2016-05-01

A Multiple-Cutoff Regression-Discontinuity Analysis of the Effects of Tier 2 Reading Interventions in a Title I Elementary School

Eli A. Jones
Brigham Young University

A Multiple-Cutoff Regression-Discontinuity Analysis of the Effects of Tier 2 Reading Interventions in a Title I Elementary School

Eli A. Jones

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard R Sudweeks, Chair
Erika Feinauer
K. Richard Young
Gordon S. Gibb
Ross Larsen

Educational Inquiry, Measurement, and Evaluation

Brigham Young University

May 2016

ABSTRACT

A Multiple-Cutoff Regression-Discontinuity Analysis of the Effects of Tier 2 Reading Interventions in a Title I Elementary School

Eli A. Jones
Educational Inquiry, Measurement, and Evaluation, BYU
Doctor of Philosophy

Reading failure in elementary school is highly correlated with future academic and social problems. Schools commonly use Tier 2 reading interventions in Response to Intervention (RtI) frameworks to help close the gap between at-risk readers and their peers who read on grade-level. This dissertation presents the findings of a quasi-experimental research study of the effects of three Tier 2 reading interventions in an urban Title I elementary school's RtI framework.

A regression discontinuity design (RDD) with two cutoff points was used to assign 320 students in grades 1-6 to two types of Tier 2 reading interventions administered by paraeducators: direct instruction (DI) and computer-assisted instruction (CAI). Students were assigned using normal curve equivalent reading composite scores on the Kaufman Test of Educational Achievement II, Brief Form (KTEA-II BFR). Students scoring below a lower cutoff were assigned to a DI reading intervention, while students scoring at or below an upper cutoff and above the lower cutoff were assigned to CAI reading interventions. January and May posttest iterations of the KTEA-II BFR served as outcome measures for all students. Results of the analysis indicated that the DI intervention was more effective than the CAI interventions at the lower cutoff ($p < .01$). Participation in CAI interventions was not any more or less effective than business-as-usual reading activities ($p > .10$). These findings suggest that that CAI programs may not be as helpful in closing the achievement gap between struggling students and their peers as DI interventions, and should be implemented with deliberation.

Keywords: computer-assisted instruction, direct instruction, English language learners, reading interventions, regression discontinuity design, response to intervention.

ACKNOWLEDGMENTS

Helen Keller once said, “I would rather walk with a friend in the dark, than alone in the light.” I wish to express my gratitude to the many friends and colleagues who walked with me through every stage of my doctoral studies leading up to the completion of my dissertation. First, thank you to my advisor and mentor Dr. Richard Young who encouraged me to pursue this topic and enabled my success with his concern, generosity, and ever-present support. Thank you also to Dr. Richard Sudweeks for providing clarity and understanding, to Dr. Gordon Gibb for his energy and cheerful encouragement, to Dr. Ross Larsen for his willingness to answer my questions at a moment’s notice, and to Dr. Erika Feinauer for her insights. A big thanks to Leslie Williams, Laurie Ottehenning, A. J. Griffin, Hayley Roberts, Aislin Pickett, and the other members of the research staff without whom this dissertation would not have come to fruition. Finally, I wish to thank my wonderful wife/friend/editor/helper, April, and my patient children (Andrew, Ben, Evan, Caleb, and Jay) who offered their unfailing love, support, and encouragement throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

Chapter 1. Introduction

One of the key concerns that schools face is how to respond to the needs of students who struggle in reading in order to provide them the support needed to become proficient readers. Reading failure is highly correlated with increased high school dropout rates, increased poverty, and unemployment (Jennings, Caldwell, & Lerner, 2013). Certain subpopulations of students who read below grade level are particularly at risk for reading failure. These include students for whom English is a second language, commonly known as English Language Learners (ELLs).

In recent decades, a strategy known as Response to Intervention (RtI) has become a standard practice in responding to sub-standard reading performance (Berkeley, Bender, Peaster, & Saunders, 2009; Fuchs, Mock, Morgan, & Young, 2003). With its tiered levels of support and data-driven methods, RtI provides a model structure for schools to follow in order to differentiate instruction according to student need. While RtI has been extensively adopted, some concern exists over how well schools implement their individual RtI programs effectively (Keller-Margulis, 2012; Lane, Bocian, MacMillan, & Gresham, 2004).

Problem Statement

Although RtI has become a standard practice in many schools, thorough experimental evaluation of the effects of RtI programs is rare (Swanson, 2012; Tran, Sanchez, Arellano, & Swanson, 2011). Hughes and Dexter (2011) conducted a meta-analysis of research studies evaluating the effects of RtI programs over the past decade. They found 13 studies that evaluated the effectiveness of multi-tier, multi-component RtI programs. These field studies included single-case studies, historical studies, quasi-experimental studies, and descriptive studies. However, the researchers noted the fact that many of the studies were not designed appropriately to control for threats to internal or external validity or to support strong causal claims. They concluded that “more longitudinal efficacy research is needed, as well as examination of factors necessary for developing and sustaining RtI, to assist educators as they consider adoption of this approach” (Hughes & Dexter, 2011, p. 10).

A hurdle in estimating treatment effects in reading interventions is that, while beneficial for controlling for threats to validity, randomized experiments are not conducive to the nature of RtI programs. Shadish, Cook, and Campbell (2002) note that randomized experiments can be unethical in situations where randomization withholds treatment from individuals in need of such treatment. In the case of reading interventions, it is clearly unethical to withhold treatment from students that are in need of reading help. This can lead to the conundrum where students receive interventions but the school is unable to accurately evaluate the effectiveness of its RtI program.

Last year, researchers from Brigham Young University were asked by the administration of a Title I urban elementary school to assist them in evaluating the effectiveness of their RtI reading program. During the 2013-2014 school year, the school added an intervention developed by the Exemplary Center for Reading Instruction (ECRI; Reid, 1997) to their RtI framework. They were interested in estimating the effects of the new intervention compared to their existing reading interventions, two of which were computer-assisted instruction (CAI) programs (i-Ready and Reading Plus) and one of which was a direct instruction (DI) group (reciprocal teaching). Students identified as at-risk readers were randomly assigned to ECRI intervention or to the control condition, which comprised the original school interventions (the two CAI and one DI intervention). The analysis of data indicated that the ECRI method resulted in a greater reading gain than the other reading interventions (Jones, Young, Gibb, & Ottehenning, 2014).

The ensuing school year, the school administration again asked for assistance in evaluating the effectiveness of their reading intervention programs, but in a way that would eliminate the need for randomization. The research team took the opportunity to incorporate a number of improvements into the study design. This dissertation documents the changes to that study. Rather than being a two-treatment model, this study included the addition of a control group since one of the main purposes for conducting the research was to compare the reading gains of students in the RtI reading interventions to those of the general school population. This was particularly important because one goal of RtI programs is to help at-risk students catch up to their peers (Martínez, Nellis, & Prendergast, 2006). In order to do this, the current study used a regression discontinuity design (RDD) to estimate a treatment effect without the need for random assignment.

Research Questions

The primary purpose of this study was to evaluate the effectiveness of the school's use of reading interventions as used in their RtI framework by comparing student outcomes of students across reading intervention assignment groups. Specifically the study explored the following questions:

1. What effect does participation in ECRI have on the reading ability of students as measured by scores on the KTEA-II BFR?
2. What effect does participation in i-Ready have on students' reading ability as measured by scores on the KTEA-II BFR for students in grades 1-3?
3. What effect does participation in Reading Plus have on students' reading ability as measured by scores on the KTEA-II BFR for students in grades 4-6?
4. What effect does participation in the three reading interventions have on English Language Learners' reading ability as measured by scores on the KTEA-II BFR in grades 1-3 and grades 4-6?

Chapter 2. Background and Literature Review

The literature review presented in this dissertation consists of a synthesis of the research on RtI programs in elementary school and juxtaposes it with a particular quasi-experimental research design (the regression discontinuity design). The researcher focused primarily on elementary reading RtI programs, and the primary focus of the literature review was at-risk students, including English Language Learners. ERIC, EBSCO, Google Scholar and EconLit were searched using terms such as *response to intervention, intervention, elementary education, reading, reading comprehension, vocabulary, reading ability, English language learner, at-risk,* and *regression discontinuity.* Because of the school's method of reading intervention delivery, terms such as *direct instruction, computer-assisted instruction, education technology, paraprofessionals,* and *paraeducators* were also used.

The Response to Intervention Model

RtI refers to a schoolwide method of decision-making in which teachers make judgments about students based on relevant data (Berkeley et al., 2009; Fuchs et al., 2003). In RtI, school personnel assess whether or not a student has responded effectively to a given intervention. Teachers then make data-based decisions about the student's further need for additional instructional interventions and systematically provide those interventions to the student (VanDerHeyden, Witt, & Gilbertson, 2007). Because of its use in responding to student academic deficiencies, RtI has become a staple of supporting academic achievement for students at differing levels of ability (Goss & Brown-Chidsey, 2011).

Typical RtI involves a determination of the level of intensity that students need in order to master academic content. Interventions are normally divided into three levels of instructional intensity (Brown-Chidsey & Steege, 2010) in order to "[implement] increasing tiers of targeted instruction… based on student progress” (Kamps, et al., 2007, p. 155). Tier 1 is instruction that is taught to all students, usually in a general education classroom context. Tier 2 instruction involves more focused instruction for students with greater academic need, often administered via regularly meeting small groups. Tier 3 instruction is reserved for individualized, intensive remediation for students who do not respond effectively to Tier 1 and Tier 2 instruction. The RtI model allows and promotes movement between tiered interventions as students progress and meet learning targets (Kamps et al., 2007).

Not all studies have shown that RtI is effective in closing the achievement gap between at-risk readers and their peers. In their meta-analysis of RtI effectiveness for students at risk for reading failure, Tran, Sanchez, Arellano, and Swanson (2012) evaluated 13 recent studies that provided pretest-posttest comparisons of RtI reading programs as well as reported effect sizes. They concluded that RtI practices in general did not reduce the achievement gap between high responders and low responders at posttest. In a formal response, Stuebing, Fletcher, and Hughes (2012) argued that the methodology used in the meta-analysis was overly complicated and did not adequately support the conclusions presented in the paper. Swanson (2012) responded in turn by defending the original methodology, but also noted that better-designed research was needed to explore the effectiveness of RtI in a defensible manner.

Direct Instruction

The RtI framework allows schools the flexibility of choosing intervention approaches that they feel best meet their academic goals. DI is a structured method of teaching that is frequently used in order to teach academic skills such as language, reading, and mathematics (Flynn, Marquis, Paquet, Peeke, & Aubry, 2012). DI has been widely implemented in RtI interventions (Daly, Martens, Barnett, Witt, & Olson, 2007; Kamps et al., 2007; Linan-

Thompson, Vaughn, Prater, & Cirino, 2006; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). Ryder, Burton, and Silberg (2006) describe three common instructional attributes of DI: (a) the teacher breaks down the skill into smaller parts that may be taught in isolation; (b) the teacher actively directs the learning activity; and (c) students have minimal input in lesson delivery. DI is characterized by clear content presentation, carefully sequenced and supported instruction, systematic feedback, and high opportunities to respond (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008).

When implementing DI, teachers focus on using teaching strategies that activate prior knowledge, explain the importance of the target skill, and facilitate student use of the skill by modeling, providing step-by-step instructions, and gradually releasing control of the skill to the students (Rupley, Blair, & Nichols, 2009). Shippen, Houchins, Steventon, and Sartor (2005) generalized the DI process into three main parts: "[Teachers] model (provide the correct response), lead (have the student say the correct answer with the teacher), and test (give immediate feedback and a delayed probe on the task initially attempted)" (p. 176).

DI has been used to offer reading instruction to students at all levels of reading ability. Coyne et al. (2009) conducted a qualitative analysis of the effect of two DI programs on the listening and reading comprehension skills of student participants. They suggest that DI principles are effective for improving students' comprehension ability across a wide range of student abilities and at different developmental situations. DI practices have also been shown to have a positive effect on the reading ability of traditionally at-risk student populations including ELL students (Ralston, Benner, Nelson, & Caniglia, 2009) and students with learning disabilities (Dağseven Emecen, 2011, Flores & Ganz, 2009; Flores et al., 2013; Wilson & Sindelar, 1991).

Computer-Assisted Instruction

CAI is another method often used to supplement regular classroom instruction (Cheung & Slavin, 2012), and has become extremely popular in educational settings over the last several decades (Liu, Moore, Graham, & Lee, 2002). CAI programs can be defined as “interactive learning method[s] in which a computer is used to present instructional material, monitor learning and help in selecting and accessing additional material in accordance with individual learner needs” (UNESCO, 2016, Computer-Assisted Instruction, para. 1). CAI programs may include features such as student assessment and performance reporting, learning activities and games, and other activities that provide targeted learning support (Cheung & Slavin, 2012).

The effectiveness of CAI programs has been a continuing matter of debate. A meta-analysis provided by Chambers (2003) suggested an overall positive effect of CAI programs on the reading ability of students who participated in their use. Conversely, Kulik (2003) reviewed 27 controlled evaluation studies and found no significant effect of CAI programs on the reading ability of elementary or secondary students. In their meta-analysis of CAI reading programs for upper-grade students, Slavin, Lake, Cheung, and Davis (2009) evaluated 31 studies that met the following criteria: (a) the study focused on upper-elementary reading, (b) the study had a control group, (c) the study lasted for at least 12 weeks, and (d) the study was published after 1970. The majority of the CAI programs in these studies were used as a supplement to classroom instruction in doses of 30 minutes, one to three times per week. Based on the results reported in these studies, the research team concluded that CAI programs had a minimal effect on upper-grade students’ reading ability. In a follow-up synthesis, they evaluated supplementary reading CAI programs in early elementary school. Overall, the findings of the second synthesis agreed with those of the first and prompted the researchers to conclude that “research on the use of technology in beginning reading instruction does not support use of the types of software that have been most commonly used" (Slavin, Lake, Chambers, Cheung & Davis, 2009, p. 19).

ELL Students and RtI

Schools are often particularly concerned with how reading interventions are meeting the needs of their ELL population. This group of students typically has reading deficits that are much greater than their English-only counterparts. An analysis of fourth-grade scores on the National Assessment of Educational Progress (NAEP) reading test indicated that the reading gap between ELL students and non-ELL students was 38 points (National Center for Education Statistics, 2015). ELL students have been shown to have significant and persistent vocabulary deficits compared with their non-ELL peers, which may limit an ELL student's ability to comprehend text at grade level and to learn early reading skills such as phonics (August, Carlo, Dressler, & Snow, 2005).

Past research has suggested that ELL students respond positively to intervention efforts (Crevecoeur, Coyne, & McCoach, 2014; Van Staden, 2011). ELL students benefit from DI interventions at similar rates when compared to English-only students (Kamps et al., 2007). The RtI model has also been used as a method to address the increasing number of ELL students in special education classrooms (Ybarra, 2012). In the interest of offering students a less restrictive learning environment, schools often opt for RtI as "an alternative to special education" (Ybarra, 2012, p. 33).

In his comparison of English-Only Learners with ELL students in a large urban school district with a significant ELL population, Ybarra (2012) found a significant relationship between reading achievement and ELL involvement in school RtI programs. In another study, Kamps et al. (2007) studied 318 elementary school students, including 170 ELL students (mostly native Spanish speakers). Overall, ELL students in the experimental (RtI) schools, and specifically those participating in DI interventions, experienced greater outcomes than students in the comparison schools.

Paraprofessionals in Interventions

Schools who implement RtI or other highly structured intervention models must address the critical question of how to administer appropriate interventions using available resources (Sansosti, Noltemeyer, & Goss, 2010). Traditionally, classroom teachers have administered interventions. However, with the expanding use of RtI, schools often rely on paraprofessional personnel, otherwise known as paraeducators, to support teachers in the administration of interventions (Hauerwas & Goessling, 2008). Until recently, studies examining the use of paraprofessionals in intervention settings were rare. Because of the trend toward using paraprofessionals more frequently in intervention settings (French, 2001; Giangreco, Edelman, Broer, & Doyle, 2001), studies evaluating their role in delivering interventions are increasingly prevalent and relevant to instructional practice.

Giangreco (2013) reviewed the current literature and practice of paraeducators in interventions and noted that their use has increased both nationally and internationally. He identified several keys to using paraeducators to successfully support instructional interventions. First, paraeducators should be used to teach supplemental instruction and not as a replacement for primary instruction. Second, paraeducators should teach from professionally prepared plans based on evidence-based practices. Third, paraeducators should be trained to implement the interventions with fidelity. Paraeducators should also be trained to support student learning by constructively managing and responding to behavior challenges. Finally, paraprofessionals should be provided ongoing support and training from professional educators.

Previous studies have shown that paraprofessional-led interventions can have significant and lasting effects on student achievement (Lane, Fletcher, Carter, Dejud, & Delorenzo, 2007; Savage & Carless, 2005). However, the effectiveness of paraprofessionals acting as the primary source of intervention delivery is debated. Webster, Blatchford, and Russell (2013) noted that students who received most of their support from teaching assistants showed less engagement in class, and made significantly less academic progress than a student who received less paraprofessional support. They stressed that the impact of paraprofessionals hinged on the effectiveness of the school in their deployment, preparedness, and employment conditions.

Causality and Educational Research

One of the chief purposes of educational research is the search to explore causal relationships between instructional treatments and student achievement. The validity of causal claims, while central to much research, relies on the ability of the research design to support such claims by eliminating bias in effect estimates (Shadish, Cook, & Campbell, 2002). The chief concern when attempting to establish causality lies in the fact that the circumstances for clear causality to be identified are impossible in the real world. Namely, researchers can never have the same students simultaneously in a treatment and a control condition. Ideally, as Murnane and Willet (2010) explain, such a comparison would allow for a comparison of the same individual with and without the treatment, thereby perfectly establishing the individual treatment effect (ITE). Let y_{1i} denote the value of the i^{th} student's outcome with treatment ($T_i = 1$), and y_{0i} denote to the value of the i^{th} student's outcome when assigned to the counterfactual condition ($T_i = 0$), and $i = 1, 2, \ldots, n$, then the individual treatment effect (ITE) for that child would be:

$$ITE_i = y_{1i} - y_{0i} \tag{1}$$

Were the individual treatment effects of all students to be identified in this manner, we could then identify the average treatment effect for all students in the target population:

$$ATE_i = E[y_{1i} - y_{0i}] \tag{2}$$

Realistically, however, such a comparison is not possible, since one individual cannot experience two different conditions (treatment and control) concurrently in the real world (West, & Thoemmes, 2010).

The potential outcomes framework (also known as Rubin's causal model) provides a basis for understanding the subject of causality and for estimating an average treatment effect (ATE) notwithstanding the lack of a concurrent counterfactual. Rubin's causal model allows researchers to estimate the counterfactual via random selection and random assignment (Murnane & Willet, 2010). Of the available research designs, researchers view the randomized controlled trial (RCT) as the gold standard in research design for estimating the counterfactual under the potential outcomes framework (Shadish, Galindo, Wong, Steiner, & Cook, 2011). In general, RCT experiments address the internal validity concerns from selection bias by randomly selecting participants from a population and then by randomly assigning them to treatment or control groups, thus theoretically equating both groups on both observed and unobserved covariates.

However, random assignment is not always feasible in education due to both ethical and practical reasons (Shadish et al., 2002). This is especially true if the treatment to be studied is intended to benefit a specific portion of a population, as in the case with a school's Tier 2 interventions, when randomizing would mean that some of the at-risk population would not receive a needed treatment (Lesik, 2006). In such cases, it is impossible to accurately fulfill the requirements of Rubin's causal model through the assignment mechanism of randomization.

The Regression Discontinuity Design

Another research design that has been presented as an alternative to the RCT when randomization is not feasible is the regression discontinuity design (RDD), otherwise known as a cutoff-based experiment (Cook, 2008). First introduced by Thistlethwaite and Campbell (1960), the RDD uses an alternate assignment mechanism rather than relying on random assignment to a treatment condition. Shadish et al. (2002) explain that the assignment mechanism is based on a participant's score on an appropriate cutoff variable. When designing a regression-discontinuity (RD) analysis, researchers select a cutoff score from a chosen assignment variable, which can be any non-dichotomous measure that occurs before the start of treatment. The cutoff score is used to divide the study units into different treatment conditions (Shadish et al., 2002). Participants who score below the cutoff, also known as the threshold, are assigned to one treatment condition, while participants who score at or above the threshold are assigned to the other treatment condition (Lee & Munk, 2008). Following treatment, participants are assessed via an outcome variable and a regression is computed. Any discontinuity in the regression line at the cutoff point may indicate a treatment effect (Lesik, 2006).

In terms of RDD, the regression equation is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon \tag{3}$$

where y_i is the outcome score for the i^{th} unit, β_0 is the coefficient for the intercept, x_i is the centered assignment score, β_1 is the linear assignment variable coefficient, and z_i is the dummy variable for treatment such that $z_i \in (0,1)$ and $z_i = 1(x_i \leq x_0)$. The local average treatment effect (LATE), which is the ATE at the cutoff point, is expressed by β_2, which is the mean difference for treatment at point x_0. The discontinuous point x_0, in the case of RDD, is assumed to be a known factor and is the cutoff point at which assignment to treatment condition is made

(Hahn, Todd, & Van Der Klaauw, 2001). The LATE equation for an RDD can be modified from Equation 2 to reflect the influence of the cutoff point:

$$LATE_i = \lim_{x_i \downarrow x_o} E\,[y_{1i}|x_i = x] - \lim_{x_i \uparrow x_o} E\,[y_{oi}|x_i = x] \tag{4}$$

RDD assumptions. A critical assumption of the RDD is that the probability of receiving treatment changes perfectly from 0 to 1 at the cutoff (Hahn, Todd, & Van Der Klaauw, 2001). In order for this assumption to hold, participants must adhere to their predetermined assignment group for the duration of the treatment (Lee & Munk, 2008). *Crossovers* are defined as participants who are assigned to one treatment but subsequently receive another treatment. Participants who remain with the initial treatment assignment are known as *compliers*. When all participants are compliers, the probability of assignment to treatment changes from 0 to 1 at the cutoff. In this case, the RDD is classified as a *sharp* design. In other words, the probability of receiving treatment is perfectly discontinuous at x_0. Otherwise, the presence of significant numbers of crossovers may add bias to the effect estimate, and necessitates the use of a *fuzzy* regression-discontinuity analysis (RD analysis), which means treating the forcing variable as an instrumental variable (Imbens & Lemieux, 2008). While an RD analysis can still be done using a fuzzy design, because the probability does not experience a perfect shift at the cutoff point the researcher must take into account the effect of the crossovers on the estimated treatment effect (Bloom, 2012).

In addition to the question of crossovers, the RDD functions under a number of other assumptions. First, a single regression model can be found for the pre-treatment relationship between the outcome variable y_i and the score variable x_i, with a correctly specified functional form. Second, no other factor exists that might cause the discontinuity at the cutoff. Lee and Lemieux (2010) warn that the integrity of the assignment variable may be at risk if individuals can precisely manipulate the assignment variable. For example, if students knew that a test score *x* would ensure (or prohibit) their participation in a certain treatment, and were able to choose their score through more or less effort (fueled by the knowledge of the importance of the score), then students just above and below the cutoff point could vary in unobservable ways.

However, if students do not have the ability to manipulate the assignment variable, the variation just around the cutoff point would approximate random assignment without endogenous interference. In other words, the RDD can be characterized as a case of local randomization (Bloom, 2012; Lee & Lemieux, 2010). The validity of this randomization can be tested as one might test a randomized experiment. This can be accomplished, according to Lee and Lemieux (2010), by testing the baseline covariates of the students just above and below the cutoff. If the baseline covariates themselves show discontinuities, then the local randomization of the RDD may be rendered invalid.

Another assumption of the RDD is that treatment of one subject must not interfere with the outcome of another. This assumption, which is shared with the RCT, is known as the stable unit treatment value assumption (SUTVA; Rubin, 1990). According to SUTVA, the outcome of a student in the treatment group cannot depend on the assignment and treatment of another student. The RDD also assumes that the conditional average treatment effect is the same for all values of the score variable. This signifies that the continuation of the regression line beyond the cutoff for treatment and control groups is parallel (Bloom, 2012). These assumptions are difficult to meet in reality, meaning that the internal validity of a study could be compromised if not followed with precision (Lee & Munk, 2008).

A final drawback of the RDD is that it requires a much larger sample size than a traditional randomized control design to achieve equivalent statistical power (Schochet, 2008).

Notwithstanding the drawbacks, studies that employ RDD have become more common over the past several decades in the social sciences, especially since it is able to estimate the ATE of a treatment without the need of random assignment (Shadish et al., 2011).

Lee and Munk (2008) suggest methods with which to respond to violations of assumptions in RDD. First, overfitting the model by including higher-order polynomials can reduce the bias due to model misspecification. They indicate that this bias can also be reduced by using a nonparametric RDD method (Hahn et al., 2001; Imbens & Lemieux, 2001), although this method requires a larger sample size than the parametric method. Second, the addition of covariates that are strongly correlated to the outcome variable can increase the efficiency of the RDD model. This can be extremely valuable since many questions on how to handle bias in an RDD boil down to a variance-bias trade-off.

Another method of addressing the limitations of RDD is to use a local linear regression approach (Lee & Lemieux, 2010) by using a window of data around the cutoff. This method functions under the assumption that the RDD is essentially an RCT in the immediate vicinity of the cutoff (Bloom, 2012). Using this method, however, means that any treatment effects may only be described sufficiently within the narrow range specified by the window. The treatment effect may or may not hold outside of that window. Using a window, therefore, limits the researcher to obtaining a local average treatment effect (LATE) instead of the ATE.

RDD evidence standards. Schochet et al. (2010) classify the assumptions underlying RD analysis into four criteria that must be met in order for a RDD to meet high evidence standards. First, there must be no systematic manipulation of the forcing variable. This can be ensured through a thoroughly endogenous assignment process, and can be validated graphically (via a histogram) or statistically by evaluating the density of the forcing variable around the cutoff (McCrary, 2007). Second, both overall and differential attrition rates must meet RCT standards. Third, as previously stated, the forcing variable must be continuous at the cutoff point. While not strictly testable, this can be checked by evaluating the means of available covariates directly above and below the cutoff point. Fourth, the functional form must be properly identified.

RD analysis and discrete data. One particular concern with the RDD arises when the assignment variable is discrete (e.g., has a finite number of possible values). In practice, test scores are rarely truly continuous. Rather, students have a finite number of values that their score may take. As an example, in the current study within a window of 21-40 on the NCE scale only a limited number of discrete values are possible above and below the cutoff. A student could not possibly score a 31.27, for example. Thus, one could defensibly classify this data as discrete rather than continuous owing to the relatively small number of discrete possible values.

Lee and Card (2008) suggest that standard errors may be underestimated when the assignment variable is discrete. This is due to the assumption that the RDD is equivalent to an RCT just above and just below the cutoff point. The authors note that "[in the case of discrete data] it is no longer possible to compute averages within arbitrarily small neighborhoods of the cutoff point, even with an infinite amount of data. . ." (Lee & Card, 2008, p. 656). To do so requires assuming that the chosen parametric functional form correctly models the underlying function. To account for the necessary assumption of the functional form, they suggest regressing the outcome variable (y_i) on the chosen polynomial and dummy assignment variable (x_i) and by using the assignment variable (x_1) as a clustering variable. The clustered standard errors from this procedure can be then compared with the conventional ones to verify the credibility of the inferences (Imbens & Lemieux, 2007).

Use of the RDD in educational contexts. After Thistlethwaite and Campbell (1960) introduced the RDD to research, it was most often used in economics. However, the unique ability of a RDD to estimate treatment effect without withholding treatment made it particularly interesting to the social sciences, since it is particularly useful when treatment is necessary for the well-being of those who participate. These types of treatment occur frequently in education and other social sciences (Shadish et al., 2011). Because of this, the RDD approach has been adopted by many fields including education to evaluate many academic programs and treatments. For example, Matthews, Peters, and Housand (2012) demonstrated the use of the RDD approach in the context of gifted education. Their study highlighted the approach when the cutoff threshold and treatment is on the upper end of the assignment variable's continuum. Tuckwiller, Pullen, and Coyne (2010) took the reverse approach, and used the RDD to assign students below the cutoff point to the treatment group. Their study involved 92 kindergarten students across six classrooms, and used the Peabody Picture Vocabulary Test, Fourth Edition as the selection measure. Students below the cutoff received Tier 2 interventions, while students above the cutoff only received Tier 1 instruction. They found that the interventions were effective in improving student reading levels on the outcome variable.

With the overall scarcity of studies evaluating the effects of school-wide RtI programs, the RDD has not been frequently used in evaluating RtI interventions on a school-wide basis (Hughes & Dexter, 2011). Only three RDD studies evaluating Tier 2 interventions had been published in academic journals during the past decade. Of these studies, all were limited to a single grade. One used an archival data set that was produced from a previous RCT study in order to evaluate the effects of a vocabulary intervention on at-risk students in first grade at three elementary schools (Ashworth & Pullen, 2015). The other examined the effect of a Tier 2 reading intervention on the oral reading fluency of at-risk readers in first grade (Baker, Smolkowski, Chaparro, Smith, & Fien, 2015). The third was a pilot study evaluating the effects of a vocabulary intervention prototype on the vocabulary skill of at-risk kindergarten students (Tuckwiller, Pullen, & Coyne 2010).

Recently, Balu et al. (2015) conducted a large-scale RD analysis of Tier 2 and Tier 3 reading interventions in RtI programs at 146 elementary schools in multiple states. Only students who maintained the treatment status for the duration of the study were included in the RD analysis. They found that assignment to Tier 2 reading intervention had a negative impact on the reading achievement of first grade students in 81 schools, with 15 of those schools showing a significant negative difference at the cutoff point. However, they also found that achievement in RtI programs varied significantly across schools, and cautioned that the findings could only be generalized to students directly above and below the cutoff.

The Current Study

The study described in this dissertation is similar to the study by Tuckwiller et al. (2010) in that it applies the RDD as a tool in evaluating interventions at a single elementary school's RtI program. The treatments, in this case reading interventions, were administered to students who fell below two predetermined cutoff points on the assignment variable. This study included the entire school population (excluding kindergarten), and was designed to evaluate the multiple interventions that were used in the target school. The primary emphasis, therefore, of this study was to analyze the effects of the school's reading interventions on the reading achievement of at-risk students, using the achievement of the general school population as a control.

The current study adds significantly to the research on RtI programs and on reading interventions by applying a robust quasi-experimental design to evaluate the effectiveness of such interventions in an applied setting. Specifically, the use of the RDD to evaluate the reading intervention program in the school's RtI framework is important because it helps to fill a void of experimental and quasi-experimental studies exploring the causal relationship between student achievement and assignment to tiered levels of reading interventions.

Chapter 3. Method

Design

This study applied a sharp RDD to explore the existence of a causal relationship between treatment and student reading achievement in order to estimate a LATE for the reading interventions in the school's RtI program. The RDD is a quasi-experimental design that does not rely on random assignment, although it may incorporate such elements into its design (Shadish et al., 2002). In place of random assignment, the design used an assignment variable to identify a cutoff point, which was then used to determine how subjects were placed into control or treatment conditions (Murnane & Willet, 2010).

Although often used to select a single cutoff point, the assignment variable can also be used to select more than one cutoff point in the case of multiple treatments (Shadish et al., 2002). Because the RtI program at the elementary school included more than two levels of treatment, two separate cutoff points were used to assign students to the various levels of reading instruction. The following diagram represents the study design:

O_A	C_3	X	O_2	X	O_3
O_A	C_2	X	O_2	X	O_3
O_A	C_1		O_2		O_3

where O_A is the assignment variable, C_1 indicates assignment to the control group, C_2 and C_3 indicate assignment to respective treatment levels, X indicates treatment, and O_2 and O_3 indicate posttest iterations.

Assignment variable and cutoff scores. Normal Curve Equivalent (NCE) scores on the KTEA-II BFR were used to select cutoff points for treatment assignment. Following the assessment, two cutoff points were selected based on frequency distributions of the NCE reading scores. Because of the benefit of having a large cluster of students around each cutoff, the cutoff points should ideally be placed close to the mean, while still allowing for a significant amount of students both above and below the cutoff (Shadish et al., 2002; Trochim, 1984). Additionally, a larger sample below the cutoff improves the number of subjects in each treatment condition, a factor that can have an effect on the statistical power of the data (Schochet et al., 2010).

Assignment process. The research team's testers completed administration of the forcing variable one week prior to the selection of cutoff points. During this time, school personnel in charge of assigning students to treatment were not allowed access to the raw test data to maintain the integrity of the cutoff selection process. To determine cutoff points, researchers met with the school administration to assess the number of students that the intervention team could feasibly service given the resources available to the school. Data were de-identified with any possible identifying information and demographics removed; only NCE scores were available during the cutoff selection meeting.

Figure 1 indicates the two cutoff points selected by the research team. During the assignment meeting, the cutoff assignment team (comprised of the school administrator and research personnel) discussed several practical factors including the availability and number of paraeducators and intervention space. The team tested several different cutoff score positions, noting the size of each treatment group at various cutoff combinations.

Final cutoffs were selected by the team based on the school's capacity to offer intervention services to the various numbers of students in each treatment. After evaluating several different scenarios, the upper cutoff point was set at the NCE score of 40 and the secondary cutoff point was set at the NCE score of 30. These cutoff scores resulted in treatment group sizes that were feasible for the school while also taking into account the sample size requirements for the design as much as possible.

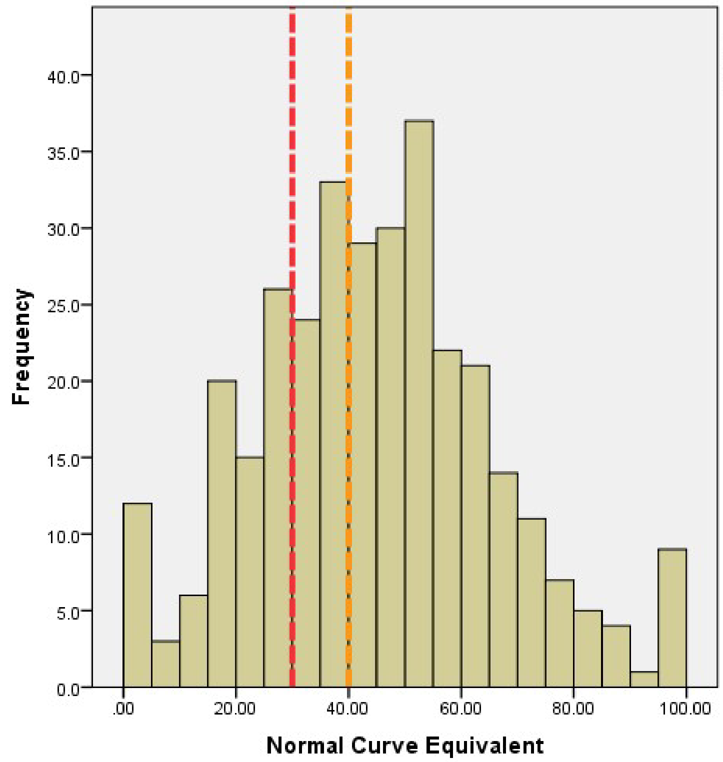


Figure 1. Cutoff points relative to assignment variable distribution.

Sample

Participants. The subjects for this study were 321 elementary school students in grades 1-6 at an urban Title I school. One first grade classroom that was part of a separate research study was excluded from the study. Additionally, two groups of students not involved in school-wide RtI instruction were also excluded. These two groups were the district-wide special education unit for students with severe disabilities and the district-wide program for students identified as being gifted. Due to the unique nature of the programs, these two groups of students were not eligible for inclusion in the school-wide RtI framework of reading interventions. The sample, therefore, included all of the general education population of the school who were participating in the three tiers of the school's RtI reading program.

Attrition and differential attrition. The overall attrition rate for the study was 11.6%, which was less than the historic mobility rates reported at the school (17% for the previous school year). The control group's attrition rate was 10.61%, with the CAI group attrition rate (13.79%) and the DI group attrition rate (12.34%) being slightly higher (Table 1). The difference in group rates was low across all groups, with the greatest difference being between the CAI and control groups (d = 2.85%). However, this level of differential attrition is still considered low, which indicates a minimal amount of expected bias (What Works Clearinghouse, 2013).

Table 1

Attrition and Differential Attrition Statistics

Treatment Condition	Initial N	Final N	Attrition	% Attrition
Control *(NCE > 40)*	179	160	19	10.61%
CAI *(30 < NCE ≤ 40)*	58	50	8	13.79%
DI *(NCE ≤ 30)*	81	71	10	12.34%
Total Sample	318	281	37	11.63%

Note. CAI = Computer-assisted instruction; DI = direct instruction.

Crossovers. Three of the 321 participants were classified as crossovers, or participants who were assigned to one treatment but received another treatment. Of these crossovers, one was initially assigned to the DI treatment (NCE of 21) and two were assigned to the control group (NCE of 43 and 44). These three crossovers account for less than one percent of the total number of the sample. Trochim (1984) noted that if the number of misaligned cases is less than 5%, then deletion of these cases would not significantly affect the probability of obtaining accurate treatment effects. Since the percentage of misaligned cases in the current study was less than .01%., these cases were removed from the study to maintain the integrity of a sharp RDD.

Summary statistics. After attrition and the removal of the three crossovers, the final sample included 281 students, with 160 students in the control group, 50 students in the CAI treatment group and 71 students in the DI treatment group (Table 2). Fifty-six percent of the students were male and 44% were female. The study included students in first through sixth grades: 39 students were in first grade, with 67 in second grade, 48 in third grade, 51 in fourth grade, 35 in fifth grade, and 41 in sixth grade. Sixty-four students were currently classified as ELL, and 26 students had Individual Education Programs (IEPs). The ELL students were predominantly native Spanish-speakers.

Table 2

Summary Statistics for Full Sample

	Before Attrition				After Attrition			
Covariates	DI n = 81	CAI n = 58	Control n =179	Total n =318	DI n = 71	CAI n = 50	Control n = 160	Total n = 281
Age (Avg.)	8.94	9.28	9.08	9.08	8.94	9.28	9.08	9.08
Male	60%	62%	50%	55%	62%	58%	52%	56%
ELL	48%	18%	12%	22%	49%	18%	13%	23%
SPED	23%	19%	2%	10%	20%	18%	2%	9%

Note. DI= Direct Instruction ($x \leq 30$); CAI = Computer-assisted instruction ($30 < x \leq 40$); x = NCE Score on August KTEA-II BFR.

Fifty-eight students scored at or below 40 and at or above 31 on the NCE and were assigned to the CAI reading intervention group. Eighty-one students scored at or below 30 and were assigned to the DI intervention group. One-hundred and seventy-nine students scored above 40; these students were assigned to the control group.

The mean NCE score on the KTEA-II BFR was 44.39. In all, 139 students were assigned to a more focused reading intervention. The number of students in interventions during this study was approximately 35% larger than the number of students who received reading interventions the previous year. This increase was partially due to the research team's intentional overestimation of student need so as to service all students at-risk for reading failure.

Grade-level sample sizes ranged from a minimum of 35 students to a maximum of 67 students, with the average grade size being 46.8 students. The proportion of students in reading interventions varied by grade, with third grade having the smallest proportion of students in treatment (25%) and first grade having the largest (64%).

In the lower grades, 29.4% of participants were assigned to the DI intervention, 18.5% were assigned to the CAI intervention, and 52.1% were assigned to the control. For the upper-grade stratum, 39.5% of participants were assigned to the DI intervention, 25.7% were assigned to the CAI intervention, and 34.7% were assigned to the control. After attrition, 45.8% of students received Tier 2 reading interventions; 58% were assigned to DI intervention and 42% were assigned to CAI intervention (Table 3).

Independent Variable

The independent variable in this study was the type of reading instruction in which students participated. This study included three types of reading intervention instruction: (a) direct-instruction, (b) computer-assisted instruction, and (c) business-as-usual school reading instruction. As part of their RtI model, the school had selected three reading interventions for use in their Tier 2 instruction: ECRI (DI), Reading Plus (CAI), and i-Ready (CAI).

Direct instruction (ECRI). ECRI uses teacher-directed instruction to help students achieve high mastery criteria for reading fluency, reading comprehension, and spelling (Reid, 1996). ECRI teaches new vocabulary according to the nature of the word (e.g., phonics, word structure, sight word). ECRI also employs word spelling practice with the list of vocabulary words from the unit acting as a mastery test.

Table 3

Summary Statistics for Full Sample by Grade Level

	First Grade				Second Grade			
Covariates	*DI* *n = 20*	*CAI* *n = 5*	*Control* *n =14*	Total *n = 39*	*DI* *n = 12*	*CAI* *n = 10*	*Control* *n = 45*	Total *n = 67*
Age (Avg.)	6.80	6.80	6.64	6.74	7.92	7.50	7.67	7.69
Male	70%	80%	64%	69%	38%	80%	49%	55%
ELL	45%	0%	7%	26%	46%	30%	17%	24%
SPED	5%	0%	0%	3%	8%	20%	2%	6%

	Third Grade				Fourth Grade			
	DI *n = 3*	*CAI* *n = 9*	*Control* *n = 36*	Total *n = 48*	*DI* *n = 19*	*CAI* *n = 11*	*Control* *n = 21*	Total *n = 51*
Age (Avg.)	9.00	8.67	8.61	8.64	9.74	9.82	9.71	9.75
Male	100%	42%	55%	54%	57%	79%	45%	57%
ELL	67%	42%	28%	35%	37%	9%	0%	16%
SPED	67%	33%	0%	13%	21%	0%	0%	8%

	Fifth Grade				Sixth Grade			
	DI *n = 9*	*CAI* *n = 9*	*Control* *n = 17*	Total *n = 35*	*DI* *n = 8*	*CAI* *n = 6*	*Control* *n = 27*	Total *n = 41*
Age (M)	10.67	10.80	10.65	10.69	12.11	12.00	11.86	11.62
Male	67%	33%	53%	51%	64%	63%	42%	50%
ELL	67%	0%	0%	17%	77%	29%	0%	17%
SPED	33%	22%	0%	14%	45%	14%	9%	15%

Note: Grade-level summary statistics are for sample after attrition. DI = Direct Instruction ($x_i < 40$); CAI = Computer-assisted instruction ($30 < x_i \leq 40$); x_i = NCE Score on August KTEA-II BFR for the i^{th} individual.

Instruction is teacher-directed, with lessons following a rigid scripted sequence based on the structure of the word being taught. To pass a mastery test, students must read the words with 100% accuracy, one word per second, and spell the words with 100% accuracy. In ECRI, new vocabulary words are pre-taught to students and then read in the context of an expository or informational text. During preteaching lessons, students must respond both orally and in writing to teacher prompts.

Computer-assisted instruction. The school used two CAI reading programs as part of its RtI framework: Reading Plus (Taylor Associates/Communications, 2014) and i-Ready (Curriculum Associates, 2014). Both programs were online reading programs accessible via school laptop and desktop computers.

I-Ready. I-Ready is an adaptive computer program with reading instruction, practice, and assessment components. The program content is correlated with the Common Core State Standards, and offers reading instruction across a wide level of grades. An independent study by the Educational Research Institute of America indicated strong correlations between i-Ready and scores on the 2013 New York State Assessment (Curriculum Associates, 2013).

I-Ready classifies lessons into five categories: (a) phonology lessons, (b) phonics lessons, (c) high-frequency word lessons, (d) vocabulary lessons, and (e) comprehension lessons. Lessons consist of a teaching section, which introduced students to new material, followed by a practice section that is often in the form of a game. Each lesson included an assessment that measured the students' accuracy and retention of the target skill.

Reading Plus. Reading Plus is a web-based reading intervention that provides scaffolded silent reading practice (What Works Clearinghouse, 2010). The program includes reading fluency, comprehension, vocabulary, and reading assessment components and is adaptive to the students' reading skill level. In a study conducted in Miami-Dade County in Florida, Reading Plus was found to have "potentially positive effects on comprehension for adolescent learners." (What Works Clearinghouse, 2010. p.1).

Reading Plus classifies lessons as one of three types: (a) See Reader, (b) I-Balance, and (c) Read Around. When completing lessons, data were collected and divided into the following categories: (a) close reading, (b) main idea identification, (c) language, (d) structure, (e) point of view, (f) imaging, (g) reasoning, (h) comparative reading, and (i) mastered words. See Reader lessons were the initial activity and all students were required to participate in them. These 15-minute lessons offer scaffolded reading support by way of a "guided window" that assists in text tracking. At the completion of the See Reader lessons, students answer assessment questions based on the text read during the lesson.

The I-balance lessons are 10 minutes in length and were assigned to students with a silent reading rate of 140 words per minute or less. These lessons were divided into activities that offered students support in scanning text as well as activities intended to strengthen visual discrimination and visual memory. Finally, Read Around lessons introduce students to new words and include lessons intended to pre-teach vocabulary words.

Comparison of learning targets. Table 4 provides a comparison of the literacy skill targets and the focus areas of each of the reading interventions. While both CAI programs focus on foundational reading skills, they differ in their approach and skill focus. I-Ready covers a much broader scope of skills, including practice in phonemic awareness, phonics, vocabulary, fluency, and text comprehension. Reading Plus provides instruction and practice in fluency, comprehension and vocabulary skills, but does not provide specific practice of phonics or phonemic awareness. ECRI provided instruction targeting phonemic awareness, phonics, vocabulary, fluency, and text comprehension.

Dependent Variable

Normal curve equivalent scores of the Reading Composite (which comprised Reading Part 1 and Reading Part 2) from the KTEA-II BFR were used as the dependent variable. Two iterations of the dependent variable were used: a posttest midway through the school year (the January KTEA-II BFR) and at the end of the school year (the May KTEA-II BFR).

Table 4

Reading Skill Components/Focus by Reading Program

Learning Target	ECRI (Grades 1-6)	Reading Plus (Grades 4-6)	i-Ready (Grades 1-3)
Phonemic Awareness	Yes	No	Yes
Phonics	Yes	No	Yes
Fluency	Yes	Yes	Yes
Vocabulary	Yes	Yes	Yes
Comprehension	Yes	Yes	Yes

Note. ECRI = Exemplary Center for Reading Instruction.

Instruments

The primary instrument used in the study was the KTEA-II BFR. This instrument was divided into four subtests: (a) Reading Part 1 (Recognition), (b) Reading Part 2 (Comprehension), (c) Math, and (d) Writing. The reading subtests were combined in the scoring to form a reading composite score. Because the scope of the study was limited to reading achievement, only Reading Parts 1 and 2 (comprising the reading composite score) were used. The reading subtest included 67 items, split into 46 word recognition items and 27 comprehension items.

Composite scores on the KTEA-II BFR were reported to have an internal-consistency reliability coefficient of .94. This coefficient was obtained using the split-half method.

Composite scores on the KTEA-II BFR were also reported to have an estimated test-retest reliability of .95 for first and second grades and .91 for fourth through seventh grades. This coefficient represents the reliability of the reading composite scores. Scores on the KTEA-II BFR were also reported to be highly correlated with other well-known and widely-used instruments. The KTEA-II BFR was reported to be highly correlated (.78) with the Woodcock-Johnson III (WJ III) Broad Reading cluster, as well as with the WJ III Basic Reading Skills Cluster (.89) (Kaufman & Kaufman, 2005).

Procedures

Paraeducators. Ten paraeducators were employed by the school over the course of the schoolyear to assist in the RtI reading interventions. Of these, three paraeducators quit during the schoolyear and were replaced by newly-hired paraeducators. One was male and the remainder were female. Three paraeducators had current teaching licenses, and four had worked at the school for over five years. Of the paraeducators who had not had prior experience, two paraeducators were in teacher preparation programs at local universities. Paraeducators were assigned by the school administration to administer either DI interventions, CAI interventions, or both. The school team made these assignments arbitrarily and not based on the paraeducators' previous experience.

The paraeducators assigned to the DI intervention were trained on ECRI implementation at the beginning of the school year. As part of this training, the paraeducator instructors received four hours of preparation and practice provided by an ECRI instructor. The initial session included information on teaching procedures such as DI, small-group management, record keeping, and student assessment. After one week of practice time teaching students, the ECRI trainer observed each of the paraeducators and provided formative feedback and coaching as needed over three teaching sessions for each. The ECRI trainer monitored treatment fidelity and provided booster sessions to the paraeducators throughout the year as needed.

Paraeducators assigned to CAI interventions also received an initial training session on the use of i-Ready and Reading Plus. A paraeducator who was an experienced user of the two computer programs provided training and support to those assigned to the CAI programs as needed. Since these programs were self-directing, the paraeducators' role in their use was supportive rather than directive. Therefore, paraeducator training focused on redirecting student misbehavior, assisting with logon and technical issues, evaluating usage reports, and maximizing time on task. Paraeducators were also trained to monitor student usage data and to make necessary changes to lesson sequences or assignments as necessary. A member of the school staff acted as a facilitator and offered feedback and instruction throughout the year as needed.

Throughout the year, paraeducator performance was monitored by trained treatment integrity observers. Observers met with the researcher biweekly to discuss concerns about treatment fidelity and to discuss specific paraeducator concerns (e.g., punctuality, pacing, and behavior management). These discussions were used to inform subsequent ongoing paraeducator training activities conducted by the research team and the school. While the school and research team made every effort to train paraeducators at the beginning of the school year, as previously noted several of the paraeducators left during the study. The school and research personnel trained replacement paraeducators within two weeks of their hire date.

Intervention materials. Reading materials for the ECRI reading intervention group were adapted to intervention use by the ECRI trainer. This adaptation included focusing the ECRI lessons on specific reading selections taken from a basal already in use by the school. Basal reading selections and mastery test vocabulary (Appendix A) for ECRI were taken from the Reading Treasures series published by Macmillan/McGraw-Hill (2007), which was used by the school as a reading basal. The specific word lists and stories were based on the Reading Triumphs section of the basal, which was intended for use in remediation settings. The ECRI trainer provided the paraeducators with the necessary ECRI materials for teaching each lesson (Appendix B).

The CAI programs were both web-based and were accessed in one of two ways: via in-classroom computers or via mobile laptops. Based on the size of the intervention group, CAI sessions were held either in the elementary school's computer lab or in the general education classrooms using the school's mobile computer lab. The classroom teacher provided CAI students a username and password at the beginning of the year (or when they moved in).

Intervention administration. Reading interventions at the elementary school were scheduled in one-hour increments for each grade level and occurred daily Monday through Thursday. During this time, students in the control group participated solely in teacher-directed and independent business-as-usual activities. Center activities varied by class, but examples included guided reading practice, vocabulary games, writing labs, completing past reading work, and other reading-themed activities. Reading block schedules varied depending on grade level, but students at all grade levels participated in the one-hour reading block four times per week.

Students in the ECRI intervention group participated in approximately 30 minutes of ECRI instruction, with the remainder of the literacy block spent in guided reading and independent reading centers. Students in the CAI groups at participated in approximately 30 minutes of CAI, and spent the remaining time in guided reading and independent reading centers. Tier 2 students in the upper grades averaged slightly more time in interventions (35-40 minutes) and slightly less time in business-as-usual activities. All students regularly met in a 15-minute teacher-led guided reading groups as part of the one-hour reading block, although the number of times each week varied from classroom to classroom.

Treatment integrity. Observers used several treatment fidelity checklists to ensure that the treatment described in the study was implemented with fidelity (Appendix C). These forms were used as part of treatment integrity observations that occurred during the regular school day and intervention time. Graduate and undergraduate research assistants were trained to conduct weekly written treatment fidelity observations of each DI and CAI session during the first week of the school year. Observations took place four to eight times per month per intervention session for a total of 10-15 minutes each observation. Observers ensured that at least 25% of all intervention sessions were checked for fidelity. On average, paraeducators maintained treatment fidelity levels of 87% for the DI intervention and 86% for CAI interventions.

Observers also participated in regular inter-rater reliability checks. Ten percent of all observations were conducted with a second observer. Inter-rater agreement was initially established at 94% in the initial month of the study, and was maintained at 96% for the remainder of the study. Throughout the study, observers met bi-weekly to evaluate observations, address concerns, and to reestablish shared understandings of treatment integrity.

Test administration. The KTEA-II BFR was administered at the beginning of the 2014-2015 schoolyear to all students in the study. The test administrators were selected from university staff and undergraduate and graduate students from within the university's School of Education. Each test administrator received two hours of initial training and was required to administer the assessment to the testing supervisor prior to testing any student. Trained test administrators then administered the test to students in a one-on-one setting. Testers were trained to follow the scripted test administration booklet without deviation. Participants who were absent for the initial testing were tested the day they returned to school.

Students retook the KTEA-II BFR following the completion of the second school term. This mid-year measure served as both dependent variable and as assignment variable to enable the research team to be able to track students' progress at the mid-year point and to compare their progress on the KTEA-II BFR with progress monitoring instruments. Additionally, the school used a number of progress monitoring tools to track the academic reading growth of students. These included DIBELS Next (Good & Kaminski, 2002) for lower grades and Student Tutoring Achievement for Reading (STAR) for upper grades (Renaissance Learning, 2014). The second administration of the KTEA-II BFR occurred in January. It was used to evaluate the effect of the reading interventions during the first two terms. It was also intended to be used to assign students who had moved into the school boundaries during mid-term to the various reading interventions according to the pre-decided cutoffs, although these students were not used as part of the full-year study.

This initial posttest test administration was requested by school personnel to assist in tracking student achievement as part of the RtI framework as a comparison for the school's progress monitoring efforts (STAR reading and DIBELS). Because this study attempted to balance both validity and practicality, the school felt it was important to reevaluate all students in order to ensure that no students were being deprived of needed intervention. In addition to move-ins, the school intervention team discussed any students whose performance was drastically lower in January than in August as candidates for reassignment.

Teachers consulted with the team individually to discuss student needs on a case-by-case basis. While some participants did score lower than the initial cutoff on the second administration of the KTEA-II BFR, most teachers felt that their reading performance was not severe enough to merit a shift in intervention, except for one of the crossovers (as discussed in the attrition section of this dissertation). The second posttest administration of the KTEA-II BFR was conducted at the end of the fourth term. Because of state-mandated testing, the May posttest was delayed until the second-to-last week of the school year. Trained testers administered the posttest measure of the KTEA-II BFR in a similar fashion to the previous two measures, in a one-on-one setting with a trained test administrator.

Analysis

The data were analyzed in Mplus version 7.4 (Muthén & Muthén, 2016), SPSS (IBM Corp., 2015) and R (R Core Team, 2016) using a local linear regression approach (Lee & Lemieux, 2010). Trochim (2006) described the analytic model for the RDD in its expanded polynomial form as being modeled by the following equation:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i^2 + \cdots + \beta_p x_i^p \ \varepsilon_i \qquad (5)$$

where $\beta_p x_i^p$ is the specified level of polynomial. In evaluating the effect of treatment, the null hypothesis can be represented as $H_0\colon \beta_2 = 0$.

One of the keys to correctly analyzing a RDD is the accurate identification of the functional form of the data. Misspecification of the functional form can lead the researcher to interpret a flexion point in the data as a discontinuity when in fact there is none. To avoid this mistake, Shadish et al. (2002) recommend overstating the functional form by two degrees more than what is expected. Therefore, if the functional form were expected to be linear, one would first fit a cubic function. The polynomial functions that were not significant would then be removed sequentially until a significant model was identified.

In the current study, the functional form of the relationship between the assignment variable and the outcome measure was expected to be linear. Therefore, the analysis included square and cubic functions in the equation, so that:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x z_i + \beta_4 z_i^2 + \beta_5 x z_i^2 + \beta_6 x_i^3 + \beta_7 x z_i^3 + \varepsilon_i \quad (6)$$

where $\beta_4 z_i^2$ and $\beta_6 z_i^3$ represent the added polynomials for main effect, $\beta_5 x z_i^2$ and $\beta_7 x z_i^3$ represent the polynomials for interaction effect, and ε_i represents the error term. The polynomial terms were then removed sequentially (cubic interaction term first, then the cubic main effect, etc.) until only those terms that were statistically significant were left in the model equation. The sequential removal of polynomials represented an effort to obtain an unbiased and efficient treatment estimate (Trochim, 2006).

Because of the interest in treatment effect on certain subpopulations, the RD analysis included covariates that were hypothesized to predict achievement on the outcome variable. Two covariates were included in the analysis: special education status and ELL status. An interaction term for ELL with treatment status was also included in the final analysis in order to respond to Research Question 4.

In a RDD, the information distribution directly above and below the cutoff is of much more immediate importance to the analysis than information that is located at the tails of the distribution. As such, restricting the parametric analysis to a smaller window around the cutoff (by means of local linear regression) may remove any noise that may be generated by the tails of the distribution (Lee & Munk, 2008) with a potential tradeoff in efficiency.

Following Murnane and Willet's (2011) example, the principal researcher tested various sizes of window widths above and below the cutoff. Initially, a window of 20 points was used. The window was then reduced in size at each iteration until a minimum spread of six points was reached in order to test the robustness of the RD estimates (Lee & Lemieux, 2010). Table 5 indicates the various bandwidths at the lower and upper cutoffs, the *N* size for each bandwidth and the proportion of participants assigned to each treatment condition relative to the cutoff point.

Data preparation. The preprogram measure (cutoff score) was centered to facilitate interpretability. Squared and cubic main effect and interaction effect terms were also calculated to test the functional form of the data. To account for missing outcome variables and covariates, a series of 100 imputations was run using the Mplus software. The clustered nature of the model was taken into account during the imputation process. The three crossovers were removed from the data set prior to running the RD analyses. Outliers were analyzed by calculating standardized residuals, leverage values, and Cook's distance.

Table 5

Window Widths with Corresponding Sample Sizes and Treatment Proportions

Lower Cutoff (NCE = 30)			*Upper Cutoff (NCE = 40)*		
Window Width	*N*	Proportion of *N* below cutoff	Window Width	*N*	Proportion of *N* below cutoff
16-40	116	.500	31-52	121	.479
19-40	108	.463	31-51	120	.483
20-40	103	.437	31-50	113	.513
21-39	92	.489	32-49	106	.547
22-38	81	.543	33-48	82	.598
23-37	76	.513	34-47	77	.571
24-36	61	.525	35-46	69	.638
25-35	52	.615	36-46	63	.603
26-36	56	.482	37-46	54	.537
26-35	47	.574	38-45	36	.583
27-34	36	.611	39-43	35	.600

Note. NCE = Normal curve equivalent.

RD analyses. A series of separate analyses were run to account for the multiple cutoffs and treatment groups. Lee and Munk (2008) indicate that studies with multiple treatments can pool the regression equations into one RDD model. However, doing so assumes that the multiple models are very similar in terms of their shape. When this is not the case, the overall model's shape may be quite different from the individual RDD models. To avoid complications from the potential misclassification of the model functional form, the researcher analyzed separate regression models for the two cutoffs.

Figure 2 represents the spread of pretest scores with a visual representation of the three separate treatment conditions, with each bar representing a bandwidth of five. The shaded area "A" represents assignment to the DI (ECRI) treatment group, while the shaded area "B" represents students assigned to the CAI reading instruction interventions (i-Ready for lower grades, Reading Plus for upper grades). The unshaded portion of the histogram "C" represents assignment to the control condition.

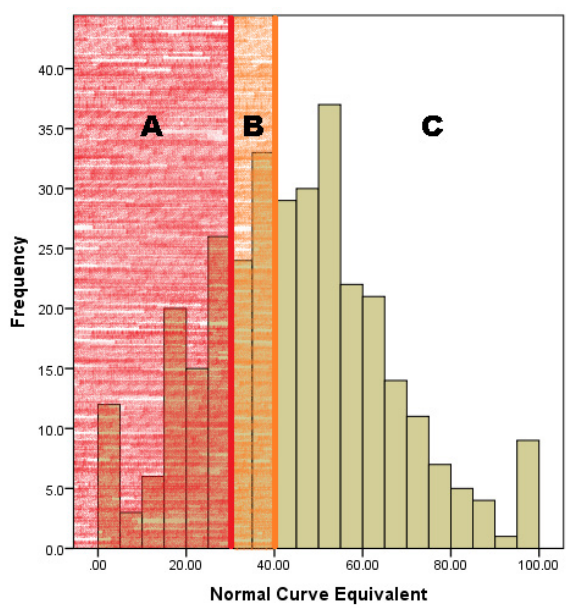


Figure 2. Histogram of the August KTEA-II BFR showing cutoff points and treatment conditions, with treatment identified as (A) direct instruction, (B) computer-assisted instruction, and (C) business-as-usual reading instruction.

Three regression models were analyzed separately for both cutoff points. The lower cutoff analyses included only sections A and B from the diagram while the upper cutoff analyses included sections B and C. The first analysis for each cutoff included all grade levels. To account for the difference in age of students and for the differences in CAI program and literacy targets for upper- and lower-grade students, the researcher also performed two subsequent analyses, stratified by grade in school: the lower-grade stratum included students in grades 1-3; the upper-grade stratum included only students in grades 4-6. These models had a much smaller sample size than the previous analyses, which resulted in reduced efficiency.

At the inception of the study, the research team had hoped to conduct single grade-level analyses of the data. However, the reduced sample size for each of the treatment conditions was an insurmountable barrier to performing an analysis with enough statistical power to detect potential treatment effects. Because of this, the analyses were restricted to the full sample and to the grade-level strata (upper grades and lower grades).

Another way of evaluating the results of an RDD is to compare the mean outcome variables of students directly above and below the cutoff. This is represented by the equation:

$$\bar{y}_{c-1} - \bar{y}_c \tag{7}$$

where $\bar{y}$ indicates the mean outcome variable, and the subscript indicates position relative to the cutoff. Were there to be any crossovers, the equation would also need to take into account the probability of receiving treatment (Jacob & Lefgren, 2004). However, since the current study uses a sharp RDD, the probability of receiving treatment shifts perfectly from 0 to 1 at the cutoff. Because of the small sample size due to narrower window widths, a one-way ANOVA was used to test the treatment effect for each RD analysis. The mean of the outcome variable for students directly below the cutoff was compared with the mean of the students directly above the cutoff.

Baseline characteristics. This study includes four baseline characteristics for the participants: (a) age, (b) gender, (c) ELL status, and (d) special education (SPED) status. Three of the covariates were dichotomous (Gender: 0 = female, 1 = male; ELL: 0 = non-ELL, 1 = ELL; and SPED: 0 = Non-SPED, 1 = SPED) while age was an interval-level covariate. To test the integrity of the local randomization near the cutoffs, robust means modeling (RMM) was employed as described by Fan and Hancock (2012) in order to identify any significant differences in the covariates above and below the cutoff points. RMM is more flexible than simple ANOVA since it does not assume equality of variance. The RMM made use of the Bonferroni corrected alpha because of the multiple comparisons.

Treatment of missing data. While every effort was made to obtain posttest scores for all participants, certain circumstances such as absences and mobility resulted in not all participants receiving one or both posttest scores. Due to these circumstances, thirty-seven students who participated in the pretest measure did not participate in the posttest. The missing data were a concern because of the resulting decreased statistical power.

Initially, the researcher intended to use Full Information Maximum Likelihood (FIML) to account for missing data. However, due to the inclusion of cubic terms in the regression equation, the FIML approach did not converge, even when using a Bayesian estimator. In place of FIML, multiple imputation (MI) with 100 iterations was employed, using all students who participated in the pretest and who had an assignment score. This method is the most preferable method for missing data imputation (McKnight, McKnight, Sidani, & Figueredo, 2007).

Move-ins. The research team hoped to be able to use the January iteration of the KTEA-II BFR as a starting point for students who moved in during the first two semesters, thereby increasing the overall sample size. Discussions with the school staff led to the decision to assign students immediately to treatment conditions upon being admitted to the school. A pretest was administered to these students, however they were not included in the final data analysis.

Effect size. In addition to estimating the LATE, R^2 statistics and Cohen's f^2 effect sizes were calculated for each RD analysis. Cohen's f^2 was calculated using the following equation:

$$f^2 = \frac{2R^2}{1 - R^2} \tag{8}$$

For the f^2 statistic, effect estimates of 0.02, 0.15, and .035 are classified as *small*, *medium*, and *large*, respectively (Cohen, 1988).

Clustering. During the intervention, participants were clustered by intervention leader. To test the effects of clustering and the necessity of multilevel considerations in the analysis, the researcher specified the intervention leader as the clustering variable (Table 6). Both intraclass correlations (ICC) and design effects (D_{eff}) were calculated for the truncated data at various window widths around the cutoffs. The D_{eff} was calculated using the equation:

$$D_{eff} = 1 + (m - 1)\rho \tag{9}$$

where m is the average number of observations in each cluster and ρ is the ICC.

Table 6

Intraclass Correlations (ICC) and Design Effects (D_{eff}) for January and May Posttest Iterations

January Posttest				May Posttest			
Window	*Mean Cluster Size*	*ICC*	D_{eff}	*Window*	*Mean Cluster Size*	*ICC*	D_{eff}
16-40	8.923	0.179	2.418	16-40	8.923	0.086	1.681
17-40	8.692	0.132	2.015	17-40	8.692	0.002	1.015
18-40	8.692	0.132	2.015	18-40	8.692	0.002	1.015
19-40	8.308	0.138	2.009	19-40	8.308	0.003	1.022
20-40	7.923	0.003	1.021	20-40	7.923	0.001	1.007
21-39	7.077	0.001	1.006	21-39	7.077	0.001	1.006
22-38	6.231	0.003	1.016	22-38	6.231	0.001	1.005
23-37	5.846	0.002	1.010	23-37	5.846	0.001	1.005
24-36	4.692	0.001	1.004	24-36	4.692	0.003	1.011

Note. Number of clusters is equal for all window widths (n = 13).

Only one window (16-40) had a D_{eff} of a magnitude more than .02 points greater than 2.00, indicating that the single-level design was more efficient in most cases. The ICC and the D_{eff} for the windows at the upper cutoff also indicated that the single-level model was more appropriate; all of the D_{eff} for the upper cutoff were below 2.00. Because of this, and due to the small sample size (Maas & Hox, 2005; Snijders, 2005), the multilevel model was rejected in favor of a single-level model.

Correction for discrete data. As described earlier, the assignment variable in the study could potentially be classified as discrete rather than continuous. In order to explore any potential misspecification issues and corresponding underestimated standard errors, the researcher applied cluster-robust standard errors (CRSE) as suggested by Lee and Card (2008). An analysis using CRSE was conducted using a series of window widths around the lower cutoff point and using the May posttest in order to act as a comparison for the original May analysis. The CRSE analysis was run in R using the *plm* package as described by Bluhm (2013). The CRSE results were compared with those of the original estimates to validate the significance of the findings.

Social validity. Because the purpose of the study was to evaluate the effectiveness of reading interventions on a schoolwide basis, the researcher believed that some measure of social validity would be important to the interpretation of the findings. Although outside the scope of this dissertation, the researcher was particularly concerned about the feasibility of the research design in the intervention setting as well as the appropriateness of paraeducators leading the interventions. Since the paraeducators were generally less trained than the licensed teachers, the question became whether or not the paraeducators could teach a rigorous DI intervention and run CAI sessions with high levels of fidelity. More specifically, the researcher was interested in the perceptions of both the paraeducators and the licensed teachers about the effectiveness of the interventions and research design.

To assess the social validity of the findings, the researcher conducted a series of focus groups and surveys intended to measure paraeducator and teacher impressions of the study. The survey consisted of 27 items and was delivered electronically to all paraeducators and licensed teachers in grades 1-6. Three focus groups were conducted at the end of the study (in May). The first group consisted of paraeducators. The second and third focus group consisted of lower-grade teachers (1-3) and upper-grade teachers (3-6), respectively. Focus groups lasted between 45 and 120 minutes in length. The focus groups were conducted by the researcher in the school's conference room and library, and were video-recorded. Focus groups began with a metaphor activity eliciting descriptions of the school's intervention program during the year. This was followed by an open-ended, loosely structured discussion.

The evaluation and analysis of the social validity results fall outside the scope of this dissertation. While some of the preliminary findings will be shared in the discussion section, the full results of the qualitative analysis will be submitted for publication in another paper.

Power analysis. Because of sample size limitations, a post-hoc power analysis was used to detect the minimum detectable effect size (MDES) that the sample was able to support (Lee & Munk, 2008). This was undertaken to evaluate the likelihood of Type II error if results were determined to have no statistical significance (May, Sirinides, Gray, & Goldsworthy, 2016). The statistical program G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) was used to conduct the post-hoc power analysis.

Chapter 4. Results

Model Assumptions and Data Descriptions

Normality and variance. The variance of the posttest measures supported the assumption of homoscedasticity. A slight ceiling effect and floor effect were apparent in the scatterplots of the predicted values and the residuals (Figure D1). [1] While the ceiling and floor effects were present in both measures, they appeared slightly more pronounced in the May iteration of the posttest. This may have been due to the greater spread of the residuals on the May posttest. With the use of local linear regression, the ceiling and floor effects did not pose a concern to the analysis, since the window widths used excluded these data points. Normal Q-Q plots indicated that the data were approximately normally distributed (Figure D2).

Continuity and integrity of the assignment variable. A graphical analysis of the histogram of the forcing variable did not indicate the presence of any large discontinuities at the cutoff points. Slight discontinuities at several other points on the histogram, including at the NCE values of 20 and 50 were tested for significance. Statistical tests at these points did not indicate that any of them were significant. The continuity of the assignment variable as indicated by the histogram was validated by a McCrary (2007) density plot which showed no discontinuities in the density of the forcing variable at the cutoff points (Figure D3).

Results of the robust means modeling indicated that none of the four baseline covariates (age, gender, special education status, ELL status) varied significantly above or below the lower cutoff. At the upper cutoff, special education status showed a significant difference above and below the cutoff ($\delta = .172$, $p < .01$), indicating a greater number of students receiving special education services below the upper cutoff. The Bonferroni corrected alpha was .0125.

---

[1] To facilitate simplicity of result reporting, tables and figures that provide helpful details but are not critical to the reader's understanding of the research questions have been included in Appendix D.

Correlations. Table 7 reports correlations for the assignment variable, outcome variables, and covariates included in the RD analyses. The assignment variable (August NCE) was highly correlated with both the January and May posttests. Special education status and ELL status were negatively correlated with the assignment variable and both posttest measures.

Table 7

Correlations among the Variables and Covariates used in the Regression-discontinuity Analyses

	August NCE	Jan. NCE	May NCE	SPED	ELL
August NCE	1.000				
January NCE	.851**	1.000			
May NCE	.791**	.884**	1.000		
SPED	-.312**	-.334**	-.349**	1.000	
ELL	-.385**	-.374**	-.368**	.115	1.000

Note. NCE = normal curve equivalent; SPED = special education status; ELL = English Language Learner.
** $p < 0.01$

Treatment dosage. Tables 8 and 9 summarize the usage statistics for the CAI Groups (Reading Plus and i-Ready, respectively). All students using Reading Plus completed a minimum of 55 lessons during the school year. Students using i-Ready completed a minimum of 60 lessons. All students in the DI intervention completed at least five units of ECRI teaching during the school year. The average number of units completed was 8.94 (SD = 2.92). The maximum number of units completed was 15. Word counts differed between lessons (Table 10), with completed word lists varying between 6 and 25 words in length.

Table 8

Dosage Statistics for Reading Plus Lessons by Lesson Type and Literacy Target

	Average Lessons		
Lesson Type	Attempted (SD)	Completed (SD)	>80% Acc. (SD)
Reading	113.05 (19.92)	93.20 (25.72)	51.20 (20.78)
I-Balance	56.80 (9.92)	38.40 (13.51)	53.15 (22.17)
Read Aloud	85.85 (14.81)	—	—

Note. Lessons completed and lessons completed with >80% accuracy were not reported for Read Aloud lessons. All targets are out of a possible 100, with the exception of mastered words.

Table 9

Dosage Statistics for i-Ready Lessons by Type of Lesson

	Number of Lessons		Lesson Accuracy	
Type of Lesson	*M*	*SD*	*M*	*SD*
Phonology	19.00	3.46	76.54	15.37
Phonics	22.62	6.05	96.38	8.10
High Frequency Words	22.85	5.38	89.92	6.43
Vocabulary	11.08	2.84	97.00	5.89
Comprehension	11.15	3.08	95.38	9.06
Total	86.69	18.46	90.15	6.20

Table 10

Lesson Statistics and Word Counts for Completed ECRI Units by Grade

	No. of Units Completed		Words per Unit			
Grade Level	*M*	*SD*	*Min*	*Max*	*M*	*SD*
First	7.17	1.67	11	25	15.43	4.05
Second	10.45	1.23	7	21	12.50	3.26
Third	13.33	2.36	10	21	14.60	2.99
Fourth	8.95	2.91	6	20	13.00	4.33
Fifth	11.78	3.04	7	16	11.20	2.59
Sixth	9.33	2.50	11	18	14.00	1.96
Lower Grades	8.61	2.62	7	25	13.99	3.59
Upper Grades	9.69	3.08	6	20	12.70	3.44
Total	8.94	2.92	6	25	13.46	3.59

Note. ECRI = Exemplary Center for Reading Instruction.

Functional form. The functional form of the regression was specified in two ways: using the full set of data and using a bandwidth of 10 points above and below the lower cutoff. The regression model tested the significance of cubic, quadratic and linear interaction and main effect terms. The researcher removed terms that were not significant sequentially until a statistically significant model was obtained. The results of the sequential testing of polynomial terms indicated that the cubic and quadratic terms were not significant. This was also true of the linear interaction term. Therefore, the remaining statistical analyses were run using only first-order terms, excluding the linear interaction effect. When the covariates were included, interaction effects for special education and ELL status were not significant. The final RD equation for the lower cutoff is shown in Equation 10.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 SPED_i + \beta_4 ELL_i + \varepsilon_i \quad (10)$$

where $SPED_i$ and ELL_i are special education and ELL covariates, respectively. Visual inspection of the January scatterplot confirmed the linear nature of the data (Figure 3).

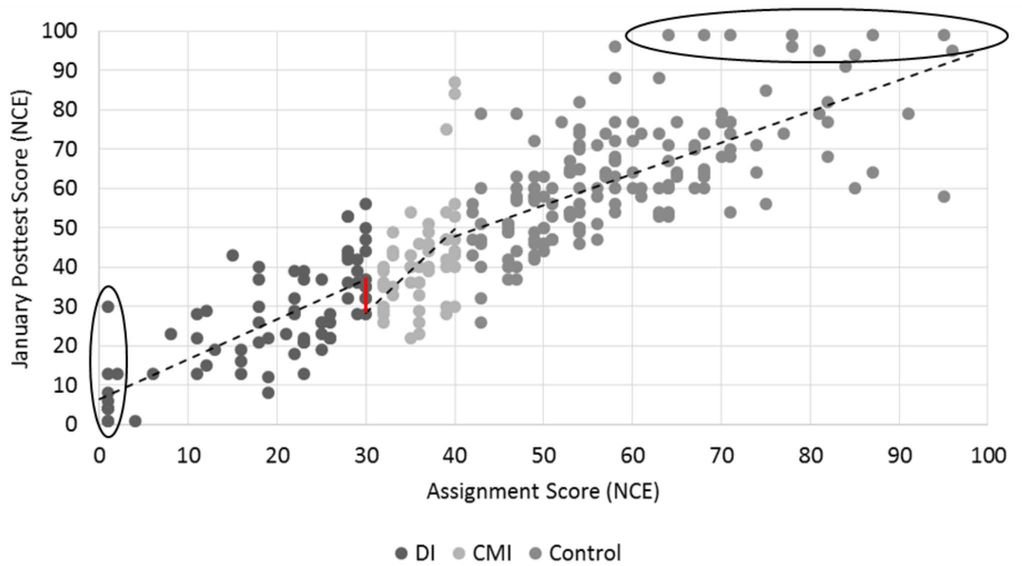


Figure 3. Scatterplot of assignment scores and January posttest scores for the full sample. Ovals indicate slight ceiling and floor effects, with red line indicating potential discontinuity.

Summary of Results

Table 11 summarizes the results of the full-sample RD analyses. In general, the main effect for the DI treatment was significant at both January and May posttests, but the main effect for CAI treatment was not. Significant DI treatment effects exceeded 10 points for both posttests and for grade-level strata. May treatment effects tended to be greater in magnitude than those in January. For the full sample, effect sizes were medium to large and tended to decrease as sample size decreased (Cohen, 1988). For grade-level strata, significant DI treatment effects were present but less stable across window widths, with medium-to-large effect sizes.

Table 11

Regression-discontinuity Estimates and Effect Sizes from the Analysis of the January and May Data.

		January Posttest								
		Full Sample			Lower-grade Stratum			Upper-grade Stratum		
Window	*N*	β	R^2	f^2	β	R^2	f^2	β	R^2	f^2
16-40	116	7.42¯	0.341	0.517	5.89	0.382	0.618	9.02	0.340	0.515
19-40	108	10.43*	0.337	0.508	10.01*	0.387	0.631	11.53	0.332	0.497
20-40	103	10.05*	0.278	0.385	10.58*	0.368	0.582	10.20	0.250	0.333
21-39	92	8.89*	0.225	0.290	10.43*	0.339	0.513	8.70	0.203	0.255
22-38	81	9.14*	0.201	0.252	10.68*	0.291	0.410	8.15	0.176	0.214
23-37	76	10.71**	0.202	0.253	12.72*	0.282	0.393	9.14	0.180	0.220
24-36	61	9.34*	0.164	0.196	10.31¯	0.143	0.167	9.66	0.282	0.393
25-35	52	11.75**	0.283	0.395	12.44*	0.206	0.259	12.35*	0.483	0.934
26-35	47	11.22*	0.207	0.261	11.16¯	0.149	0.175	12.61¯	0.412	0.701
27-35	42	5.01	0.095	0.105	-1.31	0.024	0.025	8.83	0.370	0.587
27-34	36	2.66	0.152	0.179	-3.77	0.058	0.062	10.43	0.358	0.558
		May Posttest								
Window	*N*	β	R^2	f^2	β	R^2	f^2	β	R^2	f^2
16-40	116	11.42*	0.276	0.381	8.70	0.286	0.401	14.51	0.293	0.414
19-40	108	13.30*	0.241	0.318	8.92	0.213	0.271	17.36¯	0.287	0.403
20-40	103	14.63**	0.247	0.328	11.69¯	0.248	0.330	17.94*	0.282	0.393
21-39	92	15.02**	0.221	0.284	13.36*	0.212	0.269	17.92*	0.273	0.376
22-38	81	16.60**	0.228	0.295	14.90*	0.204	0.256	19.34*	0.311	0.451
23-37	76	17.32**	0.206	0.259	15.87*	0.174	0.211	20.05**	0.305	0.439
24-36	61	16.93**	0.205	0.258	15.41¯	0.114	0.129	20.91**	0.493	0.972
25-35	52	18.27**	0.227	0.294	15.38¯	0.101	0.112	22.35**	0.533	1.141
26-35	47	14.59*	0.166	0.199	11.14¯	0.149	0.175	21.73**	0.544	1.193
27-35	42	9.67	0.158	0.188	-4.96	0.106	0.119	22.48*	0.455	0.835
27-34	36	8.09	0.203	0.255	-11.30	0.180	0.220	34.11**	0.607	1.545

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

Research Question 1: Lower Cutoff

A scatterplot of the data showed a visual discontinuity for both posttests (Figure 4). Results indicate that a significant discontinuity in the outcome measure was present at the lower cutoff across multiple window widths (Table D1). Treatment status was a significant predictor of performance on the January reading posttest (β = 10.05, p < .05). Treatment status was also a significant predictor of performance on the May reading posttest (β = 14.63, p < .01). Cohen's f^2 estimates indicated a medium effect size for both posttest iterations.

A one-way ANOVA was run using a bandwidth of four points (27-34) to compare the means of the outcome variable directly above and below the lower cutoff point. With a bandwidth of 3, the ANOVA results indicated the presence of a significant effect of treatment on reading ability as measured by the January posttest [F(1, 31) = 4.56, p = .041].

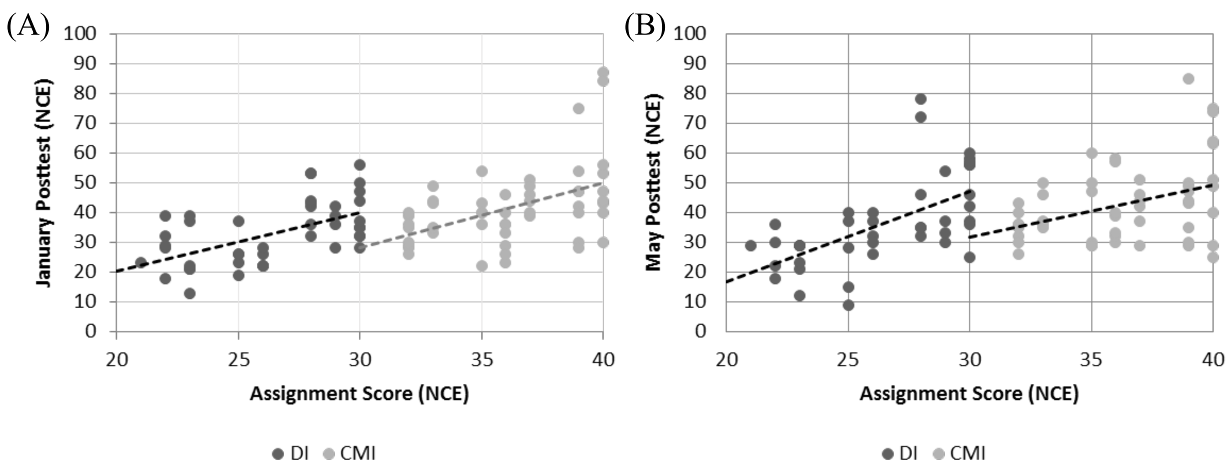


Figure 4. Scatterplots of assignment scores and outcome scores for (A) January and (B) May posttests for the full sample at the lower cutoff.

Lower-grade stratum. Scatterplots for the lower-grade stratum indicated a possible discontinuity at the lower cutoff point (Figure 5). The RD analysis substantiated the presence of a significant discontinuity at the lower cutoff point for both the January and the May posttest (Table D2). For a window width of 21-39, treatment status was a significant predictor of reading ability as measured by the January posttest (β = 10.43, $p < .05$) and the May posttest (β = 13.36, $p < .05$). Cohen's f^2 estimates indicated a medium effect size for both posttests.

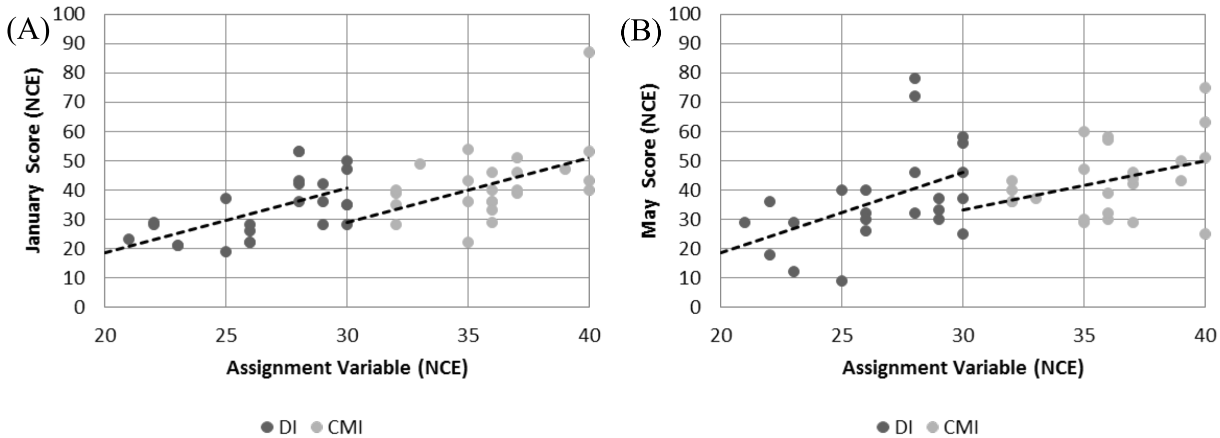


Figure 5. Lower-grade scatterplots of assignment scores and outcome scores for the (A) January and B) May posttest at the lower cutoff.

Upper-grade stratum. For the upper-grade stratum, discontinuities were also evident in the scatterplots (Figure 6). The RD analysis indicated that treatment status was not a significant predictor of reading ability in January (Table D3), except at a window width of 25-35 (β = 10.43, $p < .05$). Treatment status at the lower cutoff point was a significant predictor of the outcome variable in May (β = 17.92, $p < .05$). Cohen's f^2 indicated a large, unstable effect size in May.

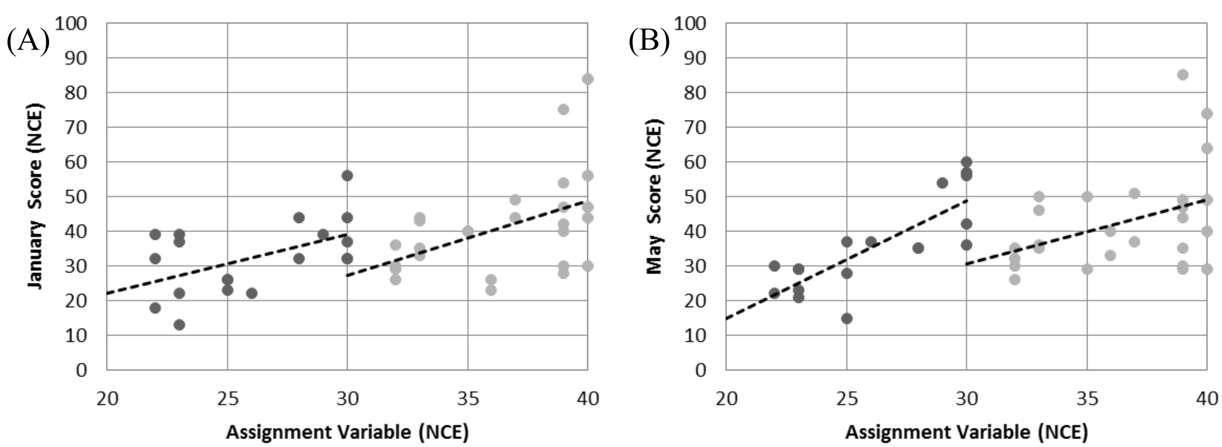


Figure 6. Upper-grade scatterplot of assignment scores and outcome scores for (A) January and (B) May posttests at the lower cutoff.

Research Questions 2 and 3: Upper Cutoff

Scatterplots of the upper cutoff did not indicate a significant discontinuity between CAI treatments and control groups (Figure 7). The results of the regression discontinuity analysis indicated that no significant CAI treatment effects were present at the January posttest across any of the window widths tested (Table D4). This was true of the May posttest results as well (Table D5). ELL and special education covariates were not significant predictors of achievement on the outcome in the presence of the pretest and treatment.

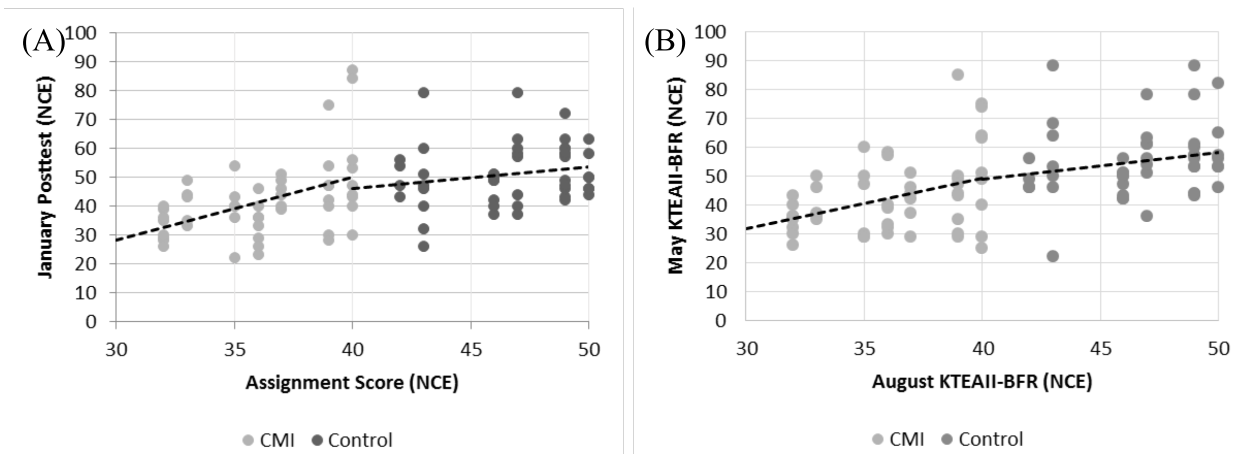


Figure 7. Scatterplots of assignment scores and outcome scores for the (A) January and (B) May posttests for the full sample at the upper cutoff.

Upper cutoff grade-level strata. Scatterplots of the grade-level strata at the upper cutoff point did not indicate a treatment effect for either of the CAI interventions. Likewise, results of the RD analysis indicated no significant treatment effect for either the upper- or lower-grade strata.

Post-hoc power analysis. The results of the post-hoc power analysis indicated that the MDES at the upper cutoff for the full sample was .145, indicating minimum possible detection of medium-to-large effect sizes. The MDES was .298 for the upper-grade stratum and .250 for the lower-grade stratum, indicating minimum possible detection of large effect sizes.

Research Question 4: ELL Analysis

At the lower cutoff, a series of RD analyses with an ELL interaction term and main effect for ELL status indicated that neither was a statistically significant predictor of posttest outcome scores in the presence of the pretest and treatment. The treatment effect did not vary significantly for students classified as ELL compared with the treatment effect for their peers. This was true for both iterations of the posttest. Similar results were obtained for both upper- and lower-grade strata (Table D6). The RD analyses at the upper cutoff did not indicate significant main effects for ELL status or interaction effects between ELL and treatment status. This was the case for both iterations of the posttest and for both upper and lower cutoff points. The stratified analyses at the upper cutoff did not indicate any significant ELL main effects or interaction effects with treatment assignment for either stratum.

CRSE Analysis

Results of the CRSE analysis followed the same pattern as the RD analysis without cluster-adjusted standard errors (Table 12). Treatment assignment was a significant predictor of achievement on the May posttest ($\beta = 13.34$, $p < .05$), as was special education status. These findings were not significantly different from the analysis conducted without CRSE adjustment, although standard errors were higher. The discontinuity was still significant across a wide range of window widths.

The estimates for treatment main effect using the cluster-adjusted standard errors were generally smaller in magnitude than estimates obtained without adjusting for the clustered nature of the data. Likewise, the R^2 and corresponding Cohen's f^2 values were also smaller in magnitude, ranging between .126 and .293. Main effects for special education status were greater in magnitude when using CRSE.

Table 12

Regression-discontinuity Estimates of the Impact of Direct Instruction Treatment on the May Posttest at the Lower Cutoff Using Cluster Robust Standard Errors

		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*P*	*β*	*β*	*β*	*p*	R^2
16-40	116	10.67*	5.07	.038	-8.45*	-0.52	1.51**	.000	.207
19-40	108	13.62*	5.76	.016	-7.64*	-1.10	1.86**	.000	.197
20-40	103	13.34*	5.64	.019	-6.98¯	-3.59	1.80**	.001	.171
21-39	92	14.66*	5.34	.011	-7.20¯	-0.35	2.03**	.000	.169
22-38	81	15.83**	5.98	.010	-4.93	-0.56	2.20**	.001	.147
23-37	76	14.57**	6.01	.018	-5.22	-0.58	1.99**	.004	.112
24-36	61	14.89**	6.24	.020	-9.89*	0.26	2.30**	.019	.140
25-35	52	17.20**	6.33	.009	-11.45*	-1.01	3.28**	.003	.227
26-35	47	16.05*	6.58	.019	-10.97¯	-2.80	2.96*	.035	.182
27-35	42	17.41¯	9.05	.061	-12.20¯	-2.28	3.41	.193	.175
27-34	36	22.02¯	12.69	.092	-16.20*	-2.43	5.34	.213	.244

Note. SPED = special education status; ELL = English Language Learner

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

Chapter 5. Discussion

Because of the unique nature of the RDD, any discussion of the implications of the results must be interpreted in light of how well the study meets the regression standards identified by Schochet et al. (2010). The first criterion was the integrity of the assignment variable. Statistical tests did not suggest that any manipulation of the forcing variable occurred, as indicated by both the histogram and the density plot (Figure D3). The assignment process precluded manipulation by the school staff or participants themselves, since participants and school staff were unaware of the cutoff points at the time of assessment, and the assignment to treatment was blind. The second criterion concerned attrition and differential attrition rates. Overall, attrition rates were lower than the school's traditional mobility rate, and attrition occurred equally for all treatment conditions.

The third criterion, the continuity of the assignment variable, was tested using RMM. The lower cutoff passed the test of continuity, but the proportion of students receiving special education services was not equal above and below the upper cutoff. Although this could be an indication of a discontinuity in the forcing variable at the upper cutoff, it is equally if not more likely that the increase in the number of students receiving special education services below the cutoff is due to the nature of why students receive special education services. A major factor in deciding whether or not a student receives special education services at the school was a persistent lack of academic achievement. As such, the distribution of students with IEPs by nature was skewed to the lower end of the forcing variable scores.

The last criterion of interest was the functional form of the regression. The researcher followed the recommendation of Shadish et al. (2002) to include two polynomial orders higher than the likely true functional form. Both the scatterplots of the data and the statistical tests indicated that the functional form was linear. Based on these criteria, the current study appears to meet the evidence standards suggested by Shochet et al. (2010) and the results can be interpreted with an assurance that the underlying research design was sound.

This study attempted to answer four research questions regarding the effect of the DI and CAI reading interventions on the reading ability of participating students. The following section discusses these questions in relation to the findings. These questions have been listed below in order to facilitate their discussion:

1. What effect does participation in ECRI have on the reading ability of students as measured by scores on the KTEA-II BFR?
2. What effect does participation in i-Ready have on students' reading ability as measured by scores on the KTEA-II BFR for students in grades 1-3?
3. What effect does participation in Reading Plus have on students' reading ability as measured by scores on the KTEA-II BFR for students in grades 4-6?
4. What effect does participation in the three reading interventions have on English Language Learners' reading ability as measured by scores on the KTEA-II BFR in grades 1-3 and grades 4-6?

Research Question 1: Direct Instruction (ECRI)

The results of the study indicate that participation in the ECRI intervention led to greater gains than participation in either CAI intervention for students at the lower cutoff. At midyear, a student at the lower cutoff point who participated in the DI intervention would have been expected to perform on average 10 points better on the KTEA-II BFR than a student in the CAI interventions. By the end of the school year, this treatment effect increased to approximately 14 points and exhibited more stability across window widths.

This effect was present for both upper and lower grades. Lower-grade students who participated in ECRI showed significant progress over their peers in i-Ready by January. This effect increased in May, although it did not appear to be as stable across window widths. While a positive effect was present for both upper and lower-grade ECRI participants, it appeared that the effect on upper-grade participants was larger but slower to appear than that for lower-grade students. A stable significant LATE was not present for upper-grade ECRI students at midyear, but by the end of the year, a very large LATE was present for the upper-grade students in the DI intervention.

These findings are similar to findings of other studies evaluating the effectiveness of DI reading programs. Many studies have found that DI instructional programs have positive effects on student learning outcomes compared with traditional methods of instruction (Jones-Carey, 2013; Stockard & Engelmann, 2012; Van Staden, 2011). This study adds credence to those findings. It should also be noted that DI effects on reading ability generally occur over extended periods of time and after prolonged intervention efforts, especially for at-risk readers (Alnahdi, 2015; Comings, 2015; Vaughn, Wanzek, Murray, & Roberts, 2012). This may explain why the treatment effect was not apparent for the upper grades until the latter part of the research study. The fact that a statistically significant effect was already present for the lower grades at mid-year indicates that the ECRI intervention may provide relatively early reading gains for younger students.

Research Questions 2 and 3: Computer-assisted Instruction

The second and third research questions focused on the effectiveness of the CAI programs. The results of this study indicate that these programs were not as effective as the ECRI intervention for students at the lower cutoff. This was true for both students who participated in i-Ready and in Reading Plus. The findings also suggest that the two programs were not statistically better or worse than business-as-usual reading activities. Students participating in these programs performed comparably to their peers just above the upper cutoff on both posttest measures. The lack of a significant effect calls into question whether or not the two CAI reading interventions in this study can effectively improve the reading ability of struggling readers better than business-as-usual reading activities. In the scope of the larger body of research on CAI reading program effectiveness, these findings align with previous studies that suggest that CAI programs have minimal effects on reading ability (Kulik, 2003; Slavin et al., 2009).

Research Question 4: English Language Learners

Results on the effectiveness of both DI and CAI programs on the reading ability of ELL students did not suggest that either intervention interacted with student ELL status either positively or negatively. The RD analysis using the full sample did not indicate that ELL status was a significant predictor of reading ability on the outcome measure in the presence of treatment. Treatment assignment did not appear to interact with ELL in a way that augmented or hindered student performance on the posttest measure. Analyses of the upper- and lower-grade strata did not indicate any significant ELL main or treatment effects.

Previous research has indicated that effective RtI programs for ELL students include components such as preteaching of vocabulary, language modeling, opportunities to respond using academic language and systematic, explicit instruction (Richards-Tutor, Baker, Gersten, Baker, & Smith, 2016; Sanford, Brown, & Turner, 2012). DI has traditionally been an effective method of providing these components, so the lack of an interaction effect between ELL status and DI participation leads one to question whether or not an effect was actually present, but undetected due to design or data limitations.

Unfortunately, relatively little research has been conducted on the impact of CAI reading interventions on the reading ability of ELL students. Research on ELL reading interventions has focused mainly on effective methods of instruction and not on the actual mode of delivery (Slavin & Cheung, 2003). More research is necessary to estimate specific impacts of CAI programs on the reading ability of ELL students.

In light of previous research, these results should not be interpreted to mean that the DI and CAI interventions had no differential effect on ELL students. As is shown in Figure D5, the ELL student distribution was positively skewed, with the majority of ELL students receiving Tier 2 intervention. In the current study, ELL status was identified as a dichotomous variable (1= ELL; 0 = non-ELL). Also, the students represented by the ELL designation were only students currently classified as ELL. The school did not provide the researcher with data on former ELL students, which means that the data on ELL students may be deficient. Because of this, the ability to detect any significant effects may have been hampered, especially considering the sample size limitations discussed later in this section.

Treatment Integrity

An important piece of these findings is that both types of intervention were implemented at approximately the same level of fidelity. This common level of treatment integrity is a strength of the study and adds credence to the findings. Paraeducators administered both DI and CAI interventions with 87% and 86% accuracy, respectively. In addition, the high level of interrater reliability indicates that the integrity measures are accurate representations of the level at which the interventions were administered in practice. While some level of deviance from a prescribed program is almost unavoidable, the treatment integrity data suggests a high level of adherence to the program.

Implications for Practice

School-specific implications. These findings are especially valuable as they pertain to the school in which the study took place because the results provide strong evidence that the DI program was more effective for struggling readers than the CAI programs that were used for Tier 2 interventions. In previous years, the school used these CAI programs for all levels of struggling readers. While we cannot make a direct comparison between the effectiveness of the ECRI intervention and the business-as-usual activities, these results are a strong message for school personnel to thoughtfully select and implement appropriate reading interventions for at-risk readers.

As previously stated, the current study was a follow-up to a semi-randomized control trial that attempted to evaluate the effectiveness of the same reading programs at the school the previous year (Jones et al., 2014). The treatment effects identified in this study are similar to those that were found in the previous study. However, the current design was far superior to that of the previous study. The improvement in design is partially due to the high level of internal validity that was achieved. The study exhibited high levels of internal validity, including a well-executed assignment process, few crossovers, and careful attention to the requirements of a strong RDD. These findings should give the school confidence that using the DI intervention for the most at-risk students is a better option than using the CAI interventions.

General implications. The importance of these findings also lies in the purpose of RtI programs. One of the major objectives of Tier 2 reading interventions is to close the achievement gap between at-risk readers and students who read on grade level. In the current study, the results do not suggest that the CAI programs helped the students who participated in them to catch up to their peers. On the other hand, while no direct comparison between control and DI interventions is possible, the results do suggest that, at least for students near the lower cutoff, ECRI did help participating students improve their reading ability much more than would have been expected if they had used the CAI programs.

The use of computer-based instructional programs in schools has been growing at an astounding rate (Cheung & Slavin, 2012). Teacher education programs and initiatives place a high priority on integrating technology into the classroom. Indeed, many companies promote their technology and programs as "evidence based" interventions, prompting the What Works Clearinghouse to offer reports on the effectiveness of specific computer-based programs, such as Reading Plus. However, the current emphasis on computer-based interventions does not mean that computer instruction is or will be able to replace certain effective instructional practices, especially for student populations with more specific needs, such as ELL students.

It should be noted that the two CAI programs in the study make no claims as to their ability to replace or to supersede effective teachers and school personnel in interventions, although they do market themselves as evidence-based reading programs suitable for interventions. While technology may indeed offer educators helpful remediation tools, the current study offers a word of caution in using CAI as a Tier 2 intervention when resources may be more effectively used to provide better interventions to struggling readers, and substantiates the findings of Slavin et al. (2009). It may also suggest that such programs may be more effective in combination with other instructional practices (Cheung & Slavin, 2012).

The use of paraeducators is particularly pertinent to the interpretation of these findings because of the manner in which they were trained and utilized over the course of the study. As

Giangreco (2013) noted, paraeducators may not be effective in administering interventions without certain key components. As was discussed in the literature review, these components include using paraeducators to supplement teacher instruction using evidence-based, professionally prepared materials, with ongoing training in treatment fidelity and behavior management training.

During this study, the paraeducators received continual and focused training and feedback on their performance in administering both the DI and the CAI interventions. While the nature of the trainings differed (e.g., the DI training focused on teaching the scripted lesson while the CAI training focused on logistics and procedures), paraeducators administering both types received feedback on pacing, fidelity to the program, behavior management, and logistics (e.g., where the children sat relative to the paraeducator, setup of computers for use by students, etc.). In general, the training for both DI and CAI interventions was intended to foster high levels of treatment fidelity, and the overall framework in which the paraeducators were used contained all of the components discussed by Giangreco (2013). While the focus of this dissertation was not the effective use of paraeducators, the fact cannot be overlooked that the results of the interventions may have been very different without professionally prepared materials and continual training and support provided to the paraeducators.

Limitations

Generalizability. While it might seem tempting to extrapolate the results of the lower cutoff to make a comparison between the DI intervention and the business-as-usual control group, in reality such a comparison is impossible given the research design (Shadish et al., 2002). One might use the logic that since the DI intervention was clearly superior to the CAI interventions at the lower cutoff, then it must be also true that DI intervention would be superior to the control. However, the results do not support the inference that the treatment is equally effective or ineffective across the entire range of assignment variable scores. Nor can one make a direct comparison between students in the DI intervention and those in the control condition. The design followed in this study does not support such a comparison, and treatment effects are limited to participants located around the cutoffs.

Second, the results of the study cannot be extrapolated to all small-group DI, or CAI programs. Rather, the analysis is limited to the current interventions as administered in a 1-hour reading block. Some studies have shown that CAI programs can be more effective than teacher-led reading interventions (Martin, Elfreth, & Feng, 2014). Other studies have indicated that the use of computerized reading programs can lead to reading gains, although they do not automatically translate into improved reading skills (Rouse & Krueger, 2004). Computer-assisted reading programs are extremely diverse, and it seems ineffectual to generalize the findings of one program to another because the sheer amount of differences between programs render such comparisons problematic. Likewise, the results cannot be generalized to all forms of RtI implementation.

Assignment variable limitations. A particular concern with the KTEA-II BFR is that it may not adequately measure the construct of reading ability in ways that allow us to make specific claims about the reading interventions in question. The effect of the DI and CAI programs on specific reading aspects (e.g., vocabulary, fluency, phonemic awareness) was not tested. Rather, the KTEA-II BFR tested a more general level of reading ability as measured by word recognition and comprehension/recall items. As such, the results do not indicate if the DI intervention was more effective across all reading targets or only some of them.

The imprecision of the outcome variable is particularly limiting when evaluating the effects of the interventions on the reading ability of ELL students. Research suggests that some literacy skills, such as vocabulary, may have much more of an impact on the reading ability of ELL students than others (Richards-Tutor et al., 2016; Sanford, Brown, & Turner, 2012). However, this study does not provide a basis for making specific decisions on the paring of specific reading programs with specific ELL literacy targets.

The scope of this study was not to identify particular strengths of the reading intervention as they relate to specific language arts targets. Rather, the research was intended to provide a more general evaluation of the relative effectiveness of each reading intervention in the context of the school's RtI efforts. While the KTEA-II BFR may not have been able to identify treatment effect on specific reading skills, it still appeared to be suitable for providing a measure of comparison between the general reading ability of the various assignment groups.

A specific concern with using the same measure for both assignment and posttest is that scores on one will likely be highly correlated with scores on the other. In the current study, the assignment variable (NCE) was highly correlated with both the January posttest (r = .851) and with the May posttest (r = .791). Deke and Dragoset (2012) note that outcome measures that are highly correlated can affect the magnitude of the regression discontinuity design effect (RDDE), which is a measure of the discrepancy between RDD impact relative to an equivalent RCT impact. High level of correlation of variables in the RD equation tend to inflate the RDDE, which in turn increases the MDES for the model. This additional loss of efficiency could have increased the probability of Type II error at the upper cutoff, especially considering the small sample size.

Sample size limitations. While the RD results were similar across window widths and grade-level strata, the stability of the results varied as the sample size decreased. This instability of results may be attributed in large measure to the small sample size. As it is, the sample size in the current study limited the ability of the analysis to accurately estimate an effect size across all window widths because the statistical power was greatly reduced. While this does not negate the presence of a large treatment effect for the DI treatment (as was found to be present at the lower cutoff), it may have impacted our ability to accurately detect a small effect (either main or interaction) at the upper cutoff by essentially obscuring its presence. The results of the post-hoc power analysis indicated that the MDES at the lower cutoff was .145, suggesting that a smaller effect may have been present but undetected due to the lack of statistical power.

The limitation of sample size is also evident in the instability of the LATE in the smaller window widths. For example, the RD analysis for the upper-grade stratum of the May pretest suggests quite a range of values for the LATE, from 19.1 to 34.1, a spread of over 15 points. Because of the small sample size, individual data points gain much more influence on the outcome when sample size is reduced. With fewer than 50 individual data points, even a slight change in the window size could have a large impact on the LATE. As a result, the effect size estimates at the narrow window widths for the grade-level strata should be interpreted with caution.

Another limitation relating to the small sample size is that effect of treatment could not be reliably calculated for single grade levels. This is true for two reasons. First, grade-level sample sizes were too small to be able to conduct a RD analysis. Second, even if the number of students in each grade had been higher, the proportion of students in each treatment condition was often quite imbalanced. Since the assignment variable was not grade-specific, some grades only had a handful of students in a particular treatment condition. For example, in third grade only three students scored below the lower cutoff score. Because of this, a RD analysis was impractical for most individual grade levels.

Fidelity and dosage limitations. The variability in effectiveness of individual paraeducators was evident in both the level of treatment integrity exhibited by the paraeducators and also the qualitative notes and preliminary social validity findings. Although overall levels of treatment integrity were high, individual levels for several of the paraeducators were much lower. This was particularly true for the paraeducators who left the school during the study. Seven of the paraeducators had treatment integrity levels above 80%, while three had fidelity levels below 70%. CAI integrity was more stable, with most paraeducators averaging above 85%. This suggests that these interventions were more easily implemented with fidelity. Because of this, the actual treatment dosage for both DI and CAI programs may have varied from student to student depending on the paraeducators assigned to each grade. Unfortunately, the sample size limited the use of multilevel analysis.

Observers also noted that the actual pacing of instruction varied significantly between paraeducators. Some paraeducators struggled to maintain an effective pedagogical pace, which resulted in a discrepancy between the number of ECRI units completed by different grade levels. While not perfectly correlated, paraeducators with low treatment integrity scores tended to spend much more time teaching individual words and lessons in the DI intervention. This may have been due to both treatment fidelity issues and training concerns. While paraeducators received substantial training and reinforcement on DI instructional practices, the preliminary social validity evidence suggests that behavior management concerns were an issue that plagued many of the paraeducators throughout the interventions, both DI and CAI. This has been identified as an area that needs attention if paraeducators are to be used effectively in interventions (Giangreco, 2013).

In addition to fidelity concerns, it appeared that students in the upper grades spent 5-7 minutes more on average per session in DI interventions than those in the CAI intervention. The difference in time raises the question, at least in the upper grades as to whether or not the time rather than the intervention itself was the cause of the discontinuity. While this is a valid concern, it should be noted that qualitative observation notes indicate that some of this extra time was due to the aforementioned slow pacing and management concerns. In addition, the actual reading practice time for all students was the same each day and did not extend beyond the one-hour time allotment. No students received extra reading intervention time during the school day.

Further Research

RtI considerations. A number of practical changes might make the implementation of an RDD more feasible in school-level RtI interventions. First, schools may wish to vary the amount of time allotted to a program evaluation. In this particular study, the DI program's effectiveness showed up at the midpoint of the study. While not every intervention would show such an immediate effect, it may be the case that RD studies could be undertaken over shorter timeframes. Researchers would need to balance sample size concerns with the likelihood of a large program effect. Programs that are not likely to provide a large effect may not be good candidates for individual school RD analysis studies, especially if the time between assignment to treatment and posttest is short. In addition, the nature of the learning target would also affect whether or not a reduced timeframe would be feasible.

Another option to enable crossovers in an RtI setting would be for researchers to implement a fuzzy regression discontinuity design. In the current study, less than 15 students who originally did not qualify for RtI services based on the August pretest would have qualified based on their January posttest score. Of these, only four students were not receiving schoolwide Tier 2 reading interventions. It is common in RtI practice to move students out once they have mastered learning targets and move new students who are struggling into appropriate interventions. In order for the RDD to provide an accurate picture of a program's effectiveness, schools would need to be creative about how they managed the crossovers to limit the impact on the design's integrity, while also serving the needs of all students.

ELL considerations. One question that the study did not adequately answer was that of the effectiveness of these interventions for specific subgroups of students, ELL students in particular. As explained, sample size limitations were one of the primary causes of the study's inability to answer this question. To mitigate this concern, a multilevel RDD could be implemented across several schools, or even a district, thereby increasing the sample size. It may also be the case that the RDD is not the best tool to answer the question about the effectiveness of Tier 2 RtI programs on the reading ability of students designated as ELL. However, the RDD could be used in tandem with other strong research designs, such as the RCT, to more specifically target certain populations such as ELL students.

When targeting ELL students, the overall proficiency level of the student should also be taken into consideration. As such, the various classification levels could provide a basis for identifying levels of proficiency, a task which would be better undertaken with higher statistical power (Figure D5). Within this classification, students formerly classified as ELL could also be included in order to offer a better picture of the effects of Tier 2 interventions on the full ELL population.

August, Shanahan, and Escamilla (2009) note that current research trends heavily favor monolinguistic research “thereby placing language minority students (who are developing bilinguals) on the margins of educational theory and practice” (p .450). While it is tempting to view ELL students as simply monolingual students with a lower level of English proficiency, this view obscures the fact that bilingual students have particular reading needs that have not been fully explored in the current body of intervention research (August, Carlo, Dressler, & Snow, 2005).

Focused learning targets. As previously stated, the current study does not make claims as to the effectiveness or ineffectiveness of specific reading interventions on specific learning targets. While this study provides a positive example of using a design that is well-suited to establishing causal claims, future research is needed in order to answer questions regarding the effect of these interventions on specific learning targets such as phonics, phonemic awareness, comprehension, vocabulary, and fluency. This is partially due to the single cutoff variable and single posttest measure that was used in the study.

Advances in the field of RD analysis have also included designs that include multiple cutoff variables as well as multiple outcome variables (Shadish et al., 2011; Wong, Steiner, & Cook, 2013). Because school RtI programs typically rely on multiple measures to assign students to interventions, the RDD could be modified to include these additional assignment variables. The RDD could also be modified to explore the relationship of specific reading interventions to specific reading targets by including more precise outcome measures, or combinations of outcome variables.

Mediation analysis. Treatment dosage is of particular interest in the current study. The majority of paraeducators followed the intervention programs with high levels of fidelity.

However, a significant number of them struggled to maintain treatment integrity. While not in the scope of this dissertation, a highly interesting study may be the use of the treatment integrity data as part of a mediation analysis. It is unclear whether such an analysis would be possible given the sample size, but the results would be informative into the impact of paraeducators on the learning outcomes of the students.

Paraeducators in RtI. The effectiveness of individual paraeducators seems to support the findings of Giangreco (2013) that paraeducators need continued support, training, and structure to be effectively used in RtI settings. However, the variability between paraeducators in their treatment integrity suggests that the question still remains into how much and what types of training are most effective in fostering the successful use of paraeducators in RtI programs. The preliminary analysis of the social validity focus group data indicates that a particular need that was not adequately met during the training of the paraeducators was classroom and behavior management. It is hoped that the full analysis of the qualitative data will provide insights into the training needs and the experiences of the paraeducators, and offer suggestions to enable practitioners to improve their use of support staff in RtI settings.

Conclusion

Notwithstanding the limitations heretofore delineated, the results of the RD analyses provide insight into the impact and practice of three specific RtI reading interventions in elementary school. The stability of the ECRI treatment effect across all of the analyses is a strong support of the ECRI intervention as a tool to support and promote reading ability in at-risk readers. The CAI programs i-Ready and Reading Plus did not seem to provide additional support to student reading above and beyond business-as-usual reading activities, although the study lacked the statistical power to detect anything but a medium to large effect size.

RtI is and will continue to be a particular topic of research and investigation. As this dissertation was being prepared, the Institute of Education Sciences published a large-scale study that explored the effectiveness of RtI programs using an RDD (Balu et al., 2016). The IES study implemented an RDD across a large number of schools with a sample size of over 20,000 students, although the design was implemented after treatment had been completed and the data were observational in nature. The authors found that the effectiveness of RtI programs varied drastically between schools. This variance in effectiveness seems to be evidence of the need for additional evaluations of additional evaluations of RtI programs. While large-data studies have significant advantages over smaller-scale ones, they also may gloss over finer points of distinction between sites.

The current study suggests that the RDD can be used successfully in small-scale education settings, such as schoolwide RtI evaluations, to estimate treatment effects of interventions. While sample size considerations were a notable limitation, in cases where the effect size is expected to be medium or larger, the RDD may continue to be useful as part of both small-scale and large-scale RtI program evaluations.

This dissertation has been an attempt to offer a thoughtful and well-designed evaluation of an elementary school's reading interventions. In the context of a strong causal design, it is hoped that the findings may add to the literature on RtI reading programs in a substantial way and encourage others to think creatively about how to apply experimental or strong quasi-experimental research designs to RtI settings in order to improve research efforts in the field of multi-tiered systems of support.

References

Alnahdi, G. H. (2015). Teaching reading for students with intellectual disabilities: A systematic review. *International Education Studies*, *8*(9), 79-87. doi: 10.5539/ies.v8n9p79

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, *20*, 50-57. doi: 10.1111/j.1540-5826.2005.00120.x

Baker, S. K., Smolkowski, K., Chaparro, E. A., Smith, J. M., & Fien, H. (2015). Using regression discontinuity to test the impact of a tier 2 reading intervention in first grade. *Journal of Research on Educational Effectiveness*, *8*, 218-244. doi: 10.1080/19345747.2014.909548

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R., National Center for Education Evaluation and Regional Assistance (2015). *Evaluation of response to intervention practices for elementary school reading.* (NCEE 2016-4000). Washington, DC: Author.

Berkeley, S., Bender, W., Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities*, *42*, 85-95. doi: 10.1177/0022219408326214

Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness, 5*, 43-82. doi: 10.1080/19345747.2011.578707

Bluhm, R. (2013, October). *Clustered ses in R and Stata.* Retrieved from http://www.richard-bluhm.com/clustered-ses-in-r-and-stata-2/

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice (2nd Ed.).* New York, NY: Guilford.

Chambers, E. A. (2003). *Efficacy of educational technology in elementary and secondary classrooms: A meta-analysis of the research literature from 1992-2002* (Doctoral dissertation, Southern Illinois University at Carbondale). Retrieved from http://search.proquest.com.erl.lib.byu.edu/docview/275959473

Cheung, A. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*, 198-215. doi: 10.1016/j.edurev.2012.05.002

Cohen J. E. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. doi: 10.1037/t01756-000

Comings, J. P. (2015). An evidence-based model for early-grade reading programmes. *Prospects: Quarterly Review of Comparative Education*, *45*, 167-180. doi: 10.1007/s11125-014-9335-9

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.jeconom.2007.05.002

Coyne, M. D., Zipoli Jr., R. P., Chard, D. J., Faggella-Luby, M., Ruby, M. F., Santoro, L. E., & Baker, S. (2009). Direct instruction of comprehension: Instructional examples from intervention research on listening and reading comprehension. *Reading & Writing Quarterly*, *25*, 221-245. doi: 10.1080/10573560802683697

Crevecoeur, Y. C., Coyne, M. D., & McCoach, D. B. (2014). English language learners and English-only learners' response to direct vocabulary instruction. *Reading & Writing Quarterly*, *30*, 51-78. doi: 10.1080/10573569.2013.758943

Curriculum Associates (2013). *I-Ready diagnostic New York state validity study*. Retrieved from http://www.casamples.com/downloads/NY-iReadyValidityReportExecutiveSummary.pdf

Dağseven Emecen, D. (2011). Comparison of direct instruction and problem solving approach in teaching social skills to children with mental retardation. *Educational Sciences: Theory and Practice*, *11*, 1414-1420.

Deke, J., & Dragoset, L. (2012). Statistical power for regression discontinuity designs in education: Empirical estimates of design effects relative to randomized controlled trials. (Working paper No. 8). Retrieved from Mathematica Policy Research, Inc. website https://www.mathematicampr.com/~/media/publications/pdfs/education/statpowerregdis_wp.pdf

Fan, W., & Hancock, G. R. (2012). Robust means modeling an alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality. *Journal of Educational and Behavioral Statistics*, *37*, 137-156. doi: 10.3102/1076998610396897

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. doi: 10.3738/BRM.41.4.1149

Flores, M. M., & Ganz, J. B. (2009). Effects of direct instruction on the reading comprehension of students with autism and developmental disabilities. *Education and Training in Developmental Disabilities*, *44*, 39-53.

Flores, M. M., Nelson, C., Hinton, V., Franklin, T. M., Strozier, S. D., Terry, L., & Franklin, S. (2013). Teaching reading comprehension and language skills to students with autism spectrum disorders and developmental disabilities using direct instruction. *Education and Training in Autism and Developmental Disabilities*, *48*, 41-48.

Flynn, R., Marquis, R., Paquet, M., Peeke, L., & Aubry, T. (2012). Effects of individual direct instruction tutoring on foster children's academic skills: A randomized trial. *Children and Youth Services Review*, *34*, 1183-1189. doi: 10.1016/j.childyouth.2012.01.036

French, N. K. (2001). Supervising paraprofessionals a survey of teacher practices. *The Journal of Special Education*, *35*, 41-53.

Fuchs, D, Mock, D., Morgan, P.L., & Young, C. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct [Electronic version]. *Learning Disabilities Research & Practice*, *18*, 157-171. doi: 10.1111/1540-5826.00072

Giangreco, M. F. (2013). Teacher assistant supports in inclusive schools: Research, practices and alternatives. *Australasian Journal of Special Education*, *37*(2), 93-106. doi: 10.1017/jse.2013.1

Giangreco, M. F., Edelman, S. W., Broer, S. M., & Doyle, M. B. (2001). Paraprofessional support of students with disabilities: Literature from the past decade. *Exceptional Children*, *68*, 45-63.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/

Goss, C. L., & Brown-Chidsey, R. (2011). Tier 2 reading intervention: Comparison of Fundations and Reading Mastery. *Preventing School Failure*, *56*(1), 65-74. doi: 10.1080/1045988X.2011.565385

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*, 201-209.

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention: A research-based summary. *Theory into Practice*, *50*, 4-11. doi: 10.1080/00405841.2011.534909

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0 [Computer software]. Armonk, NY: IBM Corp.

Imbens, G. W., & Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*, 615-635. doi: 10.1016/j.jeconom.2007.05.001

i-Ready® Diagnostic & Instruction [Computer software]. (2013). North Billerica, MA: Curriculum Associates, LLC.

Jennings, J. H., Caldwell, J. S., & Lerner, J. W. (2013). *Reading problems: Assessment and teaching strategies*. Boston, MA: Pearson Higher Ed.

Jones, E. A., Young, K. R., Gibb, G., & Ottehenning, L. (2016). *The effects of tier 2 interventions, delivered by paraeducators, on elementary students at risk for reading failure.* Manuscript submitted for publication.

Jones-Carey, M. H. (2013, January 1). Direct Reading Instruction and the NYS ELA. (Doctoral dissertation, D'Youngville College). Retrieved from: http://search.proquest.com.erl.lib.byu.edu/docview/1508442605?accountid=4488

Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., … & Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly*, *30*, 153-168.

Kaufman, A., & Kaufman, N. (2005). *Kaufman Test of Educational Achievement Second Edition, Brief Form Manual*. Bloomington, MN: NCS Pearson, Inc.

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, *49*, 342-352. doi: 10.1002/pits.21602

Kulik, J. A. (2003*). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say.* SRI Project Number P10446.001. Arlington, VA: SRI International.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential– but often forgotten– component of school-based interventions. *Preventing School Failure*, *48*(3), 36–43. doi: 10.3200/PSFL.48.3.36-43

Lane, K. L., Fletcher, T., Carter, E. W., Dejud, C., & Delorenzo, J. (2007). Paraprofessional-led phonological awareness training with youngsters at risk for reading and behavioral concerns. *Remedial and Special Education*, *28*, 266-276. doi: 10.1177/07419325070280050201

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*, 655-674. doi: 10.1016/j.jeconom.2007.05.003

Lee, D. & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48,* 281-355. doi: 10.3386/w14723

Lee, H., & Munk, T. (2008). *Using regression discontinuity design for program evaluation*. Section on Survey Research Methods, Joint Statistical Meetings. Denver, CO: American Statistical Association.

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

Liu, M., Moore, Z., Graham, L., & Lee, S. (2002). A look at the research on computer-based technology use in second language learning: A review of the literature from 1990–2000. *Journal of Research on Technology in Education*, *34*, 250-273. doi: 10.1080/15391523.2002.10782348

Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation. *Computer Assisted Language Learning*, *19*, 15-45. doi: 10.1080/09588220600803998

Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86-92. doi: 10.1027/1614-2241.1.3.86

Macmillan/McGraw-Hill. (2007). *Reading Triumphs*. New York, NY: Author.

Martin, C., Elfreth, J., & Feng, J. (2014*). Effect on reading fluency of struggling third grade students: Computer-assisted intervention versus teacher-guided intervention.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN. Retrieved from: http://files.eric.ed.gov/fulltext/ED548040.pdf

Martínez, R. S., Nellis, L. M., & Prendergast, K. A. (2006). Closing the achievement gap series: Part II. Response to intervention (RtI)--basic elements, practical applications, and policy recommendations. *Center for Evaluation and Education Policy, 4*, 1-6.

Matthews, M. S., Peters, S. J., & Housand, A. M. (2012). Regression discontinuity design in gifted and talented education research. *Gifted Child Quarterly*, *56*, 105-112. doi: 10.1177/0016986212444845

May, H., Sirinides, P. M, Gray, A., & Goldsworthy, H. (2016). Reading recovery: An evaluation of the four-year i3 scale-up. *Consortium for Policy Research in Education*. University of Pennsylvania. http://www.cpre.org/sites/default/files/reading_recovery_final_report.pdf

McKnight, P., McKnight, K, Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction.* New York, NY: The Guilford Press.

Murnane, R. J. & Willet, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York: Oxford University Press.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide* (Sixth Edition). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics (2015). *The condition of education.* Retrieved May 3, 2016 from http://nces.ed.gov/programs/coe/

R Core Team (2016). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ralston, N. C., Benner, G. J., Nelson, J. R., & Caniglia, C. (2009). The effects of the "language arts" strand of the "reading mastery signature series" on the reading and language skills of English language learners. *Journal of Direct Instruction*, *9*, 47-55.

Reading Plus® [Computer software]. (2013). Winooski, VT: Taylor Associates/Communications, Inc.

Reid, E. R. (1995). *Teaching scheduling and record keeping.* Salt Lake City, UT: Cove.

Reid, E. R. (1996). *Teaching new words through the word structure methods.* Salt Lake City, UT: Cove.

Reid, E. R. (1997). Exemplary Center for Reading Instruction (ECRI*). Behavior and Social Issues*, *7*, 19-24. doi: 10.5210/bsi.v7i1.294

Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children*, *82*, 144-169. doi: 10.1177/0014402915585483

Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, *23*, 323-338. doi: 10.1016/j.econedurev.2003.10.005

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, *25*, 279-292. doi: 10.1016/0378-3758(90)90077-8

Rupley, W., Blair, T., & Nichols, W. (2009). Effective reading instruction for struggling readers: The role of direct/explicit teaching. *Reading & Writing Quarterly: Overcoming learning difficulties*, *25*, 125-138. doi: 10.1080/10573560802683523

Ryder, R., Burton, J., & Silberg, A. (2006). Longitudinal study of direct instruction effects from first through third grades. *Journal of Educational Research*, *99*, 179-191. doi: 10.3200/JOER.99.3.179-192

Sanford, A. K., Brown, J. E., & Turner, M. (2012). Enhancing instruction for English learners in response to intervention systems: The PLUSS model. *Multiple Voices for Ethnically Diverse Exceptional Learners*, *13*(1), 56-70. doi: 10.5555/muvo.13.1.0780760121003732

Sansosti, F. J., Noltemeyer, A., & Goss, S. (2010). Principals' perceptions of the importance and availability of response to intervention practices within high school settings. *School Psychology Review*, *39*, 286-295.

Savage, R., & Carless, S. (2005). Learning support assistants can deliver effective reading interventions for ‘at-risk’ children. *Educational Research*, *47*, 45-61. doi: 10.1080/0013188042000337550

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics*, *33*, 62-87. doi: 10.3102/1076998607302714

Schochet, P. Z., Cook, T., Deke, J., Imbens, G., Lockwood, J.R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs. Retrieved January 13, 2016, from http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf

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

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

Shippen, M., Houchins, D., Steventon, C., & Sartor, D. (2005). A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education*, *26*, 175-182. doi: 10.1177/07419325050260030501

Slavin, R. E., & Cheung, A. (2003). *Effective reading programs for English language learners: A best-evidence synthesis.* Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed At Risk.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). *Effective beginning reading programs: A best-evidence synthesis.* Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.

Slavin, R.E., Lake, C., Cheung, A., & Davis, S. (2009). *Beyond the basics: Effective reading programs for the upper elementary grades.* Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B.S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3. pp. 1570-1573). New York: Wiley. doi: 10.1002/0470013192.bsa492

Star Reading® [Computer software]. (2013). Wisconsin Rapids, WI: Renaissance Learning, Inc.

Stockard, J., & Engelmann, K. (2010). The development of early academic success: The impact of direct instruction's "reading mastery". *Journal of Behavior Assessment and Intervention in Children*, 1, 2-24. doi: 10.1037/h0100357

Stuebing, K. K., Fletcher, J. M., & Hughes, L. C. (2012). Meta-analysis and inadequate responders to intervention: A response. *Journal of Learning Disabilities*, *45*, 565-569. doi: 10.1177/0022219412451999

Swanson, H. L. (2012). Meta-analysis and inadequate responders to intervention: a reply. *Journal of Learning Disabilities*, *45*, 570-575. doi: 10.1177/0022219412452097

Reading Plus® [Computer software]. (2013). Winooski, VT: Taylor Associates/Communications, Inc.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*, 309-317. doi: 10.1037/h0044319

Tran, L., Sanchez, T., Arellano, B., & Swanson, H. L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*, 44, 283-295. doi: 10.1177/0022219410378447

Trochim, W. M. K. (2006). The Regression Discontinuity Design. Retrieved September 1, 2015, from http://www.socialresearchmethods.net/kb/quasird.php

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

Tuckwiller, E., Pullen, P., & Coyne, M. D. (2010). The use of the regression discontinuity design in tiered intervention research: A pilot study exploring vocabulary instruction for at-risk kindergarteners. *Learning Disabilities Research & Practice*, *25*, 137-150. doi: 10.1111/j.1540-5826.2010.00311.x

UNESCO. (2016). UNESCO Institute for Statistics - Glossary Home. (n.d.). Retrieved May 06, 2016, from http://glossary.uis.unesco.org/glossary/en/home.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0).* Retrieved from http://whatworks.ed.gov

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RtI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225–256. doi: 10.1016/j.jsp.2006.11.004

Van Staden, A. (2011). Putting reading first: Positive effects of direct instruction and scaffolding for ESL learners struggling with reading. *Perspectives in Education*, *29*(4), 10-21.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012*). Intensive interventions for students struggling in reading and mathematics: A practice guide.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Webster, R., Blatchford, P., & Russell, A. (2013). Challenging and changing how schools use teaching assistants: findings from the effective deployment of teaching assistants project. *School Leadership & Management*, *33*, 78-96. doi: 10.1080/13632434.2012.724672

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*, 18-37. doi: 10.1037/a0015917

Wilson, C. L., & Sindelar, P. T. (1991). Direct instruction in math word problems: Students with learning disabilities. *Exceptional Children*, *57*, 512-519. doi: 10.1037/a0015917

Ybarra, C. (2012). Exploring the reading achievement of English language learners who participated in the response to intervention model. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (3528262).

What Works Clearinghouse, (2010). *WWC intervention report: Reading Plus®.* Retrieved April 10, 2016 from http://ies.ed.gov/ncee/wwc/InterventionReport.aspx?sid=419

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational & Behavioral Statistics*, *38*, 107-141. doi: 10.3102/1076998611432172

APPENDIX A

Examples of ECRI Word Lists (Second Grade Lessons)

MT 2.1 bag she cat the has pig big cap hat tap can fit pick will who look	MT 2.4 baked with he and see cake game make paper tape crane truck cloth nuts	MT 2.7 have sing play show we flute band tunes pot tube jug blow drum tub bell	MT 2.10
MT 2.2 fox you dog help pup yes egg run ox us fed mess red yum said eat	MT 2.5 yellow down to bugs looked pink bike duck dive ride met hops kite	MT 2.8 are stick from yelled there best before grab lift hand box friend faster next pole gate land slope past bump last	MT 2.11
MT 2.3 black grass plop slips flat sniff frog swim skin spot what do this some	MT 2.6 where den under cave live dome warm hills logs home hive poke animal hole safe snake stone	MT 2.9	MT 2.12

APPENDIX B

ECRI Teaching Material Sample

1._can___________c_
_n___________can_
____
___________an____I_
can_see.___________
____

2._Pam_________P__m
__________Pam
____
_____________am__I_
see_Pam.___________
____

ECRI Mastery Tests Example

Name:

Date:

Mastery Test

can	nap	pat
Pam	ran	Sam
sat	cat	jump
not	and	

Underlined proper nouns **not** included in spelling test.

	Date	Time
Read: 100% in 11 sec.		
Spell: 100%		

APPENDIX C

CAI Treatment Fidelity Observer Sheet

Observer Name: ______________________________ Date: ___________________________

Intervention: I – Ready Reading Plus Grade: ___________________________

Time	Students On Task	Students Off Task	Total Students	Percentage on Task
1: _________				
2: _________				
3: _________				
4: _________				
5: _________				
Averages:				

A total of five observations should be taken over a 10-minute period. The time each walk-around is started should be written in the first column. Students on task are defined as students who:

1) Are actively interacting with the computer program.
2) Are on the appropriate computer program.
3) Are working on the learning parts of the computer program.

ECRI Treatment Fidelity Observer Sheet

PHONICS (ECRI)

D	R	F	P	
				1. You will learn to read a new word by sounding the letters.
				2. Read each sound and hold it as long as my finger is under the letter or letters that go together. Sound. *(Repeat several times)*
				3. Read.
				4. *(Teacher uses word in sentence)*
				5. Read.
				6. Spell and read.
				7. You will provide missing sounds and letters in this word.
				a. This word should be "_________ ."
				b. Say "___________ ."
				c. What sound(s) is (are) missing?
				d. What letter(s) makes (make) that (those) sound(s) in this word?
				e. Spell the sound(s) so I can write it (them).
				f. Read.
				g. Spell and read.
				8. Write, spell and read. *(remove model)*
				9. Proof and correct. *(show model)*
				10. Spell and say. Look at me. *(remove model)*
				11. Sound and read. *(show model)*
				12. Think of a sentence using the word "_________ ."
				13. Tell me/your partner your sentence.
				14. You will read the new word in a sentence. Read this sentence.
				a. *Teacher asks comprehension question(s) about the sentence)*
				15. You will identify which letters and sounds are the same and different in these words.
				a. Read each word as I point to it.
				b. What letters are the <u>same</u> in these words?
				c. What sounds are the <u>same</u> in these words?
				d. What letters are <u>differen</u>t in these words?
				e. What sounds are <u>different</u> in these words?
				f. Spell and read each word as I point to it.
				g. Read each word as I point to it.

D Directive stated correctly
R All pupils respond correctly
F Follow-up if student(s) respond incorrectly or do not respond
P Praise given

Reid, E. R. (1996). *Teaching new words through phonics.* Salt Lake City, UT: Cove.

APPENDIX D

Additional Figures

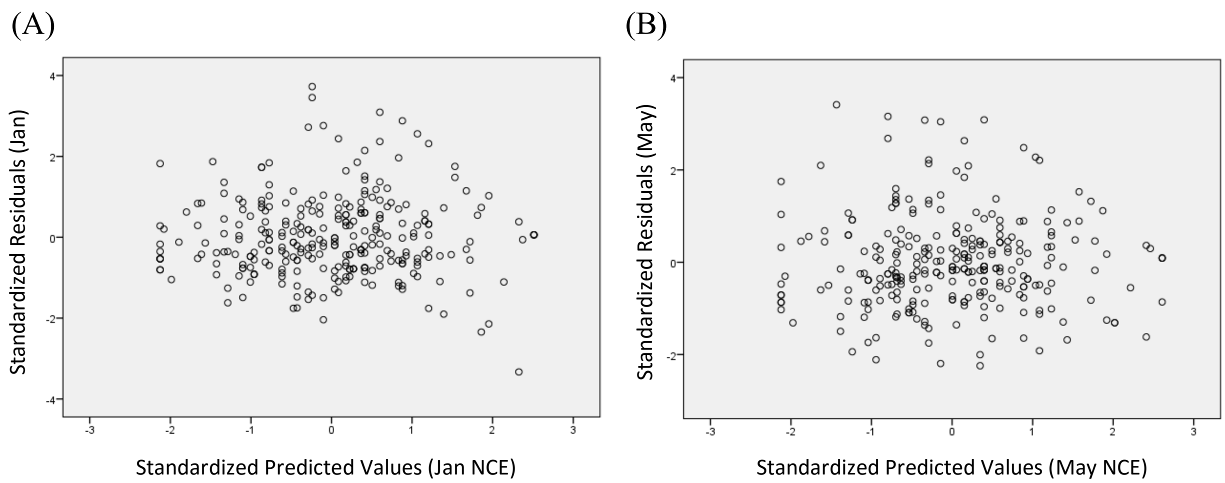


Figure D1. Residual plots for the (A) January and (B) May iterations of the KTEA-II BFR.

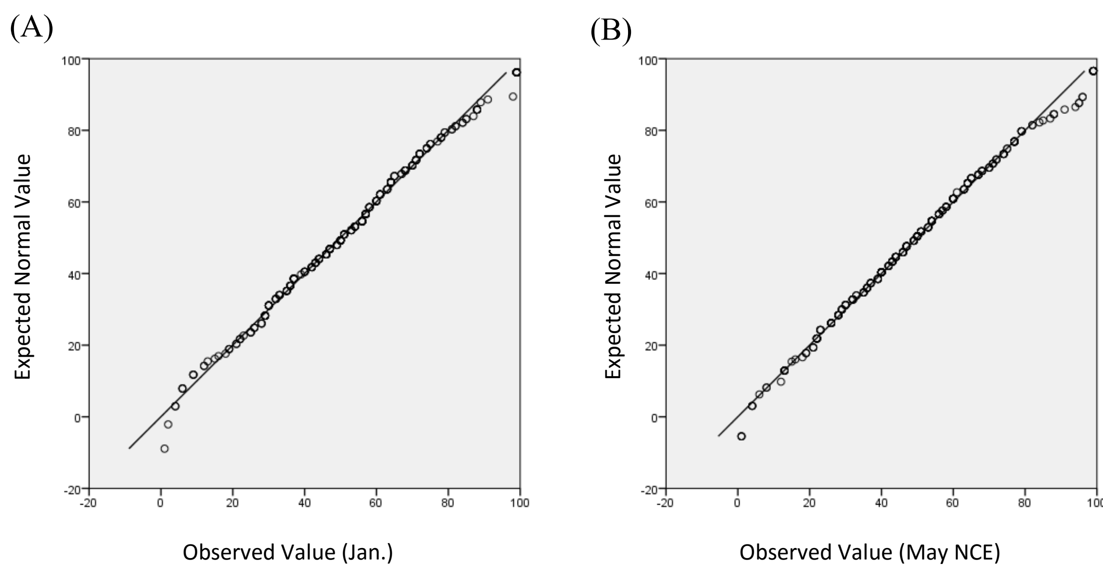
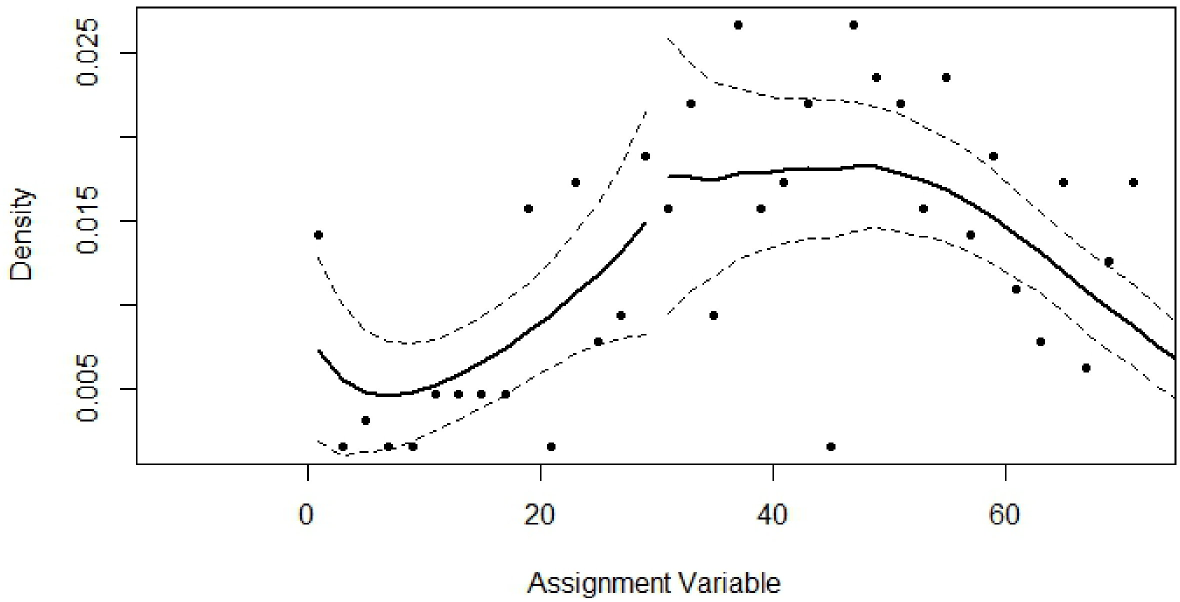


Figure D2. Normal Q-Q plots for the (A) January and (B) May iterations of the KTEA-II BFR.

(A)



(B)

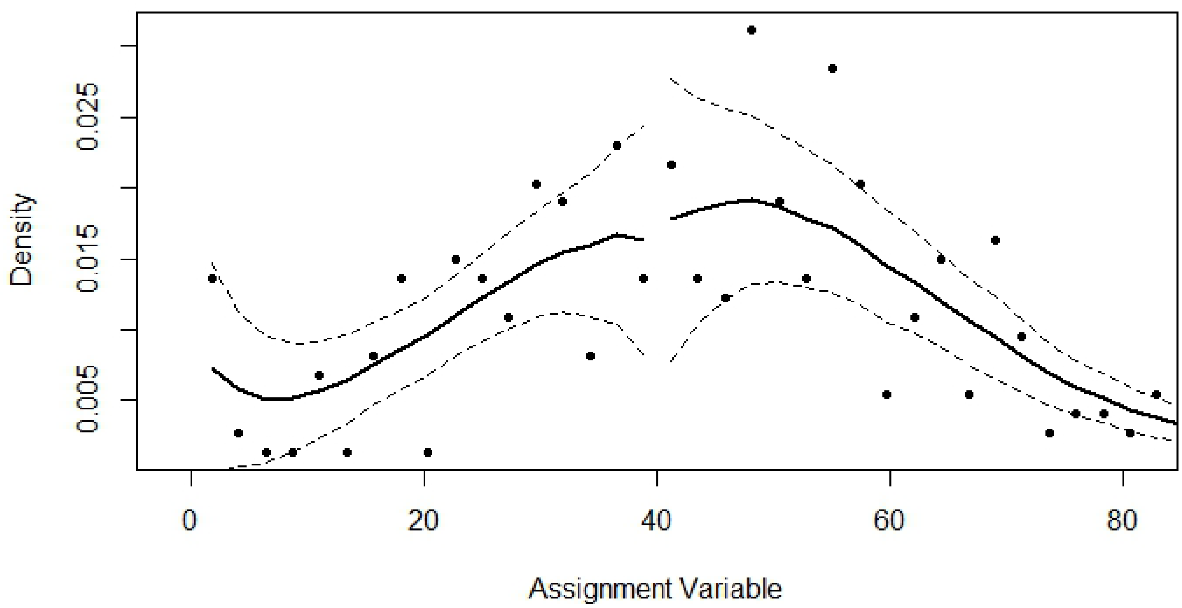


Figure D3. McCrary density plots of the assignment variable showing the (A) lower cutoff and (B) upper cutoff.

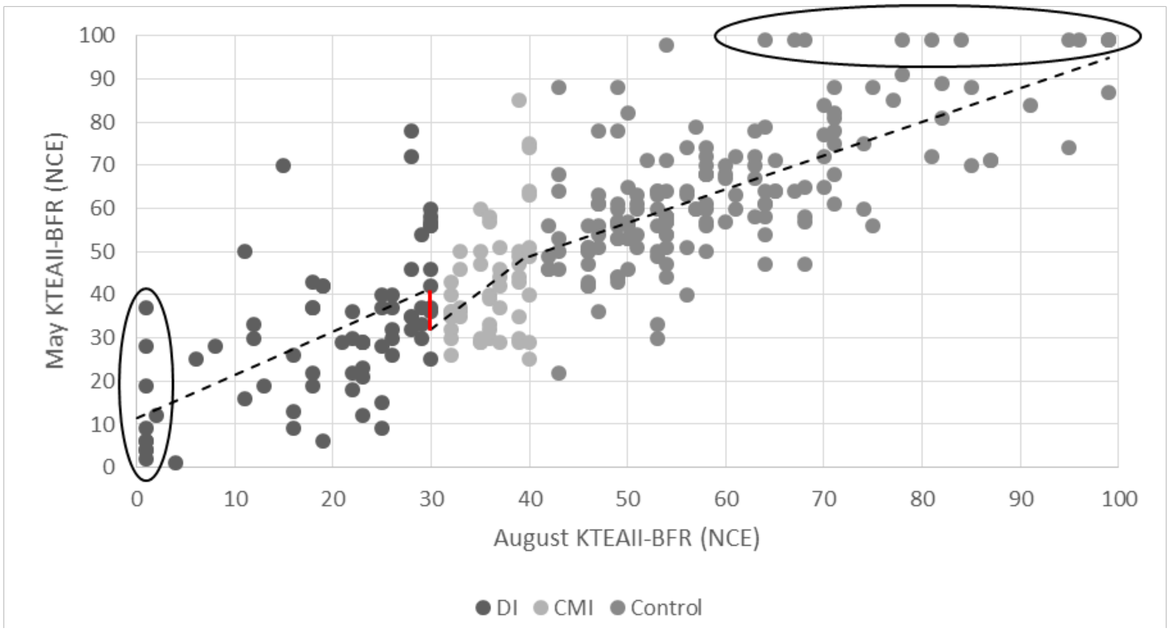


Figure D4. Scatterplot of assignment scores and May posttest scores for the full sample. Ovals indicate ceiling and floor effects, with discontinuity indicated by the red line.

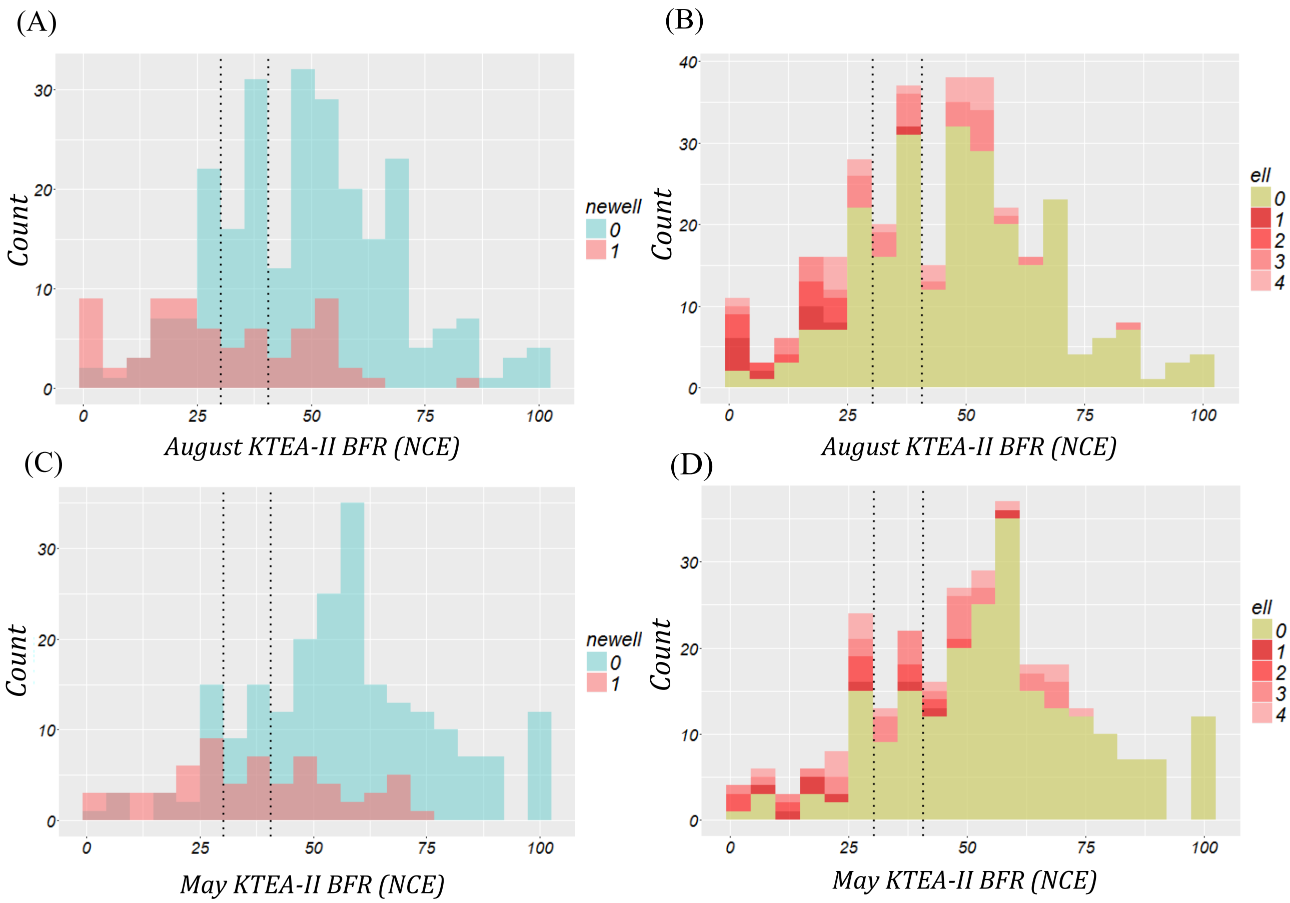


Figure D5. Histograms of ELL students on the assignment variable and on the May posttest. The left-hand histograms represent (A) August and (C) May scores and are dichotomous (0 = non-ELL, 1 = ELL). The right-hand histograms portray (B) August and (D) May scores and include non-ELL students (0) and ELL students classified by level (1-4).

APPENDIX E

Additional Tables

Table E1

Regression-discontinuity Estimates of the Impact of DI Treatment on the May Posttest

		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
16-40	116	11.42*	5.33	.032	-7.35*	-3.68	1.58**	.000	.276
19-40	108	13.30*	5.54	.016	-6.54*	-4.30	1.80**	.000	.241
20-40	103	14.63**	5.08	.004	-6.28*	-5.67	1.96**	.000	.247
21-39	92	15.02**	5.11	.003	-6.28*	-4.10	2.03**	.000	.221
22-38	81	16.60**	5.21	.001	-5.91¯	-4.26	2.33**	.000	.228
23-37	76	17.32**	4.91	.000	-5.92¯	-3.60	2.46**	.000	.206
24-36	61	16.93**	5.11	.001	-7.55*	-1.08	2.68**	.000	.205
25-35	52	18.27**	5.35	.001	-6.74¯	-1.43	3.15**	.001	.227
26-35	47	14.59*	5.34	.006	-5.58	-3.87	2.12¯	.050	.166
27-35	42	9.67	7.05	.170	-5.34	-5.02	0.63	.747	.158
27-34	36	8.09	10.11	.424	-7.93	-4.69	0.16	.961	.203

Note. SPED = special education status; ELL = English Language Learner.

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

Table E2

Lower-grade Regression-discontinuity Estimates of the Impact of DI Treatment

		January Posttest							
		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*P*	*β*	*β*	*β*	*p*	R^2
16-40	57	5.887	4.47	.188	-3.312	-.030	1.49**	.000	.382
19-40	53	10.011*	5.06	.048	-3.416	.042	2.02**	.000	.387
20-40	52	10.575*	5.35	.048	-3.350	-.040	2.09**	.000	.368
21-39	47	10.432*	5.03	.038	-2.692	1.908	2.02**	.000	.339
22-38	44	10.682*	5.16	.038	-2.890	1.903	2.07**	.000	.291
23-37	42	12.722*	5.66	.025	-2.892	2.378	2.40**	.000	.282
24-36	35	10.312^{-}	6.18	.095	-3.200	2.728	1.96*	.018	.143
25-35	28	12.435*	6.11	.042	-2.555	3.745	2.73**	.004	.206
26-35	26	11.135^{-}	6.10	.068	-2.724	2.219	2.49*	.022	.149
27-35	22	-1.308	7.08	.854	.586	-.843	-.78	.647	.024
27-34	18	-3.767	7.90	.633	-1.337	-.097	-1.87	.447	.058

		May Posttest							
		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*P*	*β*	*β*	*β*	*p*	R^2
16-40	57	8.702	5.93	.142	-6.780*	-3.742	1.51**	.001	.286
19-40	53	8.923	6.70	.183	-6.710*	-4.566	1.53*	.018	.213
20-40	52	11.693^{-}	6.66	.079	-6.387*	-4.971	1.90**	.003	.248
21-39	47	13.362*	6.39	.037	-6.290^{-}	-1.654	2.13**	.000	.212
22-38	44	14.904*	6.81	.029	-6.879^{-}	-1.822	2.41**	.001	.204
23-37	42	15.868*	7.22	.028	-6.858^{-}	-0.740	2.55**	.004	.174
24-36	35	15.405^{-}	7.96	.053	-5.992	2.109	2.53*	.037	.114
25-35	28	15.383^{-}	8.44	.068	-3.489	0.238	2.68^{-}	.080	.101
26-35	26	11.135^{-}	6.10	.068	-2.724	2.219	2.49*	.022	.149
27-35	22	-4.958	10.68	.642	-0.908	-7.035	-2.32	.418	.106
27-34	18	-11.298	15.42	.464	-3.702	-5.319	-5.14	.290	.180

Note. SPED = special education status; ELL = English Language Learner.
$^{-}$ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

Table E3

Upper-grade Regression-discontinuity Estimates of the Impact of DI Treatment

		January Posttest							
		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
16-40	59	9.016	8.35	.280	-4.581	-3.417	1.56**	.001	.340
19-40	55	11.533	8.42	.171	-4.070	-3.547	1.87**	.000	.332
20-40	51	10.202	7.55	.177	-4.953	-2.942	1.70**	.002	.250
21-39	45	8.701	7.74	.261	-4.470	-3.266	1.43**	.012	.203
22-38	37	8.152	8.00	.308	-4.790	-3.316	1.31*	.043	.176
23-37	34	9.141	7.06	.196	-4.525	-1.762	1.34*	.060	.180
24-36	26	9.663	6.48	.136	-5.062	-6.089	1.817¯	.077	.282
25-35	24	12.349*	5.85	.035	-5.558	-3.463	2.88**	.000	.483
26-35	21	12.614¯	6.48	.051	-7.663	-2.763	3.07**	.009	.412
27-35	20	8.829	6.66	.185	-8.867	-3.323	1.82	.133	.370
27-34	18	10.426	9.54	.274	-8.724	-3.410	2.39	.336	.358

		May Posttest							
		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
16-40	59	14.505	9.52	.128	-8.217¯	-4.230	1.65**	.001	.293
19-40	55	17.364¯	9.69	.073	-6.616¯	-5.372	1.99**	.001	.287
20-40	51	17.938*	8.46	.034	-6.508	-8.272	2.00**	.000	.282
21-39	45	17.919*	8.77	.041	-6.550	-8.453	2.00**	.001	.273
22-38	37	19.344*	9.02	.032	-5.327	-8.369	2.29**	.000	.311
23-37	34	20.049**	7.73	.009	-5.137	-5.137	2.42**	.002	.305
24-36	26	20.914**	6.17	.001	-7.899	-8.871*	3.26**	.000	.493
25-35	24	22.353**	6.20	.000	-8.152	-7.523	3.83**	.000	.533
26-35	21	21.734**	7.18	.002	-5.356	-8.498¯	3.49*	.025	.544
27-35	20	22.482*	8.89	.011	-5.119	-8.388¯	3.74	.116	.455
27-34	18	34.113**	8.60	.000	-4.073	-9.021¯	7.87**	.000	.607

Note. SPED = special education status; ELL = English Language Learner.

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

Table E4

Regression-discontinuity Estimates of the Impact of DI Treatment on the January Posttest

		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
31-52	121	.021	14.38	.999	-2.461	-3.230	.903	.470	.238
31-51	120	-.438	14.42	.976	-2.528	-3.045	.826	.512	.226
31-50	113	-.086	14.86	.995	-2.473	-2.900	.897	.509	.226
32-49	106	.229	15.16	.988	-2.423	-3.433	.981	.500	.233
33-48	82	-.963	16.65	.954	-2.188	-5.090	.685	.737	.173
34-47	77	-1.071	17.29	.951	-2.316	-5.179	.651	.782	.184
35-46	69	-1.115	17.09	.948	-2.346	-4.890	.486	.849	.176
36-46	63	-1.311	16.96	.938	-3.520	-5.542	.287	.918	.192
37-46	54	-1.131	16.27	.945	-4.052	-4.223	-.291	.924	.204
38-45	36	11.849	16.10	.462	-2.105	-7.000	3.892	.423	.057
39-43	35	16.645	17.28	.397	-1.687	-7.506	4.939	.358	.059

Note. SPED = special education status; ELL = English Language Learner.
¯ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

Table E5

Regression-discontinuity Estimates of the Impact of CAI Treatment on the May Posttest

		Treatment			SPED	ELL	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
31-52	121	-6.389	16.37	.696	-4.983	-4.771	.505	.725	.258
31-51	120	-6.661	16.45	.685	-5.022	-4.662	.460	.752	.251
31-50	113	-6.545	16.91	.699	-5.003	-4.563	.483	.757	.245
32-49	106	-6.791	17.16	.692	-5.052	-5.417	.418	.801	.238
33-48	82	-8.722	18.64	.640	-4.904	-7.304	-.111	.962	.204
34-47	77	-9.579	19.16	.617	-5.936	-8.012	-.379	.943	.220
35-46	69	-9.647	18.90	.610	-5.981	-7.571	-.631	.823	.214
36-46	63	-9.986	18.69	.593	-8.207	-8.735	-.988	.749	.253
37-46	54	-10.016	18.01	.578	-7.024	-11.016	-1.341	.693	.290
38-45	36	8.249	15.67	.599	-5.179	-13.317	4.215	.400	.138
39-43	35	12.518	16.28	.442	-4.537	-14.089¯	5.813	.272	.147

Note. SPED = special education status; ELL = English Language Learner.
¯ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

Table E6

Regression-discontinuity Estimates of the Impact of DI Treatment with ELL Main and Interaction Effects

January Posttest

		Treatment			ELL	ELL x Treatment	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
16-40	116	7.728	4.88	.114	-2.816	3.160	1.63**	.000	.334
19-40	108	10.612*	5.06	.036	-2.386	2.629	1.99**	.000	.330
20-40	103	10.264*	4.90	.036	-2.422	3.112	1.96**	.000	.270
21-39	92	9.271¯	4.86	.057	-0.420	0.513	1.69**	.000	.214
22-38	81	9.469¯	4.89	.053	-.650	0.933	1.75**	.000	.189
23-37	76	10.953*	4.68	.019	-0.476	2.264	2.07**	.000	.194
24-36	61	9.362*	4.61	.042	-1.369	2.572	1.82**	.003	.137
25-35	52	13.345**	4.68	.004	2.987	-1.971	2.85**	.000	.251
26-35	47	13.756**	5.15	.008	2.958	-4.084	2.83**	.001	.171
27-35	42	6.688	6.89	.331	0.131	-2.034	0.86	.541	.049
27-34	36	4.705	8.06	.559	0.330	-2.171	0.11	.958	.065

May Posttest

		Treatment			ELL	ELL x Treatment	Pretest		
Window	*N*	*β*	*SE*	*p*	*β*	*β*	*β*	*p*	R^2
16-40	116	12.690*	6.08	.037	-4.460	1.533	1.70**	.000	.248
19-40	108	14.497*	6.44	.024	-4.223	0.352	1.90**	.000	.219
20-40	103	16.429**	5.98	.006	-4.044	-2.724	2.04**	.000	.227
21-39	92	17.290**	5.97	.004	-0.025	-6.758	2.04**	.000	.203
22-38	81	18.503**	5.99	.002	-0.580	-6.205	2.29**	.000	.213
23-37	76	19.406**	5.67	.001	-0.478	-5.582	2.48**	.000	.188
24-36	61	19.954**	6.10	.001	2.008	-5.495	2.88**	.000	.166
25-35	52	21.421**	6.51	.001	1.910	-5.498	3.44**	.001	.198
26-35	47	18.543**	6.48	.004	0.584	-7.803	2.52*	.023	.154
27-35	42	13.544	8.41	.107	-1.480	-6.996	1.08	.592	.146
27-34	36	12.281	11.26	.275	-1.354	-7.082	0.59	.860	.168

Note. ELL = English Language Learner.

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