assignment variable; violation of this assumption of RDD reintroduces the sampling biases that RDD is being used to avoid.

In the language of RDD, permissive assignment would fall under the general category of error known as *misallocation of treatment*, which is one of several types of errors that may plague attempts to apply RDD in real-world settings. Misallocation of treatment happens when the rule for assigning individuals to groups fails to place all participants properly, due to grader reluctance to assign scores just below the cutoff (Shadish et al., 2002) or other situations in which professional judgment may override the strict application of cutoff score criteria. The same authors suggest that teacher subjectivity and the "bending of the rules" must be kept to an absolute minimum, ideally less than 5% of cases, in order to minimize its adverse effects (Shadish et al., 2002; Trochim, 1984).

There do not appear to be many suggestions in the literature for addressing misallocation of treatment after it has occurred. One option suggested by Davis et al. (2010) in the context of gifted education is to use IQ test subscale scores that were not part of the gifted identification process (and therefore, presumably had not been subject to improper manipulation) in order to estimate corrected full IQ scores for use in place of the original IQ scores that these authors suspected had been manipulated. Though Lohman and colleagues have pointed out some cautions related to the use of subscale scores (e.g., Lohman, Gambrell, & Lakin, 2008), further study of this approach seems warranted.

Specifying the functional form. RDD also assumes the researcher knows the functional form of the relationship between the assignment and outcome variables. In Equation 1, we assumed a linear relationship for the sake of a clear presentation. However, if the researcher assumes a linear relationship but in fact this relationship does not fit the data (such as when the true relationship is curvilinear—requiring a polynomial term in the equation in addition to the linear one), the model becomes less accurate in the same fashion as it would in a traditional regression analysis. For example, when in the true relationship the outcome variable follows a squared function (x^2) of the assignment variable, the result is a curvilinear relationship. If the researcher models this incorrectly as a linear relationship, the result could be a finding of a discontinuity where none actually exists.

In the simplest approach, nonlinearity can be addressed by the inclusion of squared or cubic functions into the general RDD equation in an attempt to better fit the model to the data. In general, Shadish et al. (2002) recommend overfitting the model any time nonlinearity is suspected. Overfitting refers to evaluating several models with varying levels of complexity, then evaluating each model's fit in comparison with the fit of the others. In this manner, the researcher can locate the model that best fits the data. Oftentimes this will address the issue and allow for greater confidence in the results than when only a single model is presented (Cook & Wong, 2008; Wong, Cook, Barnett, & Jung, 2008).

In the case of measurement error, enhancements can be made to the RDD to address nonlinearity. As stated previously, researchers should always start with a clear understanding of the functional form of the pretest data used for assignment to treatment or control, as examination of these allows one to detect any preexisting deviations from linearity (Cook & Wong, 2008; Trochim, Cappelleri, & Reichardt, 1991). When deviations are identified, the researcher has at least two options for addressing them. One option is to assign two cutoff[1] values. In this approach, students scoring above the greater of the two cutoff scores might be assigned to a treatment condition, whereas students below the lower cutoff score would be assigned to the control. Students falling between the two cutoff scores would then be randomly assigned to treatment or control. This ensures that for the range of X occurring between the two cutoff scores, there is a randomization of treatment and corresponding comparison group scores (Trochim et al., 1991). Although perhaps not entirely palatable in practice, this approach offers one way to minimize the statistical problems posed by permissive assignment in the gifted education context; it also would be particularly appropriate in settings where there are more seats available in gifted programming than there are formally identified learners to fill them.

A second variation that can be used to ameliorate functional form misspecification is to have two waves of pretreatment testing. This approach assumes that the functional form does not change between the two pretest scores, because as of yet there has been no intervention. This allows a comparison between two waves of testing. The comparison of the first two waves suggests the functional form of the relationship in the case where no treatment effect occurred. Then, once the treatment is administered and posttest scores are available, there is sufficient information to determine the functional relationship between the two measures in both cases (i.e., modeling both with intervention effects and without them). The two models can then be compared for discrepancies and the discrepancies (if any) used to identify the effect due to the treatment (Trochim et al., 1991). Applying these two approaches appropriately can minimize the risk of functional form misspecification.

One of the greatest functional form issues, and one endemic to the application of RDD to gifted education research, deals with ceiling effects. Ideally, to yield the most statistically powerful model, the cut score should be near the mean of the assignment variable distribution. However, setting a cutoff score near the median or mean is not always practical, and sometimes such placement does not fit with the underlying theory (such as with research on students with gifts and talents, who by definition are drawn from one extreme of the population distribution). Ceiling effects can be moderately problematic and cause a simple decrease in power, or can be so problematic as to prevent the estimation of the regression line (in cases where very few points lie to the right of the cut score). Because of this issue, assignment variables should yield a range of variability on both sides of the cut score. Unfortunately, in gifted education, assignment cut scores may be set at the 95th or even 99th percentile (see previous footnote), in which case very few students would fall to the right of the cut score.

When such instances occur, there may be little that can be done except to collect more data to increase the number of individuals to the right of the cut score. The application of a Tobit model correction (McBee, 2010) may also be possible in this situation, though research does not yet appear to have evaluated this approach.

The best solution to address ceiling effects, which would work well in current gifted education practice but which would need to be implemented prior to data collection, is the application of above-level tests. This practice is common in Talent Search programs, where, for example, a test typically given to eighth-grade students is given to high-ability students in Grades 3 through 6 (Lee et al., 2008; Matthews, 2008). As mentioned previously, in above-level testing, gifted and talented students (those who might score at the 95th percentile or higher on chronological-age norms) score closer to the traditional mean level on the given test (Bleske-Rechek, Lubinski, & Benbow, 2004). This results in scores that are distributed more evenly above and below the cutoff, thereby facilitating stronger power and more accurate estimation of the regression line. The same benefit could be realized using computer adaptive testing, in which ceiling effects are greatly minimized due to the substantial increase in range that most computerized adaptive tests (such as the MAP, as described in our example) offer in comparison with traditional paper-and-pencil achievement tests.

Process

Despite these limitations, RDD remains an appealing approach because it allows causal inference. Another characteristic that makes RDD attractive is that it does not require special software or extensive training or knowledge, beyond that required for conducting a traditional regression analysis. IBM® SPSS® can be set up to analyze a data set using the RDD model, as can virtually all other statistical packages. For a step-by-step description of how to use SPSS for this purpose, we encourage the reader to refer to the helpful guide by Wuensch (2010). Cook et al. (2010) also have posted a detailed handout on RDD from their Workshop on Quasi-Experimental Research and Analysis.

The process for conducting an RDD analysis involves a series of steps (Cook et al., 2010; Wuensch, 2010). First, subtract the cutoff from all participant scores to recenter the cutoff score to zero. Next, set a dummy variable to indicate treatment or control group membership for those individuals below and above the cut score (e.g., $T = 0$ for control; $T = 1$ for treatment). Produce a scatter plot of the data to look for a discontinuity around the cut score, using a different marker for each group (below and above the cut score). This will produce a graphic, similar to Figure 1, in which the researcher can visually inspect the data. However, it is important to note that visual inspection alone may not be sufficient; Shadish et al. (2002) note that it is very rare for a visually apparent discontinuity to be evident at this stage. Because of this, researchers next will need to test the independent parameters and covariates for statistical significance. Importantly, though at first glance it may seem reasonable to estimate regression lines and intercepts separately for each group, doing so will not allow calculation of the standard error that is needed to calculate the statistical significance of the gap observed at the discontinuity (cut score). Rather, the correct way to calculate the treatment effect and its standard error is to run a multiple regression across both groups, using the identification variable score and a dummy variable (in Equation 1, this would be $T = 0/1$ as an indicator of group membership) as the outcome variables. Then, the regression coefficient on T indicates the treatment effect and standard statistical regression procedures will provide its SE and associated p value.

Researchers should also create and evaluate an interaction term (using, e.g., the GLM Univariate command in SPSS) to examine the data for the presence of a significant interaction. As stated earlier, it is better to include and test for interaction terms rather than to assume that the functional form is linear with no interactions. If there is a significant interaction term, this indicates that the slope of the regression line in the group below the cutoff differs in some manner from the slope of the line for cases above the cutoff score. This might be the case in gifted education research under the so-called Matthew effect (Merton, 1968), which suggests that higher-ability individuals progress at a faster rate; such a case might show no difference in intercept but a clear difference in slope within or across treatment groups. Though too complex to explain here, the choice of bandwidth also is relevant at this point in the analysis; see discussion of this issue by Cook and Wong (2008).

Once identified, nonlinear and interaction terms can then be tested for significance to see if they help explain differences in the data. Next, all independent predictor terms (treatment dummy variable, interaction terms, etc.) are entered into the multivariate equation with the outcome variable score as the dependent variable. This will allow the researcher to determine the significance of the treatment variable in explaining any existing discontinuity. If the treatment was not effective (e.g., in our example, if TSCG does not have an effect), then the treatment term will be nonsignificant.

Conclusion

Although a number of challenges are evident as we conceptualize the application of RDDs to research in gifted education, a substantial amount of progress has been made in recent years in identifying and correcting RDD approaches to allow better inferences to be made. Some limitations, most notably ceiling effects and the issue of misallocation of treatment in the form of permissive assignment, are particularly salient in the gifted education context. Although for decades meaningful research has been conducted using regression discontinuity analysis with gifted populations, this work has not been widely cited (Karnes, Shwedel, & Lewis, 1983; Robinson, Bradley, & Stanley, 1990). Future studies should seek to address and examine approaches designed to ameliorate limitations that exist when RDD is applied to gifted education settings, which most recently has been done primarily in the context of economics (Bui, Craig, & Imberman, 2011; Davis et al., 2010).

As multiple-criteria identification systems continue to be developed and evaluated, the prevalence of permissive assignment may decrease due to increased perceptions of fairness in the identification process. In addition, we believe that small scale (e.g., 2% to 3%) permissive assignment is not necessarily a bad thing from the standpoint of both RDD theory and gifted education practice. It can be tolerated by RDD (Shadish et al., 2002), while allowing for the inclusion of some students who show a need but still may not qualify for services under even the best-designed identification systems. Our experience suggests that small-scale permissive assignment will probably always exist as a kind of appeals system in response to the natural limitations of any kind of identification procedure.

Regression discontinuity design can yield causal inferences in a similar fashion to RCT studies without the requirement for randomized assignment to treatment and control groups. Despite the potential limitations and complexities of doing research in the social sciences, RDD still may be the best solution for establishing causal inference within the limitations posed by the school-based research context. With requirements from the Institute of Education Sciences and other funding agencies leaning more and more toward supporting only those methods that can establish causation, observational and correlational methods are beginning to fall by the wayside. Gifted education research could benefit from the application of methods that permit causal inference, such as RDD, when investigating the effects of gifted education programming and services.

Acknowledgments

The authors thank Peter Steiner and two anonymous reviewers for their feedback on earlier drafts of this article.

Authors' Note

Much of the information presented in this article came from two of the authors' attendance at the Workshop on Quasi-Experimental Research and Analysis in 2008 and 2010. The authors wish to thank Tom Cook, Will Shadish, the Spencer Foundation, and the Institute for Educational Sciences for offering these workshops, and they would encourage any gifted education researchers interested in RDD to apply for future Workshops.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. It is beyond the scope of this brief to discuss the setting of cut scores. Here we assume that such scores have been determined appropriately, through a standard setting process based on local policies and available programming.

References

Bleske-Rechek, A., Lubinski, D., & Benbow, C. P. (2004). Meeting the educational needs of special populations: Advanced Placement's role in developing exceptional human capital. *Psychological Science, 15*, 217-224.

Bui, S. A., Craig, S. G., & Imberman, S. A. (2011, May). *Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement* (Working Paper No. 17089). Cambridge, MA: National Bureau of Economic Research.

Cappelleri, J. C., Trochim, W. M. K., Stanley, T. D., & Reichardt, C. S. (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: I. The case of no interaction. *Evaluation Review, 15*, 395-419.

Cook, T. D. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics, 142*, 636-654. doi:10.1016/j.econom.2007.05.002

Cook, T. D., et al. (2010). *Quasi-experimental research and analysis* [Document]. Retrieved from http://www.northwestern.edu/ipr/events/workshops/QEPPTs/QE-Day2.pdf

Cook, T. D., & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity design. *Annals of Economics and Statistics, 91/92*, 127-150.

Davis, B., Engberg, J., Epple, D. N., Sieg, H., & Zimmer, R. (2010, September). *Evaluating the gifted program of an urban school district using a modified regression discontinuity design* (Working Paper No. 16414). Cambridge, MA: National Bureau of Economic Research.

Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly, 55*, 74-79. doi:10.1177/0016986210390635

Gentry, M., & Mann, R. L. (2009). *Total school cluster grouping and differentiation*. Mansfield Center, CT: Creative Learning Press.

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *Journal of Educational Research, 83*, 313-326.

Karnes, M. B., Shwedel, A. M., & Lewis, G. F. (1983). Short-term effects of early programming for the young gifted handicapped child. *Exceptional Children, 50*, 103-111.

Lee, S.-Y., Matthews, M. S., & Olszewski-Kubilius, P. (2008). A national picture of talent search and talent search educational programs. *Gifted Child Quarterly, 52*, 55-69. doi:10.1177/0016986207311152

Lohman, D., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science Quarterly, 50*, 269-282.

Lohman, D., & Renzulli, J. (2007). *A simple procedure for combining ability test scores, achievement test scores, and teacher ratings to identify academically talented children*. Retrieved from http://faculty.education.uiowa.edu/dlohman/pdf/Lohman_Renzulli_ID_system.pdf

Marsh, H. W. (1998). Simulation study of nonequivalent group matching and regression-discontinuity designs: Evaluations of gifted and talented programs. *Journal of Experimental Education, 66*, 163-192. doi:10.1080/00220979809601402

Marsh, H. W., Chessor, D., Craven, R. G., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal, 32*, 285-319.

Matthews, M. S. (2008). Talent search programs. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: What the research says* (pp. 641-653). Waco, TX: Prufrock Press.

Matthews, M. S., Gentry, M., McCoach, D. B., Worrell, F. C., Matthews, D., & Dixon, F. (2008). Evaluating the state of a field: Effect size reporting in gifted education. *Journal of Experimental Education, 77*, 55-68. doi:10.3200/JXE.77.1.55-68

Matthews, M. S., Kirsch, L., Shaunessy, E., & Fahmy, N. R. (2007). *Diversity within diversity: Understanding our linguistically diverse gifted students*. Paper presented at the 2007 annual conference of the National Association for Gifted Children, Minneapolis, MN.

McBee, M. T. (2010). Modeling outcomes with floor or ceiling effects: An introduction to the Tobit model. *Gifted Child Quarterly, 54*, 314-320. doi:10.1177/001698621037909

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*, 56-63.

Robinson, A., Bradley, R. H., & Stanley, T. D. (1990). Opportunity to achieve: Identifying mathematically gifted Black students. *Contemporary Educational Psychology, 15*(1), 1-12.

Robinson, A., & Stanley, T. D. (1989). Teaching to talent: Evaluating an enriched and accelerated mathematics program. *Journal for the Education of the Gifted, 12*, 253-267.

Schochet, P. Z. (2009). Statistical power for regression discontinuity designs in educational evaluations. *Journal of Educational and Behavioral Statistics, 34*, 238-266.

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

Shadish, W. R., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D (2011). A randomized experiment comparing random and cut-off-based assignment. *Psychological Methods, 16*, 179-191. doi:10.1037/a0023345

Stanley, T. D., & Robinson, A. (1986). Regression discontinuity: Integrating research and program design in programs for the gifted. *Journal for the Education of the Gifted, 9*, 181-191.

Trochim, W. M. K. (1984). *Research design for program evaluation: The regression discontinuity approach*. Beverly Hills, CA: Sage.

Trochim, W. M. K., Cappelleri, J. C., & Reichardt, C. S. (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: II. When an interaction effect is present. *Evaluation Review, 15*, 571-604.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40-84). New York, NY: Cambridge University Press.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*, 122-154. doi:10.1002/pam.20310

Wuensch, K. L. (2010). *Using SPSS to analyze data from a regression-discontinuity design*. Retrieved from http://core.ecu.edu/psyc/wuenschk/spss/RegDiscont-SPSS.doc

Bios

Michael S. Matthews is an associate professor of gifted education and graduate coordinator for the Academically & Intellectually Gifted program at the University of North Carolina at Charlotte. In 2010, he was awarded the NAGC Early Scholar Award. He is coeditor of the *Journal of Advanced Academics* and is an active member of the Research & Evaluation network of NAGC; treasurer of the SIG-Research on Giftedness, Creativity, & Talent of the American Educational Research Association; and a board member with the North Carolina Association for the Gifted & Talented. He is author of three books and more than 15 peer-reviewed articles in gifted education. His research interests include gifted education policy, gifted education in science and mathematics, and the assessment and identification of learners from diverse backgrounds, especially those who are English language learners.

Scott J. Peters is an assistant professor of educational foundations at the University of Wisconsin–Whitewater where he teaches courses related to educational measurement, research methods, and gifted and talented education. His research interests include educational research methodology in addition to gifted and talented student identification, nontraditional giftedness, and secondary student programming outcomes. He has presented at the state, national, and international levels on topics dealing with gifted education and research methodology and currently serves on committees and as an officer of the National Association for Gifted Children and the American Educational Research Association. His scholarly work has appeared in *Teaching for High Potential, Gifted Child Quarterly*, the *Journal of Advanced Academics, Gifted and Talented International*, the *Journal of Career and Technical Education Research, Ed Leadership*, and *Pedagogies*.

Angela M. Housand, PhD, is the executive director of the Consortium for Research on Educational Accountability and Teacher Evaluation and an assistant professor at the University of North Carolina–Wilmington where she develops and teaches courses for the Academically and Intellectually Gifted Teacher Licensure Program. Her research involves studying the behavioral characteristics of gifted and profoundly gifted individuals and the effect that environmental factors have on emotion regulation and behavior.