Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 16-07

# Transitional Kindergarten vs. Prekindergarten: A Fuzzy Regression Discontinuity Analysis of Student Literacy Skills

AUTHORS

Christopher Doss

Stanford University

ABSTRACT

A growing body of research provides evidence that quality early childhood experiences can affect a host of life outcomes. Equally well documented is the variation in the quality of prekindergarten programs (pre-K) offered to children. In this study I employ a fuzzy regression discontinuity approach to evaluate the efficacy of Transitional Kindergarten (TK) on student outcomes in a large, urban district in California. Importantly, universal prekindergarten was already established in the city which the district serves, making this study a comparison of different prekindergarten opportunities. TK is a highly regulated, state funded, early education program meant to provide a more developmentally appropriate kindergarten curriculum. This study is a test of whether a more highly regulated and academically oriented pre-K program can provide benefits over a more traditional pre-K approach for young five year olds. I find that students who attended TK outperform their peers on a variety of foundational literacy skills. In addition I find some evidence that the gains are larger for minority children.

**Acknowledgements**: I am grateful to Carla Bryant, Pamela Geisler, Meenoo Yashar, Laura Wentworth, Michelle Maghes, Norma Ming, and all other employees of San Francisco Unified School District who provided contextual details and answered all my questions. I am also grateful to Susanna Loeb, Thomas Dee, and Benjamin York for their guidance and support. I thank the participants of the Stanford Center for Education Policy Analysis seminar and the participants of the Association for Education Finance and Policy conference session for their suggestions. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090016 to Stanford University. The opinions expressed are those of the author and do not necessarily represent views of the Institute or the U.S. Department of Education.

VERSION

March 2016

**Suggested citation:** Doss C. (2016). Transitional Kindergarten vs. Prekindergarten: A Fuzzy Regression Discontinuity Analysis of Student Literacy Skills (CEPA Working Paper No.16-07). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp16-07

# *Transitional Kindergarten vs. Prekindergarten: A Fuzzy Regression Discontinuity Analysis of Student Literacy Skills*

**Christopher Doss**[1]

**Stanford Graduate School of Education**

**March 2016**

Abstract: A growing body of research provides evidence that quality early childhood experiences can affect a host of life outcomes. Equally well documented is the variation in the quality of prekindergarten programs (pre-K) offered to children. In this study I employ a fuzzy regression discontinuity approach to evaluate the efficacy of Transitional Kindergarten (TK) on student outcomes in a large, urban district in California. Importantly, universal prekindergarten was already established in the city which the district serves, making this study a comparison of different prekindergarten opportunities. TK is a highly regulated, state funded, early education program meant to provide a more developmentally appropriate kindergarten curriculum. This study is a test of whether a more highly regulated and academically oriented pre-K program can provide benefits over a more traditional pre-K approach for young five year olds. I find that students who attended TK outperform their peers on a variety of foundational literacy skills. In addition I find some evidence that the gains are larger for minority children.

[1] Center for Education Policy Analysis, 520 Galvez Mall, CERAS Building, Stanford, CA 94305 cdoss@stanford.edu. I am grateful to Carla Bryant, Pamela Geisler, Meenoo Yashar, Laura Wentworth, Michelle Maghes, Norma Ming, and all other employees of San Francisco Unified School District who provided contextual details and answered all my questions. I am also grateful to Susanna Loeb, Thomas Dee, and Benjamin York for their guidance and support. I thank the participants of the Stanford Center for Education Policy Analysis seminar and the participants of the Association for Education Finance and Policy conference session for their suggestions. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090016 to Stanford University. The opinions expressed are those of the author and do not necessarily represent views of the Institute or the U.S. Department of Education.

## 1. Introduction

The importance of providing young children with high quality early childhood education has become increasingly clear over the past few decades. Researchers have shown that early childhood education programs can lead to short and medium term academic and socio-emotional gains and potentially improved long term outcomes (Deming, 2009; Currie & Thomas, 1995, 2000; Garces, Thomas & Currie, 2002; Gormely et al, 2005; Huang, 2012; Ludwig & Miller, 2007; Puma et al, 2010; Heckman et al 2010; Belfield et al, 2006; Campbell et al 2012; Anderson, 2012).

These encouraging results spurred states and localities to invest in prekindergarten (pre-K) programs. California took a step in this direction when Governor Schwarzenegger signed the 2010 Kindergarten Readiness Act into law. Previous to this law, all children who turned five on or before December 2 were eligible for kindergarten. Educational stakeholders in the state were concerned that the youngest of these children were not developmentally ready for the academic demands of kindergarten (Governor's State Advisory Council, 2013). Beginning in the 2012-2013 school year, the law gradually moved the cutoff date to September 2 and established a Transitional Kindergarten (TK) year for students who turn five between September 2 and December 2. Though districts were given leeway on how to implement the program, the state tasked districts to provide a more developmentally appropriate kindergarten curriculum for children of this age (Governor's Advisory Council, 2013). In doing so, the state effectively created a pre-K program for children turning five within that time frame. TK distinguishes itself from other pre-K programs in that it is funded and governed in the same manner as the K-12 system. It is more highly regulated than typical prekindergarten programs and is completely free to families. The program was projected to cost the state $675 million a year when fully implemented (Legislative Analyst Office, 2012), though a recent expansion will likely increase that amount.

In this study I leverage a fuzzy regression discontinuity (FRD) design to causally evaluate the efficacy of TK in raising student literacy skills in the San Francisco Unified School District (SFUSD).

The district provides an opportunity to compare more and less regulated preschool programs. Residents in San Francisco previously voted to establish universal pre-K. In the 2012-2013 the program served 3,225 students (Controller Office, 2013) with a budget of $19.7 million. A child turning five years old on December 2 can enroll in TK (but may enroll in pre-K), while a child turning five years old on December 3 only has the option to enroll in a pre-K program offered in the city. The next year both sets of children enter kindergarten at the same time. Figure 1 illustrates this assignment for the second cohort of students. True to the spirit of the law, the district designed their TK program to be an academic and socio-emotional middle ground between pre-K and kindergarten.

This study adds to the literature documenting the causal effects of early childhood education programs on short-term student academic outcomes. It compares the outcomes of a more highly regulated and academically focused form of pre-K to a more traditional pre-K market. This is a direct test of the developmentally appropriate nature of TK that underpins the law and the district's efforts. There is a debate as to what developmentally appropriate means in the early childhood context. Many stakeholders are pushing back at what they see as an increasing focus on academics (Bassok & Rorem, 2014; Bassok & Latham, 2014; Stipek, 2006; Elkind, & Whitehurst, 2001; Zigler & Bishop, 2006). This study contributes to that debate by examining whether a greater emphasis on academics can produce higher academic outcomes in this age group.

I analyze 6,773 kindergarteners enrolled in SFUSD in the 2013-2014 and 2014-2015 school years. These classes contain the first two cohorts of students that experienced TK in the district. Of the students in the sample, 950 were eligible for TK in the previous year and 336 enrolled in the program. The primary outcomes are the fall kindergarten and fall first grade administrations of the Fountas and Pinnell Benchmark Assessment System (BAS) and the California English Language Development Test (CELDT). The BAS measures pre-literacy skills and the ability to read books of increasing difficulty. The CELDT is given to all Limited English Proficient (LEP) students and measures reading, listening, speaking, and writing. I find that, in the fall of kindergarten, students who experienced TK outperform

their peers on both assessments. Some of this advantage may come from the fact that TK students have been previously exposed to the assessments. Fall first grade results show that the advantages in CELDT remain, though this does not translate to the ability to read more advanced books on the BAS.

## 2. Literature Review and the District Context

### *2.1 Prior Early Education Literature*

Researchers have put considerable effort in estimating the effects of specific early childhood interventions. The Perry-Preschool experiment, the Abecedarian study, and studies on the efficacy of Head Start are among the most widely cited prekindergarten studies. The Perry-Preschool and Abecedarian programs are examples of intensive programs that have been found to have large, short to medium term effects on IQ, reading, and math scores (Campbell et al, 2012; Heckman et al, 2010; Barnett, 2011). Head Start is a quintessential example of a large, federally funded program meant to provide services to a large swath of economically disadvantaged children. Though less intensive than the Perry-Preschool and Abecedarian programs, the program has positive effects on language, literacy, and math outcomes (Deming, 2009; Currie & Thomas, 1995; Puma et al, 2010). A common theme, however, is that test score gains tend to fade with time.

The establishment of TK fits into a larger trend of state and localities investing in their own pre-K programs as a response to this encouraging evidence. Causal evaluations of state and local programs in Oklahoma (Gormely et al, 2005) and five other states (Wong et al, 2008), as well as a descriptive evaluation of Georgia's pre-K program (Huang, 2012) have shown short-term benefits in academic outcomes. Wong et al (2008), however, point out that there is still considerable variation in the effectiveness programs, making continued causal evaluations of pre-K programs important. This study provides more evidence on the effectiveness of a large, state program.

In addition, TK has the potential to combat the sorting of families to prekindergarten programs along socioeconomic and demographics lines. Researchers have documented that the socioeconomic and demographic inequalities in K-12 education are reproduced in the early education sector. Economically

disadvantaged and minority families are either less likely to opt into formal early childhood programs, or enroll in less effective programs (Magnuson et al, 2004; Phillips & Lowenstein, 2011; Capizzano, 2006). These sorting patterns are also related to academic outcomes (Bassok et al, *forthcoming*; Lee et al, 1998; Loeb et al, 2004). Layered on to this sorting issue is the fact that the early childhood labor market is marked with a dramatic variation in the stability, education, and compensation of the teachers, even in the formal early childhood education sector (Bassok et al, 2013).

There is some evidence that addressing these factors can be beneficial for children. Rigby et al (2007) showed that subsidies are associated with an increase in the quality of care provided to children and have mitigated some of the selection effects by increasing the uptake of center care. Meanwhile, pre-K programs in markets that more highly regulate the early childhood services and its labor market are associated with better outcomes (Hotz & Xiao, 2011; Rigby et al 2007; Bassok et al, *forthcoming;* Fuller et al, 2004). This paper can contribute to this line of literature due to the free nature of TK and the strict regulation of its labor force. This evaluation is a test of whether these regulations are associated with increased student outcomes. If low-income and minority children attend prekindergarten programs of lower quality, combatting these selection effects can result in greater outcomes for minority children.

A final feature of TK finds itself relevant to a current debate in the literature as to what a developmentally appropriate curriculum looks like for young children. Recent studies have shown that kindergarten is becoming increasingly focused on building more academic behavior in reading and math, with the phenomenon more pronounced in schools serving low-income and minority children (Bassok & Rorem, 2014; Bassok & Latham, 2014). The phenomenon, partly spurred by the accountability movement, has begun to reach preschool (Stipek, 2006). This trend has caused parents, researchers, and practitioners to debate whether we are asking too much of children too soon (Elkind, & Whitehurst, 2001; Zigler & Bishop, 2006; Stipek, 2006; Hatch, 2002). On one hand, the link between pre-K academic performance and later outcomes has encouraged the greater emphasis on academics. On the other hand, many professionals worry that this academic growth will come at the expense of socio-emotional skills

and mental health. This study cannot speak to the socio-emotional effects of a more academic pre-K, but can speak to whether it produces stronger academic outcomes in this younger age group.

The American Institutes of Research (AIR) concurrently evaluated TK programs throughout the California (Manship, K. et al, 2015). In their report they find that TK had positive effects on literacy, mathematics, and the self-regulation of children. This study is able to build on their findings in a few ways. First, it provides a more detailed account of the efforts of one large, urban district. Because the Manship et. al. study covers a larger geographic area, it cannot clearly identify the alternatives to TK that those not receiving the program attend. With the district-specific data, I can describe the district and city context more clearly, thereby elucidating the contrast between pre-K and TK. Second, I am able to give a more complete picture of the benefit children receive in literacy skills. Manship et al measured letter and word recognition, phonologic awareness, and expressive vocabulary. I am able to report effects on similar measures, as well as the effects on reading books of increasing difficulty. Additionally, for students classified as Limited English Proficient students (LEP) I am able to look at the effect of TK on the listening, speaking, and writing abilities of students. Finally, by separating out the effects by subgroup I am able to test if minority students especially benefit from the program as predicted by the mitigation of selection effects. While the Manship study could look at differences by measured characteristics of children, these characteristics could be confounded with alternative pre-K opportunities and other geographic characteristics. To the extent that different subgroups have sorted into different districts and areas, each subgroup result will rely on different TK programs and TK/pre-K contrasts. This makes subgroup comparison more difficult to interpret. I have population data for the district, making it easier to identify differences across students.

*2.2 Prekindergarten vs. Transitional Kindergarten, The District Context*

San Francisco has a voter-approved universal pre-K market that served about 83 percent of the city's four year olds in 2011-2012 (EED, 2012). The city funds an umbrella organization which

establishes minimum criteria that all participating pre-K programs must meet including services provided, teacher qualifications, and tuition. The pre-K market, thus, is regulated to an extent that is not typical in the country. There is evidence that San Francisco's universal pre-K has been successful. In 2013, Applied Survey Research leveraged a regression discontinuity approach to evaluate the umbrella organization's programs. In this study they compared children who attended pre-K to those who did not. They found that the program produced a three-month gain in letter and word recognition, a three- to four-month gain in applied problem solving, large gains in self-regulation, and suggestive evidence that it benefited Spanish vocabulary. Overall, classrooms were rated mid-to high on the Classroom Assessment Scoring System (CLASS) assessment, with emotional support being a strength and instructional support a challenge (Applied Survey Research, 2013a, 2013b). Representatives of the organization stated that two thirds of the city's four year olds experience high quality preschool programs (Edsource, 2015).

This type of regulation is likely to provide a floor with regard to the quality of services provided to children in the city. Even in this regime there is variation in quality and the opportunity for sorting of children in higher or lower quality settings. City providers must be licensed by the state; however, providers range from school-based programs to Head Start programs, to home-based care. The teachers they employ must have a minimum of 24 early childhood or child development credits and 16 general education credits, but certainly providers can employ more highly educated teachers. Additionally, there is no minimum compensation for teachers required of providers. Therefore centers can attract teachers of varying quality, partially through compensation. The variation in the market opens the door to the sorting of families to centers. In addition, the city provides funding for 612.5 hours of instruction spread through 175 to 245 days. This amounts to 3.5 to 2.5 hour school days. The organization does not subsidize more time, making it plausible that disadvantaged families select fewer hours of instruction.

The highly regulated nature of TK can mitigate many of these lingering selection effects. First, TK is strictly a school-based early childhood program and completely eliminates the variation in types

of programs offered to families. In addition, the state requires teachers to hold a bachelor's degree and the same credentials as other elementary school teachers. This raises the floor of, and reduces the variation in, provider qualifications and education. The district also compensates TK teachers at the same rate as other teachers, effectively raising the floor of, and reducing the variation in, teacher compensation. Finally, TK is open to all residents of the city and is completely free. In SFUSD, all TK-eligible families have the opportunity to enroll in a full day early childhood program at no cost. Some variation certainly remains. There is likely to be variation in quality of TK classrooms throughout the city and selection to these classrooms is likely to be correlated to demographic and economic variables. On the balance, these selection effects are likely muted in comparison to those in the larger pre-K market.

TK further distinguishes itself from pre-K in regards to the structure of the day and the focus of the curriculum. The city offers no set pre-K curriculum, but all providers must align their curriculums to the California Preschool Curriculum Frameworks. Perhaps the best way of illustrating the contrast in programs is to distinguish the key differences between SFUSD's pre-K program, which is part of San Francisco's universal pre-K system, and SFUSD's TK program.

Figure 2 compares the key elements of the SFUSD's TK and pre-K programs. The district structures the TK day to mirror that of kindergarten. In pre-K, children start the school day at different times and parents select the number of hours of instruction. In TK all children start the day at the same time and attend for a full six hours. While breakfast is provided for morning pre-K students, breakfast is not provided for TK students, though some centers provide a morning snack. Prekindergarten students are also given nap time, whereas no naps are provided for TK students. Finally, prekindergarten students have up to 1 hour of outdoor time, while TK students have 15 to 20 minutes of outdoor time.

Second, the district is currently using a home-grown TK curriculum designed to be the middle ground between their pre-K and kindergarten curriculums. In creating the TK curriculum, district officials placed a large emphasis on foundational literacy skills and socio-emotional skills and a growing

emphasis on foundational math skills. Starting with the second cohort, SFUSD introduced visual arts. TK differs from the SFUSD's pre-K program in that it is less child-driven and more structured. In prekindergarten, student skills are allowed to guide the activities and instruction, whereas in TK activities are guided by the teachers. For example, teachers are expected to stay on a curriculum map and timeline in TK, whereas no such map or timeline exists in pre-K. In both TK and pre-K, each period of whole group instruction lasts no more than 10 minutes, but TK utilizes whole group instruction more often.

Finally, TK also differs from pre-K in the number of adults in the classroom. Qualified pre-K programs must have a maximum class size of 24 and a child-adult ratio of 8:1. In contrast the district TK is a modified kindergarten classroom with a maximum class size of 22 children, but only one paraprofessional is available for the first six weeks of class. This makes the overall child-adult ratio significantly larger in TK, though still less than that of kindergarten where there are no paraprofessionals and the maximum class size is still 22 students.

## 3. Data

This study examines the first two cohorts of Transitional Kindergarten students in SFUSD to estimate the effect of the program when compared to the other pre-K options available locally. The TK program was phased in over three years. In the first year children were eligible for TK if they turned five years old between November 2 and December 2. In the second year, children turning five years old between October 2 and December 2 were eligible for the program. Children born after December 2 were eligible for only the city's pre-K services; children born before November 2 (or October 2 in the second year) enrolled in kindergarten and are therefore not in the study.[2] The structure of the program means that a plausibly exogenous cut point, based solely on birthdate, dictates potentially very different educational experiences for children. That is, children born a few days on either side of the cutoff should,

[2]I can also compare students born on November 1 (October 1 in the second year), and therefore in kindergarten, to students born on November 2 (October 2) and therefore in TK. However, these children are not in the same classroom at the same time and will not have the same assessments administered concurrently.

on average, be similar except for the probability of enrolling in TK. A fuzzy regression discontinuity can leverage this cut point to analyze the effect of TK on outcomes. The discontinuity is fuzzy because TK-eligible children do not have to attend the program.

SFUSD provided administrative data on the universe of kindergarten students for the 2013-2014 and 2014-2015 school years. The administrative data included student background characteristics, detailed in Table 2, as well as each student's birthdate. I match kindergarten administrative data to the previous year's TK rosters to identify students who enrolled in TK. I repeated the process with pre-K rosters to identify students who attended the district's pre-K program.

The district uses the Fountas and Pinnell Benchmark Assessment System (BAS) to measure literacy skills of every student in TK to third grade. The BAS is meant to be a formative assessment tool and has been shown to be a valid assessment of literacy development in children (Fountas and Pinnell, 2012). In the fall, all teachers are required to assess their children on the foundational skills. In the 2013-2014 year these skills were: upper and lower case letter recognition, letter sounds, initial word sounds, early literacy behaviors, rhyming, blending, 25 high frequency words, 50 high frequency words, and segmenting. If students mastered eight of these ten skills they then began to read books of increasing difficulty. Students started with the easiest books (level A) and after reading with sufficient accuracy and comprehension they progressed to harder levels (levels B to Z).

In the 2014-2015 year the district incorporated teacher feedback and made segmenting and the 50 high frequency word skills optional. To advance to the leveled books, students needed to master six of the remaining eight foundational skills. For consistency, the fall kindergarten BAS outcomes in this paper are the eight foundational skills common to both years, the probability of mastering the requisite number of foundational skills to move on to the leveled reading assessment, and the probability of reading at least at level A. The test could also be administered in either English or Spanish. The scales for upper and lower case letter recognition, letter sounds, and early literacy behavior are slightly different

between the two test versions. My preferred specification therefore includes controls for year and test language. By first grade almost all children (98 percent) have been assessed on their ability to read books. The fall first grade results are whether TK students are reading more advanced books.

Both because many students in the district are English learners and because the assessments for English learners are on a continuous scale and thus easier to analyze, I assess the effects of TK on the performance of LEP students on the California English Language Development Test (CELDT). This district automatically identifies a student as LEP if the family indicates they speak a language other than English in the home. Any student in the state of California who is identified as LEP is required to take the CELDT the first year they enter the district and every year until they are reclassified as English proficient. Students are assessed in listening, speaking, reading, and writing. In the listening portion of the exam students are tested on their ability to follow directions, comprehend descriptions of situations and stories communicated orally, and identify rhyming words orally. In the speaking section students are tested on their oral vocabulary, their speech functions, their ability to orally construct a story from pictures, and their ability to communicate reasoning skills. The reading portion of the exam tests many of the same skills as the BAS including a child's ability to identify letter sounds, identify pictures associated with written words, and identify parts of a book. In the writing portion of the exam students are asked to copy letters and words, write words based on pictures, and recognize punctuation and capitalization.

The CELDT compliments the BAS in a few ways. Whereas the BAS is administered by teachers, the CELDT is administered by trained outside assessors. This mitigates any concern that the teachers expect differences in performance from former TK students and grade accordingly. In addition the CELDT subtest and overall performance is expressed in traditional scale scores, which lends itself to more traditional interpretation of the estimates. Finally, to the extent both assessments test many of the same skills, similar results reinforce our confidence in the estimates.

One caveat to the kindergarten results is that that TK students were exposed to the CELDT and BAS in their TK year (the year prior to K) while students in pre-K were not. The district uses the BAS as a formative assessment tool in TK and the state requires that all entering LEP TK students are assessed on the CELDT. The fall kindergarten results therefore contain any true learning in TK as well as any practice effects of having taken the test before. In the fall of first grade all students have been exposed to the assessments, thereby eliminating any practice effects.

Across the two years 8,717 kindergarten students matched to the fall kindergarten administrations of the BAS. Teachers varied considerably in the extent to which they followed district assessment guidelines in administering the BAS. Many students were missing individual foundational skills scores, and many teachers assessed the child's reading level if they were close to mastering the required number of foundational skills. The final analytical sample therefore consists of 6,773 out of the original 8,717 students. These students had scores for all foundational literacy skills except rhyming and blending. The missing data was largest for those two domains and the sample sizes are smaller. If the missing data is not the same for students born on either side of the birthday threshold, comparisons of outcomes between these two groups may include bias from the missing data. Table A1[3] shows reduced form regression discontinuity results run on the analytical sample of students with an indicator if a student was missing rhyming or blending scores. Results show that missing scores are not significantly related to the birthday threshold, making bias in the results unlikely.[4]

Of the 6,773 students in the analytical sample, 3,334 are LEP and were tested in the CELDT in the fall of kindergarten, 6,219 continued to first grade and were assessed in the fall with the BAS, and 2,697 LEP students progressed to first grade and were assessed with the CELDT. Again the results for

[3] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://www3.interscience.wiley.com/cgibin/jhome/34787.
[4] Furthermore, results are robust to including all students in the sample.

the LEP and first grade samples would be biased if the probability of being in those samples is discontinuous across the threshold. Table A1[3] indicates that this is not the case.[5]

Table 1 presents the descriptive statistics for the analytic sample, former TK, and former pre-K students in the sample. The students are mostly Asian (31.1 percent) and Hispanic (25.1 percent), with fewer whites (16.4 percent). African Americans (6.3 percent) make up a small part of the sample and are contained in the Other category (17.5 percent). Special education students compose 7.8 percent of the sample, while 49.4 percent has been classified as LEP. The alternative pre-K experience of students who did not enroll in SFUSD's TK program is not fully known. However, 16.9 percent attended the pre-K in SFUSD and 5 percent attended TK. In total, 22 percent was enrolled in the district in the prior year.

Compared to the former pre-K students, former TK students differ in some important ways. Due to the eligibility criteria, they are mechanically older. TK students were also significantly more likely to be minority and LEP and less likely to be special education. Overall TK students, on average, significantly outperformed pre-K students in all administrations of the assessments.

## 4 Empirical Strategy

### *4.1 Identification Strategy*

The differences in age and background characteristics between former TK students and their kindergarten peers make clear the need for quasi-experimental techniques such as a FRD approach. The fall kindergarten results provide a measure of the effectiveness of TK in relation to the other pre-K experiences of children in the sample, though it may also measure test-taking practice effects because TK students took the test while in TK. Test performance in the fall of first grade is less likely to reflect practice effects because all students took the tests in kindergarten.

One challenge in working with the BAS foundational skills is the left skewed nature of the distribution. In the fall assessment 6.5 percent to 48.5 percent of the sample achieved the highest score

[5] Very few students are reclassified as English proficient before second grade. In the analytical sample only 1 student who was designated LEP in kindergarten was reclassified in first grade.

on the foundational skills. The non-normal distribution of the outcomes make OLS inappropriate. I therefore backwards code each skill so that I have a count of how many items a student *missed*, and treat each variable as a count variable. This approach allows me to use a family of parametric regressions based on the poisson distribution that include poisson regression, negative binomial regression, and their zero-inflated versions. I present estimates from the negative binomial model.[6]

When analyzing the ability of students to read books of increasing difficulty, I use ordinal logit models due to the ordinal nature of the book levels. In addition I present linear probability models of the probability of reading at level C, level E, and level I or above. I choose these levels because they represent approximately the 20th, 50th, and 80th percentiles of the analytical sample's distribution in the fall of first grade. This strategy allows me to present an overall measure of a group's ability to read books of increasing difficulty, as well as probe certain points in the distribution for effects.

Equations (1) and (2) model my fuzzy regression discontinuity approach:

$$TK_{ict} = \beta_0 + \beta_1 \mathbf{1}\{B_{ict} \geq 0\} + \beta_2 f(B_{ict}) + \boldsymbol{X_{ict}}\beta_3 + \delta_{at} + \varepsilon_{ict} \quad (1)$$

$$Y_{ict} = \gamma_0 + \gamma_1 \mathbf{1}\{B_{ict} \geq 0\} + \gamma_2 f(B_{ict}) + \boldsymbol{X_{ict}}\gamma_3 + \delta_{at} + \epsilon_{ict} \quad (2)$$

Equation (1) regresses $TK_{ict}$, an indicator for whether student, *i*, in classroom, *c*, in year, *t*, enrolled in TK in the previous year, on the following: an indicator for TK eligibility in the previous year, a flexible polynomial, *f*, of the rating birthday rating variable, $B_{ict}$, a vector of student characteristics, $\boldsymbol{X_{ict}}$, and assessor-by-year fixed effects, $\delta_{at}$[7]. The rating variable, $B_{ict}$, is the distance, in days, a child is born from December 2. Following Lee and Lemieux's (2008) recommendation, I cluster standard errors on the

[6] In choosing from among the models I follow Cameron and Trivedi (2010) and Long and Freese (2014) and compare the Akaike Information Criterion (AIC), the Bayesian Information Criteria (BIC) and the Vuong statistic (1989) via Stata's - countfit- command. In all cases the negative binomial model was preferred to poisson model and the zero inflated negative binomial model was preferred to negative binomial model. I choose the negative binomial model because it is more easily interpretable. All inferences are consistent when using the zero-inflated negative binomial model.

[7] If the TK program causes sorting of students to classrooms or schools the assessor-by-year fixed effects may not be appropriate. If TK students sort to higher (lower) performing classrooms or schools the estimates may be biased downward (upward). My preferred estimates still include the fixed effects to account for any stable differences between assessors. To be inclusive, Table 4 presents my main results with and without covariates and fixed effects.

birthday rating variable because the exact birthday may be considered a coarse rating variable. The coefficient of interest is $\beta_1$, which represents the compliance rate with the TK eligibility requirements.

Equation (2) presents reduced form intent-to-treat (ITT) estimates of the effect of being eligible for TK on student outcomes. $Y_{ict}$ is now the literacy outcomes of the child. $\gamma_1$ in equation (2) is the coefficient of interest and represents the ITT estimate of being TK-eligible on student literacy outcomes. In both equations the vector $\boldsymbol{X_{ict}}$ includes all student characteristic variables in Table 2 and an indicator for kindergarten year. For the BAS outcome, the assessor-by-year fixed effect would account for stable differences among teachers in how they assess their students in a given year. I cannot identify the CELDT assessors, but one to three assessors were assigned to a school depending on the size of the school. $\delta_{at}$ in these cases are the school-by-year fixed effects. Once again standard errors are clustered on the birthday rating variable.[8] Finally, I leverage Akaike's Information Criterion (AIC) to determine the optimal functional form of $f$ (Schochet et al, 2010). The results indicate that a linear spline is optimal in all cases. As a robustness check I present results from many bandwidths, including local linear regressions.

*4.2 Manipulation of the Threshold*

A key identifying assumption is that the potential outcomes, $Y_{ict,}$ are independent of the treatment assignment, conditional on the forcing variable, $B_{ict}$. That is, the cut point of December 2 threshold is plausibly exogenous such that, in the limit, those students on either side of the threshold are, on average, similar. Any attempt to sort children to either side of the threshold would undermine this identification strategy. The first two cohorts of TK students were born in 2007 and 2008, two to three years before Governor Schwarzenegger signed the law. Parents were unable to make family planning decisions based on the law. It is possible that the TK program affects enrollment into kindergarten in systematically different ways around the birthday threshold. Figures 3(a) and (b) present visual depictions of the distribution of observations around the threshold. Figure 3(a) shows that there could be a significant drop

[8] In the conditional negative binomial and ordinal logit models standard errors must be clustered on the assessor-by-year fixed effect.

in observations in crossing the threshold, though there are similar natural fluctuations throughout the range of the rating variable. I follow McCrary (2008) and statistically test whether there is a discontinuous jump in the density of observations surrounding the threshold. Figure 3(b) presents the graphical results of the test and makes clear that I cannot reject the null hypothesis that there is no change in density around the threshold. The point estimate and standard error of the density discontinuity is 0.110 (0.087).

These natural fluctuations, however, are indicative of regular heaping often found in birthday rating variables. Recent work by Barreca et al (2015) shows that heaping can cause bias in RD point estimates if observations in the heaps are systematically different from observations in the non-heaped data. To test for bias they recommend estimating the effects on heaped and non-heaped data separately. As shown in the histogram in Figure 3(a), 15 to 32 students are concentrated on some values of the rating variable, thereby creating heaps. In Section 7 I test for bias by eliminating observations in values of the rating variable that contain 15 or more students. My results are robust to eliminating these heaps.

The regression discontinuity technique additionally assumes that nothing that affects the outcomes, except for the probability of enrolling in TK, is discontinuous across the threshold. I partially test this assumption by running RD regressions to see if the covariates are discontinuous around the threshold. Table 2 presents these results for the full sample and with a bandwidth restriction of 60 days and 30 days on either side of the cutoff. The covariates tested are well balanced across the threshold. No covariate is consistently unbalanced across all the bandwidths tested.

Finally, to be a valid FRD the December 2 threshold must predict a strong treatment contrast. Figure 4 presents the first stage results graphically. Virtually nobody who was TK-ineligible enrolled in TK. Only 1 child, who was born on December 3, managed to enroll into the program in the two years of the study. For those children born before December 2, the probability of enrollment increases considerably. Table 3 presents statistical estimates of this compliance rate for the full sample, and for

the sample that lies between 60 and 30 days on either side of the birthday threshold. I find a robust compliance rate of about 30 to 33 percent across models.

## 5. Main Results

Students who have previously experienced TK outperformed their peers on the foundational literacy skills in kindergarten. Figure 5 graphically presents the main fall kindergarten literacy results. This approach has the advantage of allowing the data to present the results free of any statistical manipulation. After aggregating all foundational literacy skills together, the number of items missed drops as one crosses the December 2 threshold. Figure 5(a) indicates that TK-eligible students missed about 8 items less than their peers, which represents about a 14 percent decrease in the number of items missed from a base of about 56 items missed for TK-ineligible students at the threshold. For the individual skills, similar drops in items missed are present for upper case letters, lower case letters, letter sounds, high frequency words, early literacy behaviors, and rhyming. Figure A1 in the appendix[9] illustrates these results. There is also a jump in the probability of mastering enough foundational skills to be assessed in reading, and the probability of reading at level A or above. For LEP students, Figure 5(d) shows a jump in the overall performance of students on the CELDT. Figure A2[9] shows similar jumps for the listening, reading, and writing subtests of the CELDT.

The picture changes somewhat by the fall of first grade. Figure 6 shows the advantage seen in foundational skills does not translate to the ability to read more advanced books. There are small jumps in the probability of reading at levels C, E, and I or above, but they are insignificant. However, the advantages in the CELDT remain and TK students classified as LEP still outperform their peers.

[9] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://www3.interscience.wiley.com/cgibin/jhome/34787.

Table 4 presents the results from the statistical models. I report the coefficients for the unconditional FRD results, as well as results from my preferred specification that includes a linear spline of the rating variable, covariates, and assessor-by-year fixed effects. Though this specification relies heavily on the validity of the linear functional form, I show in Section 7 that results are robust to a variety of bandwidths.[10] Columns 1 and 2 of panel A of Table 4 show that the intent to treat estimates are significant ($p<0.05$) for eight of the eleven kindergarten BAS outcomes. TK students benefited on all foundational literacy skills. TK students, however, were not more likely to master the requisite number of foundational skills to move on to the reading assessment, nor were they more likely to have been reading at level A or beyond.

The coefficients on the negative binomial models are difficult to interpret. Table 5 therefore presents incidence rate ratios versions of the coefficients in column 1 of Table 5. These estimates are obtained by: $e^{\gamma_1}$. Incidence rate ratios in this context will indicate the rate at which TK-eligible students, on average, miss an outcome compared to TK-ineligible students. Table 5 indicates that TK-eligible students were less likely to miss foundational skills by factors of about 0.91 to 0.71. This translates to a 9 percent to 29 percent decrease in the number of items missed respectively. To make these results more meaningful I calculate the average number of items missed by students in the control group born within 30 days of the threshold. I multiply the percent decrease in missed items by the control group mean. On average TK students missed nine fewer items, knew two more upper case letters and letter sounds, and knew one more lower case letter. They could also recognize 2.3 more words out of 25. Of the remaining skills, measured on a one to ten scale, TK students performed better by half a point. With about a 33 percent compliance rate, the treatment-on-the treated estimates will be roughly three times as big.

[10] In an effort to find the optimal bandwidth I also implement the procedure recommended by Imbens and Kalyanaraman (2011). For most literacy outcomes, the procedure recommended bandwidth of about 2-11 days. This highly localized bandwidth only encompasses 2.1 to 7.4 percent of the data. Instead of using this restrictive slice of data I present the results using all observations and show robustness to a variety of bandwidth restrictions.

Turning our attention to the performance of LEP students in kindergarten, column 4 of panel A in Table 4 indicates that overall students performed 0.183 standard deviations (SD) better on the CELDT exam ($p<0.05$). All subtests except speaking were also significantly better and estimates range from 0.186 SD to 0.221 SD. Overall the CELDT results corroborate the BAS results and indicate that TK students outperformed their peers on most literacy outcomes.

Because TK students entered the district a year earlier, they were exposed to the tests and, thus, some of the gains could be from having practice. The first grade CELDT outcomes seen in Column 4 of Panel B in Table 4 indicate that such a practice effect is likely not biasing the fall effects. At this point all LEP students have been assessed at least once, but the results remain the similar. Overall LEP students still outperform their peers by 0.221SD ($p<0.01$). In addition, point estimates for the listening portion of the exam is significant at the 1% level, while point estimates for the speaking and writing portions of the exam are significant at the 10% level.

The results differ for the first grade results of the BAS. Column 2 of panel B of Table 4 indicate that TK students are not reading more difficult books. The coefficient on the ordinal logit is slightly negative and insignificant, while the coefficients on the linear probability models are slightly positive and insignificant. Though there is robust evidence that TK improved pre-literacy skills it did not affect children's ability to read more complex books as measured by BAS.

## 6. Heterogeneity of Results

Aggregate results can be hiding important heterogeneity based on gender, ethnicity and English proficiency status. Despite the regulation of the universal pre-K market, sorting of families to centers of varying quality can remain. TK can mitigate some of these trends because it is free to families and decreases variation in the credentials, compensation of teachers, and the curriculum offered. In this circumstance minority students can particularly benefit from the program.

Columns 1 and 3 of Table 6 indicate that the kindergarten advantages in the BAS are seen in both genders as well as the Asian, Hispanic, LEP, and English proficient subgroups. For brevity, I present intent to treat estimates from my preferred specification for the total number of items missed, the probability of mastering the requisite foundational skills, and the probability of reading at level A or beyond. Looking at the total items missed, all subgroups, except for the white and other subgroups, benefit in the kindergarten administration of the BAS. There is some indication that the Asian subgroup benefitted the most, with the most negative coefficient on the negative binomial portion of the model at -0.363 (or missing 31percent less items). However I cannot reject the null hypothesis that all coefficients on the four racial subgroups are equal ($\chi_3^2 = 5.70$, p<0.1273). Looking at the probability of mastering the requisite number of foundational skills, only male and Asian students were more likely to move onto the leveled reading assessments. Males were 4 percentage points more likely to do so (p<0.10) and Asian students were 12.5 percentage points more likely to do so (p<0.01), and white students were actually 11.6 percentage points less likely to do so. Here I am able to reject the null hypothesis that the effects on the racial subgroups are equal ($\chi_3^2 = 13.71$, p<0.003). No subgroup, however, was more likely to move onto the leveled reading assessments and read at level A or above.

Little heterogeneity is found in the fall first grade BAS results. Here, no subgroup has an advantage in reading books of increasing difficulty. The only estimate that is significant is the probability of reading level E books or above for English proficient students (9.3 percentage points, p<0.05), though with so many first grade outcomes this may occur by chance.

Table 7 presents subgroup results for the CELDT assessment. Again for brevity only the overall results are reported. The white and other subgroup results are not reported due to small sample sizes. Here the female and minority subgroups are driving the results. Column 1 presents the kindergarten results where Hispanic TK-eligible students particularly benefit and outperform their peers by 0.365SD (p<0.01) and female TK-eligible students outperform their peers by 0.241SD (p<0.05). It is worth noting,

however, that the point estimates on the male and Asian subgroups are also positive and large, but the smaller sample size and larger standard errors make it harder to detect a significant effect. I cannot reject the null hypothesis that the male and female effect are equal ($\chi_1^2 = 0.38, p < 0.5378$), nor can I reject the null hypothesis that the effect on the Asian and Hispanic subgroups are equal ($\chi_1^2 = 2.15, p < 0.1428$). Column 2 of Table 7 indicates that in the fall of first grade the female advantage remains at 0.195SD, though the slightly smaller point estimate results in a 10 percent significance level. The Hispanic effect is now half as large and insignificant, and the Asian subgroup now has a 0.268SD ($p<0.01$) advantage. Once again, the male and Hispanic subgroup point estimates are relatively large, but imprecisely estimated due to smaller sample sizes. Again I cannot reject the null hypothesis that the male and female effects are equal ($\chi_1^2 = 0.13\ p < 0.7195$), or that the effects on the Asian and Hispanic subgroups are equal ($\chi_1^2 = 0.29, p < 0.5378$).

Taken together the data indicate that, like the full sample, TK gave most subgroups an advantage in pre-literacy skills, though this did not translate to a higher reading level in first grade. There is some evidence that the Asian subgroup in the BAS experienced larger effects and that the white subgroup benefitted the least. In SFUSD the Asian subgroup is a socio-economically diverse community with many immigrants and first generation Americans. This result is consistent with the notion that the regulation of the TK market mitigates selection effects that put minority students at a disadvantage. Because the vast majority (80 percent) of students assessed with the CELDT are minority students, these results are less useful in comparing results racial lines. However, the CELDT and BAS results reinforce each other in that the Hispanic and Asian subgroups experienced advantages on both assessments.

**7. Robustness Checks**

The results thus far employ the full set of data. While employing the full data maximizes the precision of my estimates, I am relying heavily on the assumption that a linear spline accurately models the relationship between the outcomes and the rating variable. As is standard practice (Schochet et al,

2010), I present evidence that the results are robust to different bandwidths. Figure 7 presents these robustness checks for the main outcomes. Figures A4 through A7 in the appendix[11] present robustness checks for all other results. Each plot presents ITT estimates and their 95 percent confidence intervals for bandwidths that vary from 30 days to 300 days from the threshold. Figure 7 presents robustness checks for the total number of items missed in kindergarten and the overall CELDT scores in kindergarten and first grade. The point estimates are largely stable for bandwidths as small as 30 days on either side of the threshold, though the significance tends to decrease as the bandwidths get shorter. This is expected because as the bandwidth decreases so does the sample size thereby increasing the standard errors

I employ a second robustness check by running a series of placebo regression discontinuities. The effects previously seen should occur uniquely at the December 2nd threshold. Moving the threshold to any other date should result in null effects. To test this proposition I move the threshold 30, 40, and 50 days on either side of the threshold. Table 8 presents the results of this exercise for the total items missed in kindergarten, and the overall results for the CELDT in both grades. Column 4 is bolded to emphasize the reduced form results from the original regression discontinuity. All other columns present the results for each outcome after the rating variable and instrument were centered at the specified placebo date. The significant results found in column 4 largely disappear in these placebo specifications. There are two significant point estimates, but with no clear pattern. Overall, Table 8 shows that the kindergarten and first grade effects are not present at other thresholds.

The last robustness check builds off recent work by recent work by Barreca et al. (2015) who find that in situations, such as birthday rating variables, heaping can cause bias in regression point estimates if observations in the heaped portions of the data are systematically different from observations in the non-heaped portion of the data. To investigate this bias they recommend estimating the effects on

[11] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://www3.interscience.wiley.com/cgibin/jhome/34787.

heaped and non-heaped data separately. The histogram in Figure 2(a) shows that there could be heaping in the birthday variable, with about 15 to 32 students concentrated in some values of the rating variable. These heaps are generally larger than the sample average of 18.5 students born in a day. I investigate whether this heaping is biasing the point estimates by re-estimating my main results on portions of the data that exclude successively smaller heaps. In Table 9 I present point estimates of the outcomes from portions of the data that exclude heaps with more than 25, 20, 18, and 15 students born on the same day.

The results indicate that heaping induced bias does not seem to be a concern in this study. Eliminating the biggest heaps containing more than 25 or 20 students does little to the point estimates. The estimates grow larger as smaller heaps are eliminated, but the sample size is also reduced. Point estimates are noticeably larger after heaps containing more than 18 or 15 students are eliminated, but at this point less than half the sample remains. Most importantly, in these most restrictive situations, the overall conclusion of the study remains intact: there are significant gains for TK-eligible students.

**8. Discussion and Policy Implications**

This paper presents evidence that Transitional Kindergarten produces greater literacy gains in students when compared to pre-K programs available to families as part of the San Francisco's universal pre-K program. Large gains in pre-literacy skills are found in the fall of kindergarten. Despite the causal nature of the study, one issue complicates the inference. The district uses the BAS as a formative assessment tool from TK to 3rd grade. Assuming other pre-K programs in the city do not use the assessment, TK students were exposed to the assessment up to three times in the previous year. Similarly because LEP students are assessed every year they are in the district, TK students were exposed to the CELDT a year before non-TK students. The fall results may therefore be biased due to a practice effect. The first grade CELDT results indicate that this practice effect is not likely to be an issue. In first grade all LEP students have been assessed with the CELDT at least once and the advantages remain. These pre-literacy advantages, however, did not translate into the ability to read books of increasing difficulty.

There are two main mechanisms by which TK could provide literacy benefits to children. The first is through efforts to align the curriculum to the development of children in this age range. The district created its own curriculum that academically and socio-emotionally spanned the middle ground between its pre-K and kindergarten curriculums. The greater alignment between student cognitive development and the academic demands of the curriculum may contribute to the gains seen in this study. This study provides evidence that a more academically oriented curriculum can lead to increased student learning. Given the link between student outcomes, even at a young age, and improved longer term outcomes (Chetty et al, 2011) these academic gains can be consequential. This study cannot test the effect of TK on socio-emotional skills, though SFUSD also emphasized socio-emotional development in its TK curriculum. Winship et al (2015) found that other TK programs increased executive functioning skills of children and did not affect the socio-emotional skills of children. Together these results suggest that the academic gains do not have to come at the expense of the socio-emotional health of children.

The second mechanism stems from the greater regulation that resulted from folding this pre-K program into the larger K-12 system. Parents who could only take advantage of the subsidized half day pre-K programs offered by the city now have the option to enroll their children in a free, full day program. In addition, TK teachers are more educated, more highly paid, and solely situated in schools. This regulation likely decreased the variance in the quality of programs offered and the types of teachers available to students. Prior literature has shown that minority and socio-economically disadvantaged families often select into less formal prekindergarten or lower quality prekindergarten experiences (Magnuson et al, 2004; Magnuson & Waldfogel, 2005; Phillips & Lowenstein, 2011; Capizzano, 2006). If TK provides these families with larger amounts of higher quality instruction, we would expect them to particularly benefit from this program. This study shows that among all students, there is evidence that the Asian subgroup saw the greatest benefits in the BAS, particularly in satisfying the pre-literacy requirements needed to move on to the reading assessment, while the white subgroup saw the least benefits. Furthermore the Asian and Hispanic subgroups saw benefits on both the BAS and CELDT

assessments. These results support studies such as Hotz and Xiao (2011) and Rigby et al (2007) who find that regulated markets lead to a greater uptake of formal early childhood care and improved student outcomes. This aspect of the program suggests that these results are best generalizable to other large, urban districts with significant numbers of low income and minority students.

In addition, a back-of-the envelope calculation estimates that these literacy benefits come at a lower cost. In 2012-2013 San Francisco's universal pre-K budget of $19.67 million served 3,225 students at a cost of $6,099 per student. The program provides 612.5 hours of instruction for a total cost of $9.96 per hour per student. In 2012-2013 the district spent $9,479 per pupil (California Department of Education, 2012). TK is funded at the same per pupil cost as the rest of the district and provides students with 6 hours of instruction a day for 180 days. As result, I estimate that TK costs SFUSD $8.78 per hour per student. These estimates imply that TK is more cost effective than the universal pre-K program.

Recently, the Transitional Kindergarten program has been expanded with the introduction of Extended TK. Starting in the 2015-2016 school year, children who turn five after December 2, 2015 and before the end of school year can either enter Transitional Kindergarten at the time they turn 5 (in the middle of the year), or start TK at the beginning of the school year (Torlakson, 2015). The exact details are left to the discretion of individual school districts. This study does not provide evidence on whether extending TK to all four year olds, essentially making it a form of universal pre-K, will benefit children. On one hand, offering free pre-K services to all four year olds would likely benefit families. However, if TK is meant to be a modified kindergarten curriculum, more scrutiny is needed to determine if the TK curriculums are appropriate for younger children. Like all regression discontinuity studies, the results are valid only for children born near the December 2 threshold. Extrapolation of the results to children significantly older or younger is not valid. This is especially pertinent in this case because children of this age develop rapidly in a relatively small amount of time. This study indicates that for students near the December 2 threshold this district's efforts to provide a middle ground between pre-K and kindergarten have been successful.

## References


Applied Survey Research (2013a). Evaluating Preschool for All effectiveness: Research brief. Retrieved, June 26, 2015, from http://www.first5sf.org/sites/default/files/pagefiles/Evaluating%20PFA%20Effectiveness%20%20Research%20Brief.pdf

Applied Survey Research (2013b). Evaluating Preschool for All quality: Research brief. Retrieved, June 26, 2015, from http://www.first5sf.org/sites/default/files/pagefiles/Evaluating%20PFA%20Quality%20%20Research20Brief.pdf

Anderson, M.L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. Journal of the American Statistical Association, 103, 1481-1495.

Barnett, W.W. (2011). Effectiveness of early education interventions. Science. 333, 975-978

Barreca, A. I., Lindo, J.M., Waddell, G.R. (2015) Heaping induced bias in regression discontinuity designs. Economic Inquiry.

Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S., (forthcoming). Within-and between sector quality differences in early childhood education and care. Child Development.

Bassok, D., Fitzpatrick, M., Loeb, S., & Paglayan, A.S. (2013). The early childhood care and education workforce inthe United States: Understanding changes from 1990 through 2010. Education Finance and Policy, 8, 581–601.

Bassok, D., & Latham, S. (2014). Kids today: Changes in school readiness in an early childhood era. EdPolicyWorks,University of Virginia. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/ 35_Kids_Today.pdf

Bassok, D., & Rorem, A. (2014). Is kindergarten the new first grade? The changing nature of kindergarten in the age of accountability. EdPolicyWorks, University of Virginia. Retrieved, September 14, 2015, from http://curry.virginia.edu/uploads/resourceLibrary/20_Bassok_Is_Kindergarten_The_New_First_Grade.pdf.

Belfield, C., Nores, M., Barnett, W. S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost benefit analysis using data from age 40. Journal of Human Resources, 16, 162-190.

California Department of Education. (2012). Cost per average daily attendance. Retrieved, September 14, 2015, from http://www.cde.ca.gov/ds/fd/ec/currentexpense.asp

Cameron, A.C & Trivedi, P.K. (2010). Microeconomics using Stata. College Station, TX: Stata Press.

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., Sparling, J. & Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project followup. Developmental Psychology, 48, 1033-1043

Capizzano, J, Adams, G., Ost, J. (2006). Caring for children of color: The child care patterns of white, black, and Hispanic children under 5. The Urban Institute Occasional Paper 76. Retrieved, June 26, 2015 from http://www.urban.org/sites/default/files/alfresco/publicationpdfs/311285-Caring for-Children-of-Color.PDF

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. The Quarterly Journal of Economics, 126, 1593–1660.

Controller Office. (2013). City services measure performance report. Retrieved, October 7, 2015, from http://sfcontroller.org/Modules/ShowDocument.aspx?documentid=4957.

Currie, J. & Thomas, D. (1995), Does Head Start make a difference? The American Economic Review, 85, 341-364.

Currie, J. & Thomas, D. (2000), School quality and the longer-term effects of Head Start. The Journal of Human Resources, 35, 755-774.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Econometrics, 1, 111-134

Early Education Department. (2012). PreK–3rd annual report. Retrieved, June 27, 2015, from http://www.sfusd.edu/en/assets/sfusd-staff/programs/files/Early%20Education/PreK 3rd%20Report%20Year%20One_7-18-13.pdf

Edsource (2015). San Francisco to expand preschool program. Retrieved, June 27, 2015 from: http://edsource.org/2015/sanfrancisco-to-expand-preschool-program/72818#.VLZ VivF950

Elkind, D., & Whitehurst, G. (2001). Young Einsteins: Should Head Start emphasize academic skills? EducationNext, 1. Retrieved, September 14, 2015, fromhttp://educationnext.org/young-einsteins/

Fountas and Pinnell (2012). Field study of reliability and validity of the Fountas and Pinnell Benchmark Assessement Systems 1 and 2. Retrieved, July 8, 2015 from http://www.heinemann.com/fountasandpinnell/research/BASFieldStudyFullReport.pdf

Fuller, B., Kagan, S. L., Loeb, S., & Chang, Y.-W. (2004). Child care quality: Centers and home settings that serve poor families. Early Childhood Research Quarterly, 19, 505–527.

Garces, E., Thomas, D., Currie, J. (2002). Longer-term effects of Head Start. The American Economic Review, 92, 999-1012.

Gormley, W.T., Gayer, T., Phillips, D, & Dawson, B. (2005). The effects of universal pre-K on cognitive development. Developmental Psychology, 41, 872-884

Governor's State Advisory Council on Early Learning and Care (2013). Transitional kindergarten implementation guide: A resource for California public school district administrator and teachers. Retrieved, April 3, 2015 from http://www.cde.ca.gov/ci/gs/em/documents/tkguide.pdf

Hatch, J. A. (2002). Accountability shovedown: Resisting the standards movement in early childhood education. Phi Delta Kappa, 83, 457–462.

Heckman, J.J., Seong, H. M., Pinto, R., Savelyev, P.A., Yavitz, A. (2010) Analyzing social experiments asimplemented: A reexamination of the evidence from the High Scope Perry Preschool Program. Quantitative Economics, 1, 1-46

Hotz, V. J., & Xiao, M. (2011). The impact of regulations on the supply and quality of care in child care markets. American Economic Review, 101, 1775–1805.

Huang, F.L., Ivernizzi, M.A., Drake, E. A. (2012) The differential effects of preschool: Evidence from Virginia. Early Childhood Research Quarterly, 27, 33-45.

Kulik, J.A., Kulik, C.C., Bangert, R.L. (1984). Effects of practice on aptitude and achievement test scores. American Educational Research Journal, 2, 435-447

Imbens, G.W., & Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. The Review of Economic Studies, 1-27

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142, 615-635

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

Lee, V. E., Loeb, S., & Lubeck, S. (1998). Contextual effects of pre-K classrooms for disadvantaged children on cognitive development: The case of Chapter 1. Child Development, 69, 479-494.

Legislative Analyst Office. (2012). Preschool and transitional kindergarten. Retrieved, September 14, 2015 from http://www.lao.ca.gov/handouts/education/2012/Preschool_and_Transitional_Kindergarten_41212.pdf

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. (2007). How much is too much? The effects of duration and intensity of child care experiences on children's social and cognitive development. Economics of Education Review, 26, 52-66.

Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality and stability. Child Development, 75, 47-65.

Long, J. S. & Freese, J. (2014). Regression models for categorical dependent variables using Stata. College Station, TX: Stata Press

Ludgwig, J., Miller, D.L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. The Quarterly Journal of Economics, 159-208.

Manship, K., Quick, H, Holod, A., Mill, N., Ogut, B., Chernoff, J.J., Blum, J., Hauser, A., Anthony, J., Gonzalez R. (2015). Impact of California's Transitional Kindergarten program, 2013-2014. American Institutes For Research. Retrieved, December 29, 2015 from http://www.air.org/resource/impact-californias-transitional-kindergarten-program-2013-14.

Magnuson, K.A, Meyers,M.K., Ruhm, C.J., Waldfogel, J. (2004). Inequality in preschool education and school readiness. American Education Research Journal, 41, 115-157.

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

Phillips, D.A, Lowenstein, A.E, (2011). Early care, education, and child development. Annual Review of Psychology, 62, 483-500.

Puma, M. et al (2010). Head start impact study: final report. U.S. Department of Health and Human Services. Administration for Children and Families Retrieved, July 2, 2015 from http://www.acf.hhs.gov/sites/default/files/opre/hs_impact_study_final.pdf

Rigby, E., Ryan, R.M., Brooks-Gunn, J. (2007). Child care quality in different state policy contexts. Journal of Policy Analysis and Management, 26, 887-907.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs. What Works Clearinghouse.

Stipek, D. (2006). No child left behind comes to preschool. The Elementary School Journal, 106, 455-66.

Torlakson, T. (2015). Amendement to California education code 48000(c). Retrieved, September 14, 2015 from http://www.cde.ca.gov/nr/el/le/yr15ltr0717.asp

Vuong, Q.H (1989). Likelihood ratio tests for model selection & non-nested hypotheses. Econometrica, 57, 307-33.

Wong, V.C., Cook, T.D, Barnett, W. S., Jung, K. (2008). An effectiveness-based evaluation of five state prekindergarten programs. Journal of Policy Analysis and Management, 27, 122-154.

Zigler, E. F., & Bishop-Josef, S. J. (2006). The cognitive child versus the whole child: Lessons from 40 years of Head Start. Play= Learning: How Play Motivates and Enhances Children's Cognitive and Social Emotional Growth, 15–35.

*2013/2014*

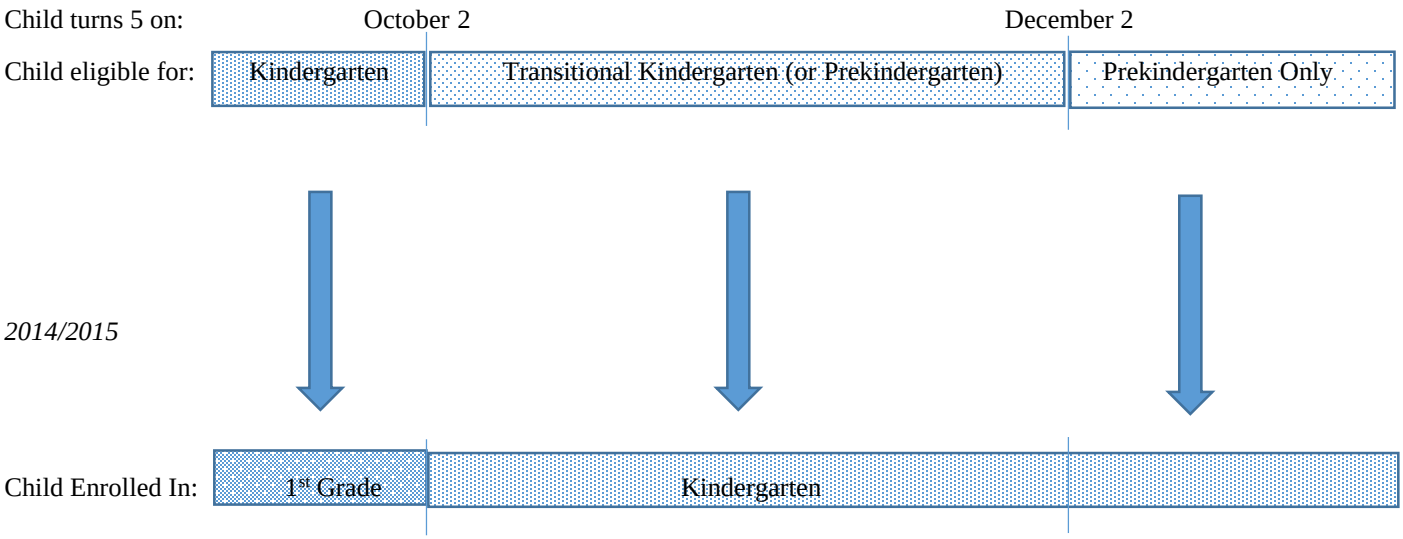
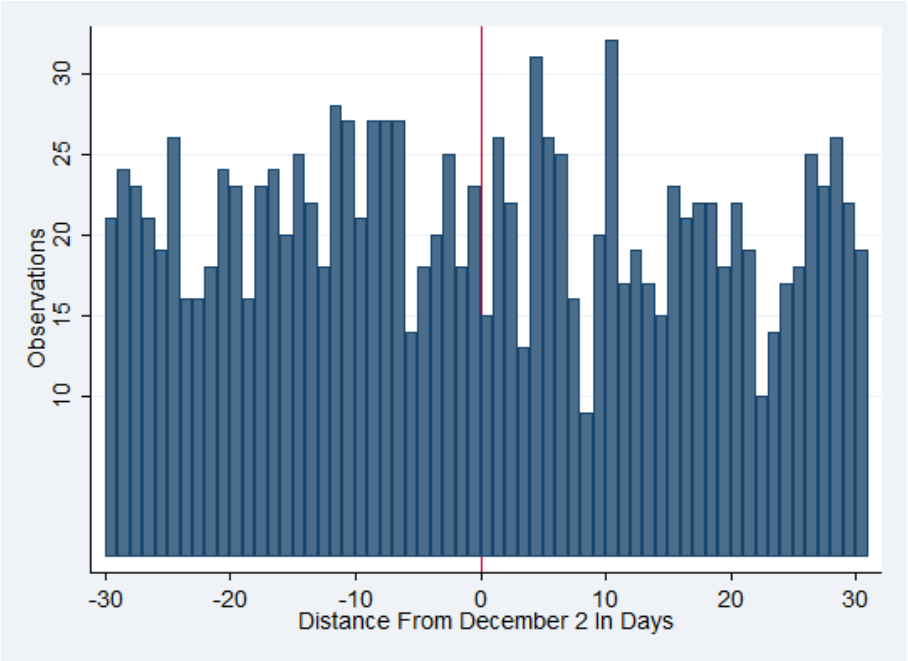



**Figure 1:** Early childhood education experience based on birthdate cut point for cohort 2

| Prekindergarten | Transitional Kindergarten |
|---|---|
| **Structure of Day** | |
| Children start at different times based on contract | Academic day starts at same time for all children |
| Families select hours of instruction | 6 hour program |
| Breakfast provided | No breakfast but may have morning snack |
| Nap time | No nap time |
| 1 hour of outdoor time | 15-20 minutes of outdoor time |
| **Curriculum** | |
| Activities and pace are based on child's skill | Activities and pace more structured |
| No curriculum map or timeline | Curriculum map and timeline exist |
| Whole group instruction lasts no more than 10 minutes | Whole group instruction lasts no more than 10 minutes |
| Whole group instruction used less frequently | Whole group instruction used more frequently |
| **Class Size** | |
| Maximum class size of 24 students | Maximum class size of 22 students |
| 1 adult for every 8 children | 1 paraprofessional for first 6 weeks |

**Figure 2:** Differences in SFUSD Transitional Kindergarten and prekindergarten programs




**(a)**


**(b)**

**Figure 3:** Histogram of observations by birthday and McCrary density test. Birthdays are centered at December 2 such that the x-axis represents the distance in days from December 2. TK ineligible students are to the left of the threshold and TK eligible students are to the right of the threshold. Figure (a) presents birthdays ranging from -30 to 30 days. Each bar indicates the number of observations born in a 1 day bin. Figure (b) presents the results from a McCrary density test. The point estimate and standard error of the discontinuity is 0.110 (0.087). Vertical lines indicate the December 2 threshold.

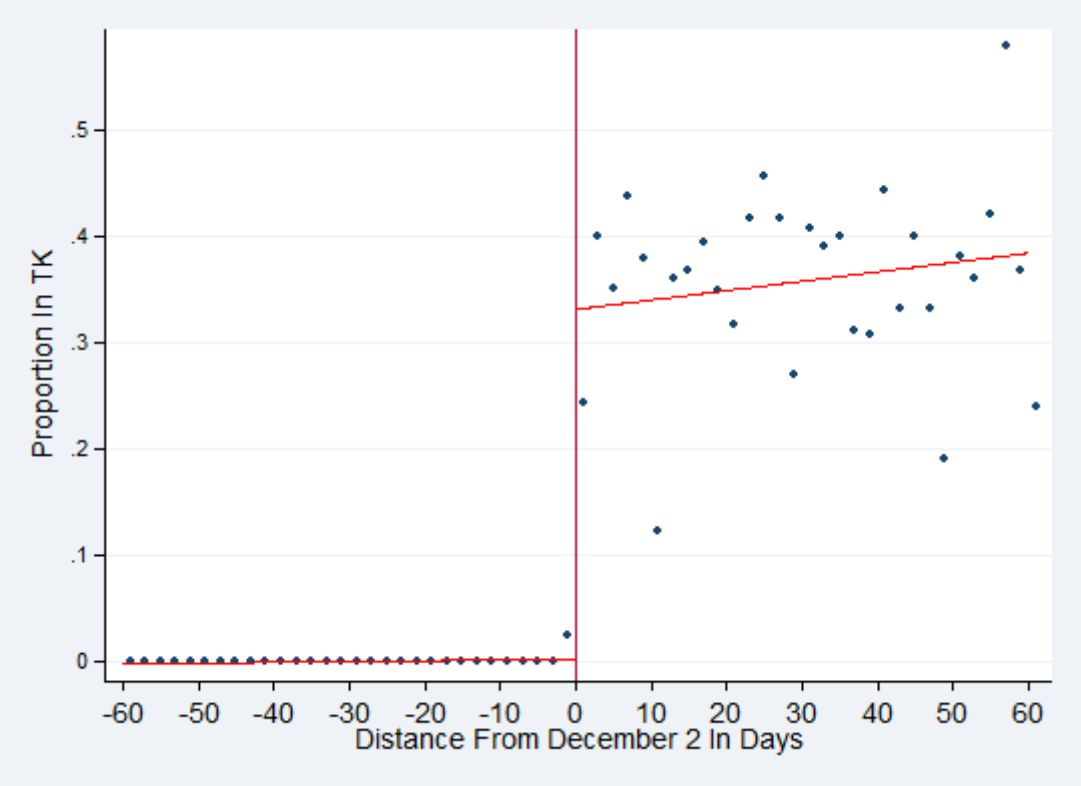
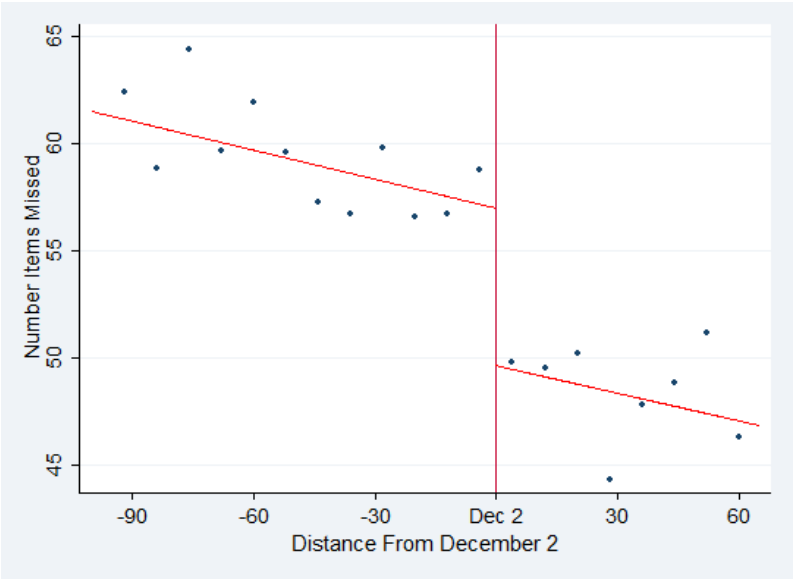
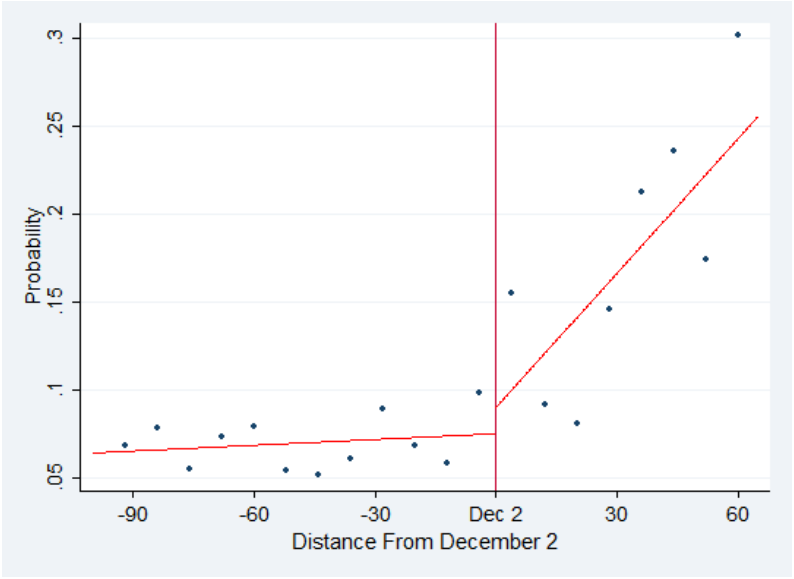


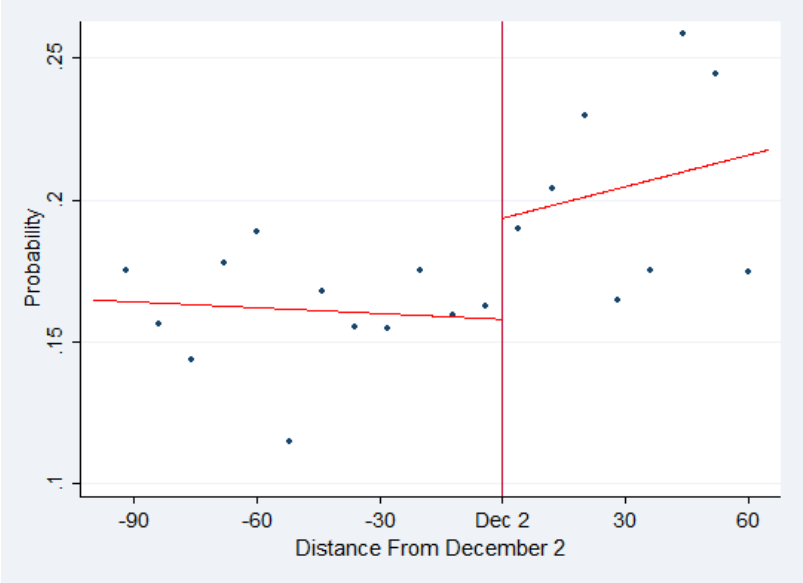

**Figure 4**: First Stage: Enrollment in TK in prior year by birthday. Each dot represents the proportion of students that enrolled in TK in the previous year within a bin of 2 days. The vertical line represents the December 2 threshold. Regression lines are estimated using local linear regression with a rectangular kernel on a bandwith of 60 days.



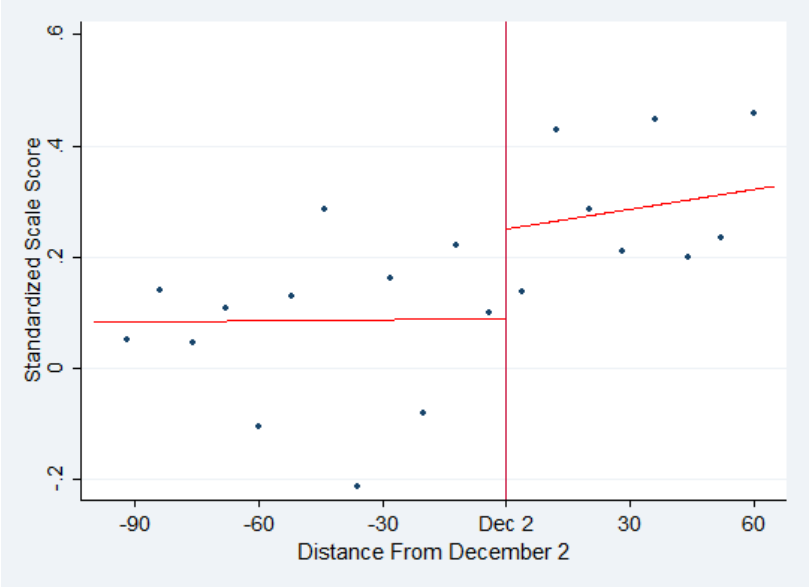

**(a) Total Items Missed**




**(b) Pr(Mastering Enough Foundational Skills)**

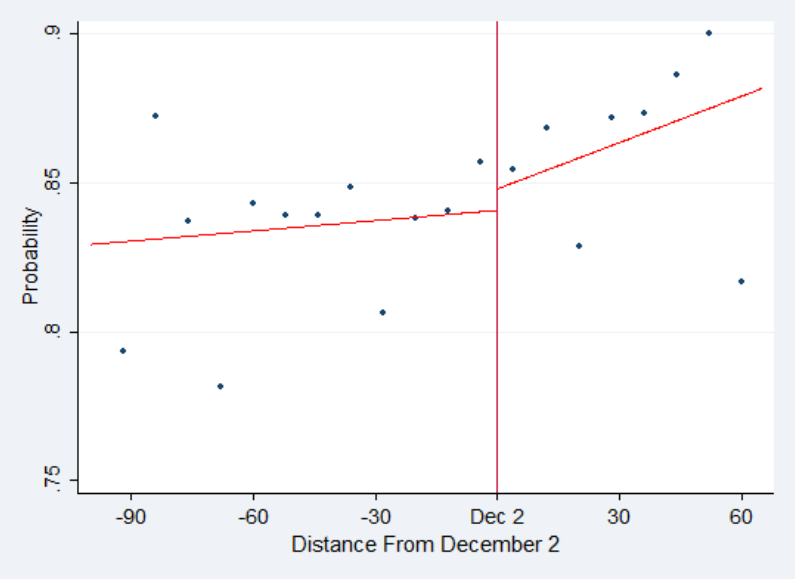



**(c) Pr(Reading At Level A or Above)**

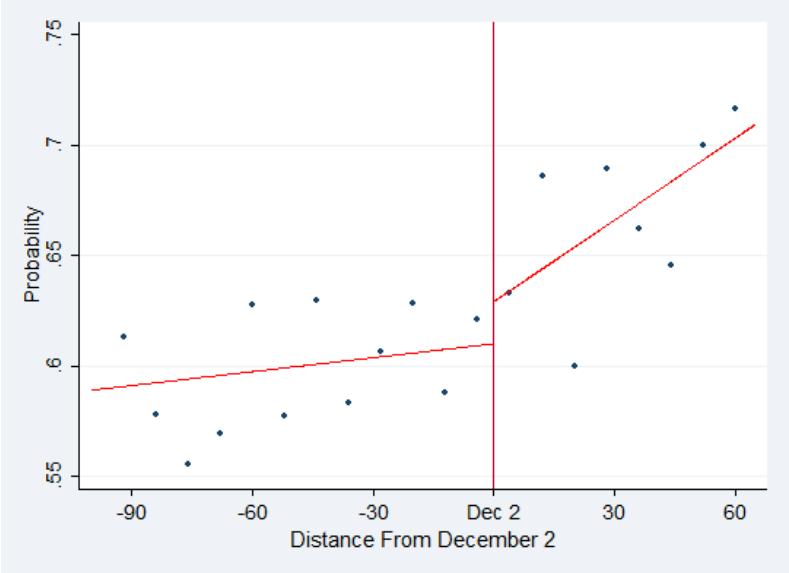



**(d) Overall CELDT Score**

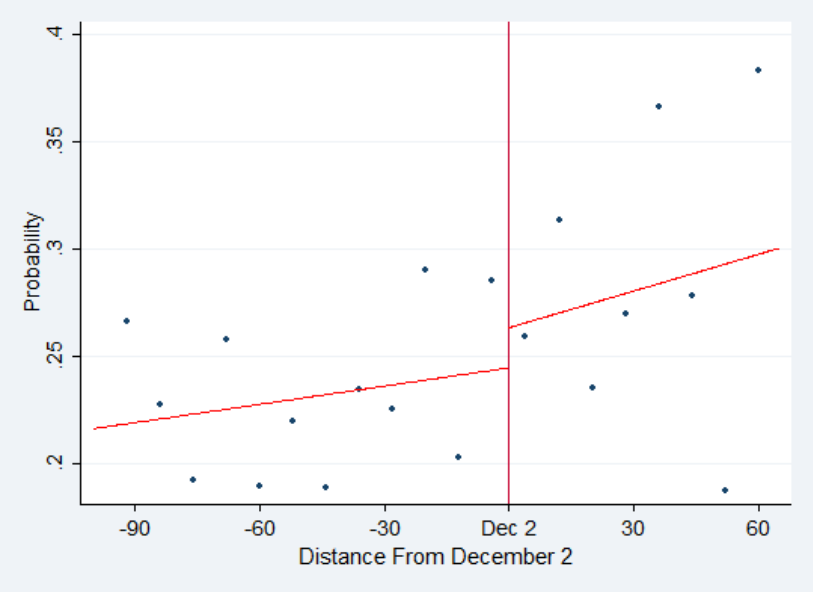
**Figure 5:** Fall kindergarten literacy outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. Birthdays are centered at December 2. The total items missed is the sum of items missed in the following skills: upper case letter recognition, lower case letter recognition, letter sounds, initial word sounds, high frequency words, early literacy behaviors, blending, and rhyming. Figures 4(a) – 4(c) are Fountas and Pinnell Outcomes. Figure 4(d) is the overall CELDT score.



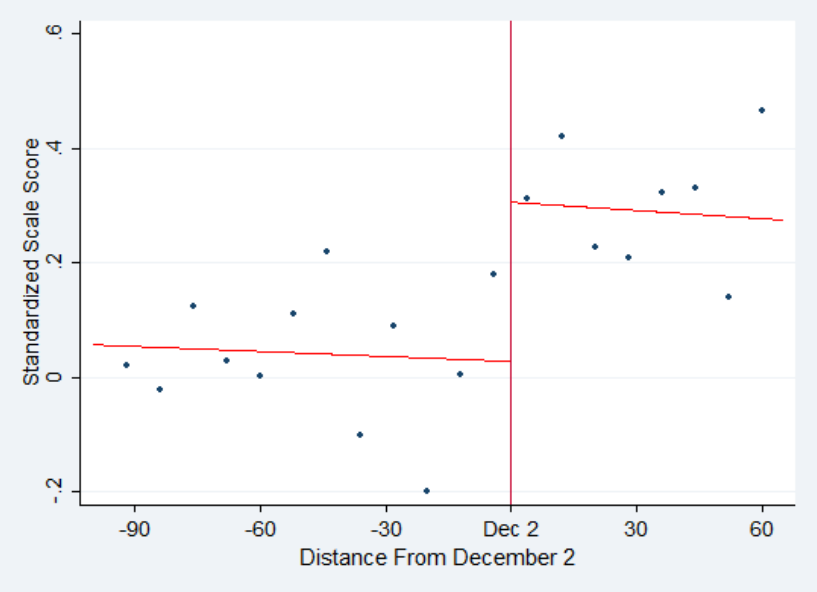

**(a) Pr(Reading at Level C or Above)**




**(b) Pr(Reading At Level E or Above)**

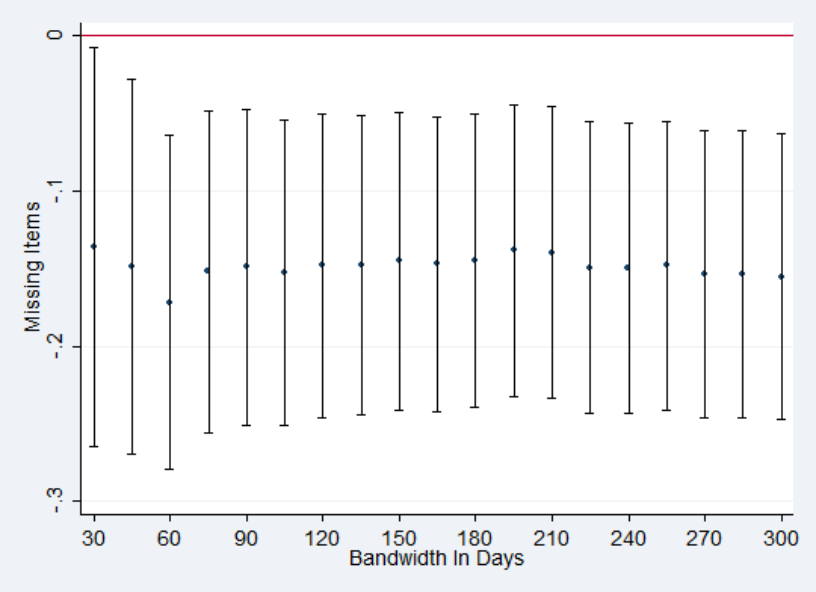



**(c) Pr(Reading at Level I or Above)**

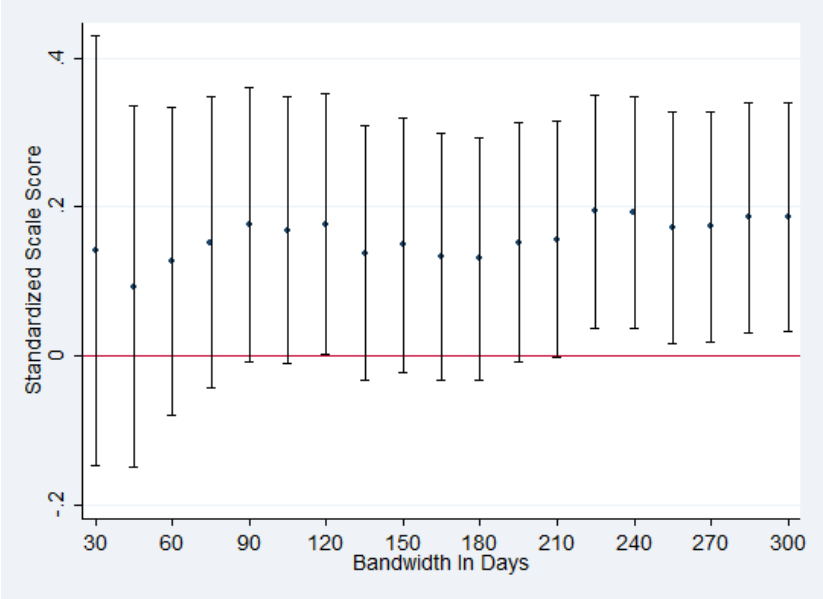



**(d) Overall CELDT Score**

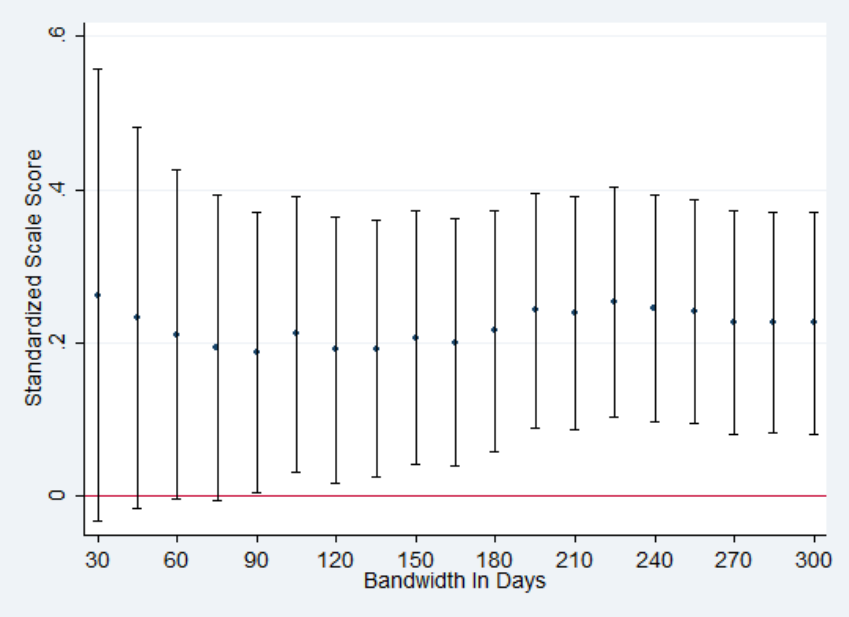
**Figure 6:** Fall first grade literacy outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. Birthdays are centered at December 2. Figures 5(a)-5(c) are Fountas and Pinnel outcomes. Figure 5(d) is the overall CELDT score.




**(a) Total Items Missed In Fall Kindergarten BAS**




**(b) Overall Fall Kindergarten CELDT Score**




**(c) Overall Fall First Grade CELDT Score**

**Figure 7:** Robustness checks of main BAS and CELDT outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relavant outcome for observations in bandwidths between 30 and 300 days. Figure 6(a) employs a negative binomial model and represents the total items missed from the fall kindergarten administration of the BAS. The total items missed is the sum of items missed in the following skills: upper case letter recognition, lower case letter recognition, letter sounds, initial word sounds, high frequency words, early literacy behaviors, blending, and rhyming. Figures 6(b) and 6(c) employ OLS models and present results from fall kindergarten and first grade administration of the CELDT. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 5. Standard errors are clustered on the birthday rating variable except in 6(a) where it must be clustered at the teacher-by-year cell.

Table 1: Descriptive statistics

| | Analytical Sample | | | | | TK Students | | Pre-K Students | | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | St. Dev. | Min | Max | N (Total) | Mean | N (TK) | Mean | N (Pre-K) | (TK-Pre-K) |
| Programmatic Characteristics | | | | | | | | | | |
| TK Eligible | 0.140 | 0.347 | 0 | 1 | 6773 | 0.997 | 336 | 0.096 | 6437 | 0.000 |
| Attended TK In Year T-1 | 0.050 | 0.217 | 0 | 1 | 6773 | 1.000 | 336 | 0.000 | 6437 | --- |
| Attended District Pre-K in Year T-1 | 0.169 | 0.375 | 0 | 1 | 6773 | 0.000 | 336 | 0.178 | 6437 | 0.000 |
| Birthday (days from December 2) | -120.263 | 98.408 | -304 | 61 | 6773 | 26.146 | 336 | -127.905 | 6437 | 0.000 |
| Student Characteristics | | | | | | | | | | |
| Female | 0.490 | 0.500 | 0 | 1 | 6773 | 0.488 | 336 | 0.491 | 6437 | 0.929 |
| Asian | 0.311 | 0.463 | 0 | 1 | 6773 | 0.423 | 336 | 0.305 | 6437 | 0.000 |
| Hispanic | 0.251 | 0.434 | 0 | 1 | 6773 | 0.259 | 336 | 0.251 | 6437 | 0.746 |
| White | 0.164 | 0.370 | 0 | 1 | 6773 | 0.098 | 336 | 0.167 | 6437 | 0.001 |
| Other | 0.175 | 0.380 | 0 | 1 | 6773 | 0.179 | 336 | 0.174 | 6437 | 0.841 |
| Declined To State Ethnicity | 0.098 | 0.297 | 0 | 1 | 6773 | 0.042 | 336 | 0.101 | 6437 | 0.000 |
| Special Education | 0.078 | 0.267 | 0 | 1 | 6773 | 0.033 | 336 | 0.080 | 6437 | 0.002 |
| Limited English Proficient (LEP) | 0.494 | 0.500 | 0 | 1 | 6773 | 0.595 | 336 | 0.488 | 6437 | 0.000 |
| Home Language: | | | | | | | | | | |
| Chinese | 0.170 | 0.376 | 0 | 1 | 6773 | 0.298 | 336 | 0.164 | 6437 | 0.000 |
| Spanish | 0.150 | 0.357 | 0 | 1 | 6773 | 0.173 | 336 | 0.149 | 6437 | 0.234 |
| English | 0.595 | 0.491 | 0 | 1 | 6773 | 0.455 | 336 | 0.602 | 6437 | 0.000 |
| Other | 0.084 | 0.278 | 0 | 1 | 6773 | 0.074 | 336 | 0.085 | 6437 | 0.497 |
| Dominant Language: | | | | | | | | | | |
| Chinese | 0.205 | 0.404 | 0 | 1 | 6773 | 0.307 | 336 | 0.200 | 6437 | 0.000 |
| Spanish | 0.176 | 0.380 | 0 | 1 | 6773 | 0.182 | 336 | 0.175 | 6437 | 0.767 |
| English | 0.504 | 0.500 | 0 | 1 | 6773 | 0.417 | 336 | 0.508 | 6437 | 0.001 |
| Other | 0.115 | 0.320 | 0 | 1 | 6773 | 0.095 | 336 | 0.117 | 6437 | 0.234 |
| Kindergarten Fountas and Pinnell Outcomes | | | | | | | | | | |
| Upper Case Letters | 20.379 | 8.379 | 0 | 29 | 6773 | 22.509 | 336 | 20.268 | 6437 | 0.000 |
| Lower Case Letters | 18.774 | 8.614 | 0 | 29 | 6773 | 21.866 | 336 | 18.613 | 6437 | 0.000 |
| Letter Sounds | 12.660 | 9.140 | 0 | 29 | 6773 | 17.530 | 336 | 12.405 | 6437 | 0.000 |
| High Frequency Words | 6.901 | 7.812 | 0 | 25 | 6773 | 13.685 | 336 | 6.547 | 6437 | 0.000 |
| Initial Word Sounds | 5.282 | 3.223 | 0 | 8 | 6773 | 6.426 | 336 | 5.223 | 6437 | 0.000 |
| Early Literacy Behaviors | 6.904 | 3.055 | 0 | 11 | 6773 | 8.402 | 336 | 6.826 | 6437 | 0.000 |
| Blending | 3.790 | 4.099 | 0 | 10 | 6460 | 5.792 | 317 | 3.687 | 6143 | 0.000 |
| Rhyming | 5.714 | 4.087 | 0 | 10 | 6026 | 7.270 | 293 | 5.634 | 5733 | 0.000 |
| Mastered Required Found. Skills | 0.070 | 0.255 | 0 | 1 | 6773 | 0.238 | 336 | 0.061 | 6437 | 0.000 |
| Reading at Level A or Above | 0.167 | 0.373 | 0 | 1 | 6773 | 0.223 | 336 | 0.164 | 6437 | 0.005 |
| Test Given In Spanish | 0.140 | 0.347 | 0 | 1 | 6773 | 0.131 | 336 | 0.141 | 6437 | 0.609 |
| Kindergarten CELDT Outcomes | | | | | | | | | | |
| Listening | 374.275 | 86.161 | 220 | 570 | 3344 | 419.660 | 200 | 371.388 | 3144 | 0.000 |
| Speaking | 387.669 | 94.781 | 140 | 630 | 3344 | 428.245 | 200 | 385.088 | 3144 | 0.000 |
| Reading | 294.282 | 57.568 | 220 | 570 | 3344 | 343.360 | 200 | 291.160 | 3144 | 0.000 |
| Writing | 306.158 | 52.411 | 220 | 600 | 3344 | 352.635 | 200 | 303.202 | 3144 | 0.000 |
| Overall | 372.429 | 77.748 | 184 | 580 | 3344 | 415.880 | 200 | 369.665 | 3144 | 0.000 |
| First Grade Fountas and Pinnell Outcomes | | | | | | | | | | |
| Reading at Level C or Above | 0.819 | 0.385 | 0 | 1 | 6219 | 0.870 | 315 | 0.816 | 5904 | 0.000 |
| Reading at Level E or Above | 0.568 | 0.495 | 0 | 1 | 6219 | 0.692 | 315 | 0.562 | 5904 | 0.000 |
| Reading at Level I or Above | 0.211 | 0.408 | 0 | 1 | 6219 | 0.308 | 315 | 0.205 | 5904 | 0.000 |
| First Grade CELDT Outcomes | | | | | | | | | | |
| Listening | 454.059 | 63.509 | 220 | 570 | 2697 | 485.486 | 181 | 451.798 | 2516 | 0.000 |
| Speaking | 456.716 | 66.331 | 140 | 630 | 2697 | 483.939 | 181 | 454.757 | 2516 | 0.000 |
| Reading | 396.269 | 76.464 | 220 | 570 | 2697 | 426.193 | 181 | 394.117 | 2516 | 0.000 |
| Writing | 400.408 | 57.651 | 220 | 600 | 2697 | 431.033 | 181 | 398.205 | 2516 | 0.000 |
| Overall | 449.188 | 57.241 | 184 | 594 | 2697 | 478.597 | 181 | 447.072 | 2516 | 0.000 |

*Note:* Former TK students are who enrolled in the district's TK program in the previous year. Former prekindergarten students are students who enrolled in the district's prekindergarten program in the previous year. 2013-2014 and 2014-2015 Kindergarten administrative data contained student characteristics, including exact birthdate. Administrative data were linked to district test files to obtain Fountas and Pinnell and CELDT outcome data. Former TK and prekindergarten were identified by linking kindergarten administrative data to the district TK and pre-K administrative data sets from the previous school year. TK stands for Transitional Kindergarten, pre-K stands for prekindergarten, and CELDT stands for California English Language Development Test.

Table 2: RD regressions of covariate balance

| Variable | Full Sample | $\|B_{ict}\| \leq 60$ | $\|B_{ict}\| \leq 30$ |
|---|---|---|---|
| Student Characteristics | | | |
| Female | 0.010 | -0.015 | -0.027 |
| | (0.029) | (0.037) | (0.050) |
| Asian | -0.017 | -0.047 | -0.040 |
| | (0.035) | (0.044) | (0.059) |
| Hispanic | 0.018 | 0.018 | -0.015 |
| | (0.028) | (0.035) | (0.045) |
| White | -0.029 | -0.032 | -0.002 |
| | (0.028) | (0.035) | (0.050) |
| Other | 0.046+ | 0.037 | 0.033 |
| | (0.025) | (0.034) | (0.055) |
| Declined To State Ethnicity | -0.019 | 0.021 | 0.018 |
| | (0.018) | (0.024) | (0.030) |
| Special Education | -0.011 | -0.018 | -0.010 |
| | (0.015) | (0.018) | (0.022) |
| Limited English Proficient (LEP) | -0.027 | -0.057 | -0.077 |
| | (0.038) | (0.046) | (0.065) |
| Home Language: | | | |
| Chinese | 0.001 | -0.015 | -0.033 |
| | (0.030) | (0.034) | (0.047) |
| Spanish | -0.001 | -0.010 | -0.020 |
| | (0.019) | (0.028) | (0.040) |
| English | -0.015 | -0.011 | 0.038 |
| | (0.035) | (0.041) | (0.061) |
| Other | 0.015 | 0.037+ | 0.015 |
| | (0.015) | (0.020) | (0.026) |
| Dominant Language: | | | |
| Chinese | -0.017 | -0.046 | -0.064 |
| | (0.029) | (0.034) | (0.046) |
| Spanish | -0.007 | 0.002 | 0.005 |
| | (0.020) | (0.027) | (0.037) |
| English | 0.027 | 0.049 | 0.071 |
| | (0.037) | (0.045) | (0.064) |
| Other | -0.002 | -0.005 | -0.012 |
| | (0.018) | (0.024) | (0.031) |
| Test Characteristic | | | |
| Test Given In Spanish | -0.023 | -0.009 | 0.032 |
| | (0.025) | (0.033) | (0.044) |
| N | 6,773 | 2,191 | 1,278 |

*Note:* Each cell represents the results of a separate regression discontinuity estimate of the covariate balance. Row headers indicate the appropriate covariate tested. Column headers indicate the bandwidth restriction. In all regressions the functional form is a linear spline. Akaike's Information Criterion indicates a linear spline is the optimal functional form for the majority of covariates. All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01

Table 3: RD regressions of first stage

| *Dependent Variable: Enrolled In TK in Year T-1* | | | |
|---|---|---|---|
| | (1) | (2) | N |
| Full Sample | 0.335** | 0.320** | 6,773 |
| | (0.032) | (0.027) | |
| $\|B_{ict}\| \leq 60$ | 0.329** | 0.304** | 2,191 |
| | (0.032) | (0.031) | |
| $\|B_{ict}\| \leq 30$ | 0.313** | 0.282** | 1,278 |
| | (0.041) | (0.045) | |
| Covariates | | √ | |
| Fixed Effects | | √ | |

*Note:* Each cell represents the results of a separate first stage regression discontinuity estimate. The dependent variable in all regressions is an indicator for enrolling in TK in the previous year. Row headers indicate the bandwidth restriction. Covariates include all variables in Table 2. Covariates also include an indicator for kindergarten year, and teacher-by-year fixed effects. The functional form in all regressions is a linear spline. Akaike's Information Criterion indicates a linear spline is the optimal functional form. All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01

Table 4: Reduced form estimates of fall kindergarten and first grade literacy outcomes

| | (1) | (2) | | | (3) | (4) | |
|---|---|---|---|---|---|---|---|
| Panel A: Fall Kindergarten Outcomes | | | | | | | |
| Fountas And Pinnell Outcomes | | | N | CELDT Outcomes | | | N |
| Total Items Missed | -0.143* | -0.182** | 6,773 | Overall Score | 0.121 | 0.183* | 3,344 |
| | (0.059) | (0.041) | | | (0.109) | (0.078) | |
| Upper Case Letters | -0.291* | -0.339** | 6,773 | Listening | 0.139 | 0.186* | 3,344 |
| | (0.132) | (0.086) | | | (0.105) | (0.079) | |
| Lower Case Letters | -0.228* | -0.163* | 6,773 | Speaking | 0.070 | 0.138+ | 3,344 |
| | (0.102) | (0.068) | | | (0.105) | (0.078) | |
| Letter Sounds | -0.131* | -0.185** | 6,773 | Reading | 0.200* | 0.221* | 3,344 |
| | (0.055) | (0.050) | | | (0.098) | (0.091) | |
| High Frequency Words | -0.101** | -0.142** | 6,773 | Writing | 0.195+ | 0.205** | 3,344 |
| | (0.035) | (0.038) | | | (0.101) | (0.077) | |
| Early Literacy Behaviors | -0.161 | -0.211** | 6,773 | | | | |
| | (0.099) | (0.060) | | | | | |
| Initial Word Sounds | -0.157 | -0.214* | 6,773 | | | | |
| | (0.109) | (0.090) | | | | | |
| Rhyming | -0.163 | -0.192* | 6,026 | | | | |
| | (0.103) | (0.080) | | | | | |
| Blending | -0.036 | -0.099* | 6,460 | | | | |
| | (0.054) | (0.049) | | | | | |
| Pr(Mastering Required Found. Skills) | 0.011 | 0.033 | 6,773 | | | | |
| | (0.022) | (0.021) | | | | | |
| Pr(Reading at Level A or Above) | 0.019 | 0.012 | 6,773 | | | | |
| | (0.028) | (0.016) | | | | | |
| Panel B: Fall First Grade Outcomes | | | | | | | |
| Fountas And Pinnell Outcomes | | | N | CELDT Outcomes | | | N |
| Reading Scale (Ordinal Logit) | -0.051 | -0.036 | 6,219 | Overall Score | 0.236** | 0.221** | 2,697 |
| | (0.120) | (0.120) | | | (0.090) | (0.073) | |
| Pr(Reading at Level C or Above) | 0.007 | 0.008 | 6,219 | Listening | 0.288** | 0.286** | 2,697 |
| | (0.027) | (0.023) | | | (0.086) | (0.078) | |
| Pr(Reading at Level E or Above) | 0.013 | 0.021 | 6,219 | Speaking | 0.139 | 0.125+ | 2,697 |
| | (0.038) | (0.030) | | | (0.092) | (0.075) | |
| Pr(Reading at Level I or Above) | 0.021 | 0.017 | 6,219 | Reading | 0.143 | 0.098 | 2,697 |
| | (0.031) | (0.028) | | | (0.113) | (0.088) | |
| | | | | Writing | 0.227* | 0.170+ | 2,697 |
| | | | | | (0.109) | (0.091) | |
| Covariates | | √ | | | | √ | |
| Fixed Effects | | √ | | | | √ | |

*Note:* Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated literacy outcome. Row headers indicate the dependent variable. Columns 1 and 2 present estimates for Fountas and Pinnell outcomes. Columns 3 and 4 present estimates for CELDT outcomes. Covariates include an indicator for kindergarten year, teacher-by-year fixed effects, and all variables in Table 2. Negative binomial models are used to estimate the effect of transitional kindergarten on foundational literacy skills, ordinal logit models are used to estimate the effect of transitional kindergarten on literacy skills, and OLS is used in all other models. The functional form of all regressions is a linear spline. Akaike's Information Criteria indicates a linear spline is optimal. All standard errors are clustered on the day of birth running variable except for the conditional negative binomial and ordinal logit models which must be clustered on the kindergarten teacher. +indicates p<0.10, *p<0.05, **p<0.01

Table 5: Reduced form incidence rate ratio estimates of fall kindergarten literacy outcomes

| Literacy Outcome | (1) Incidence Rate Ratio | (2) Avg Number of Items Missed by Control Group | (3) Fewer Items Missed By TK Students |
|---|---|---|---|
| Total Items Missed | 0.833** | 51.865 | 8.817 |
| Upper Case Letters | 0.712** | 6.212 | 1.789 |
| Lower Case Letters | 0.849* | 7.475 | 1.129 |
| Letter Sounds | 0.831** | 13.288 | 2.246 |
| High Frequency Words | 0.868** | 17.501 | 2.310 |
| Early Literacy Behaviors | 0.810** | 2.888 | 0.549 |
| Initial Word Sounds | 0.808* | 2.402 | 0.461 |
| Rhyming | 0.825** | 4.096 | 0.717 |
| Blending | 0.905* | 5.895 | 0.560 |
| Covariates | √ | √ | √ |
| Fixed Effects | √ | √ | √ |

*Note:* Column 1 presents results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated literacy outcome. Row headers indicate the dependent variable. Point estimates in column 1 represents the incidence rate ratios of the point estimates in column 2 of Table 4. Column 3 represents the average number of items missed by the control group born within 30 days of the Transitional Kindergarten threshold. Included covariates are defined in Table 4. +indicates p<0.10, *p<0.05, **p<0.01

Table 6: ITT RD estimates of kindergarten and first grade Fountas and Pinnell outcomes by subgroup

| Kindergarten | | 1st Grade | |
|---|---|---|---|
| | (1) | | (2) |
| *Panel A: Full Sample, N= 6,773* | | *N=6,219* | |
| Total Items Missed | -0.182** | Reading Scale | -0.036 |
| | (0.041) | | (0.120) |
| Pr(Mastering Required Found. Skills) | 0.033 | Pr(Level C or Above) | 0.008 |
| | (0.021) | | (0.023) |
| Pr(Reading at Level A or Above) | 0.012 | Pr(Level E or Above) | 0.021 |
| | (0.016) | | (0.030) |
| | | Pr(Level I or Above) | 0.017 |
| | | | (0.028) |
| *Panel B: Male, N=3,451* | | *N=3,144* | |
| Total Items Missed | -0.216** | Reading Scale | -0.136 |
| | (0.059) | | (0.167) |
| Pr(Mastering Required Found. Skills) | 0.046+ | Pr(Level C or Above) | 0.018 |
| | (0.027) | | (0.034) |
| Pr(Reading at Level A or Above) | 0.040+ | Pr(Level E or Above) | -0.021 |
| | (0.022) | | (0.043) |
| | | Pr(Level I or Above) | -0.010 |
| | | | (0.041) |
| *Panel C: Female, N=3,322* | | *N=3,075* | |
| Total Items Missed | -0.165** | Reading Scale | 0.078 |
| | (0.061) | | (0.177) |
| Pr(Mastering Required Found. Skills) | 0.023 | Pr(Level C or Above) | -0.017 |
| | (0.030) | | (0.034) |
| Pr(Reading at Level A or Above) | -0.021 | Pr(Level E or Above) | 0.064 |
| | (0.024) | | (0.047) |
| | | Pr(Level I or Above) | 0.039 |
| | | | (0.042) |
| *Panel D: Asian, N=2,105* | | *N=2,017* | |
| Total Items Missed | -0.382** | Reading Scale | 0.133 |
| | (0.086) | | (0.215) |
| Pr(Mastering Required Found. Skills) | 0.125** | Pr(Level C or Above) | 0.049 |
| | (0.048) | | (0.035) |
| Pr(Reading at Level A or Above) | 0.022 | Pr(Level E or Above) | 0.004 |
| | (0.028) | | (0.054) |
| | | Pr(Level I or Above) | 0.028 |
| | | | (0.054) |
| *Panel E: Hispanic, N=1,703* | | *N=1,546* | |
| Total Items Missed | -0.180** | Reading Scale | -0.146 |
| | (0.066) | | (0.241) |
| Pr(Mastering Required Found. Skills) | 0.027 | Pr(Level C or Above) | -0.091 |
| | (0.022) | | (0.065) |
| Pr(Reading at Level A or Above) | 0.016 | Pr(Level E or Above) | -0.022 |
| | (0.025) | | (0.070) |
| | | Pr(Level I or Above) | 0.018 |
| | | | (0.045) |

| Kindergarten | | 1st Grade | |
|---|---|---|---|
| | (3) | | (4) |
| *Panel F: White N=1,111* | | *N=1,001* | |
| Total Items Missed | -0.039 | Reading Scale | -0.122 |
| | (0.128) | | (0.331) |
| Pr(Mastering Required Found. Skills) | -0.116* | Pr(Level C or Above) | 0.031 |
| | (0.058) | | (0.052) |
| Pr(Reading at Level A or Above) | -0.033 | Pr(Level E or Above) | 0.039 |
| | (0.056) | | (0.089) |
| | | Pr(Level I or Above) | 0.151 |
| | | | (0.097) |
| *Panel G: Other N=1,182* | | *N=1,068* | |
| Total Items Missed | 0.021 | Reading Scale | -0.136 |
| | (0.115) | | (0.280) |
| Pr(Mastering Required Found. Skills) | -0.040 | Pr(Level C or Above) | 0.055 |
| | (0.056) | | (0.072) |
| Pr(Reading at Level A or Above) | -0.024 | Pr(Level E or Above) | -0.016 |
| | (0.044) | | (0.090) |
| | | Pr(Level I or Above) | -0.145+ |
| | | | (0.075) |
| *Panel H: Limited English Proficient (LEP), N=3,344* | | *N=3,115* | |
| Total Items Missed | -0.169** | Reading Scale | -0.084 |
| | (0.055) | | (0.173) |
| Pr(Mastering Required Found. Skills) | 0.044 | Pr(Level C or Above) | -0.011 |
| | (0.029) | | (0.036) |
| Pr(Reading at Level A or Above) | 0.012 | Pr(Level E or Above) | -0.057 |
| | (0.019) | | (0.045) |
| | | Pr(Level I or Above) | -0.026 |
| | | | (0.039) |
| *Panel I: English Proficient N=3,429* | | *N=3,104* | |
| Total Items Missed | -0.227** | Reading Scale | 0.067 |
| | (0.063) | | (0.170) |
| Pr(Mastering Required Found. Skills) | 0.019 | Pr(Level C or Above) | 0.027 |
| | (0.030) | | (0.032) |
| Pr(Reading at Level A or Above) | 0.012 | Pr(Level E or Above) | 0.093* |
| | (0.026) | | (0.043) |
| | | Pr(Level I or Above) | 0.056 |
| | | | (0.041) |

*Note:* Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated literacy outcome. Row headers indicate the dependent variable and panel headers indicate the subsample. Negative binomial models were used to estimate the effect of transitional kindergarten on the total items missed, ordinal logit models were used to estimate the effet of transitional kindergarten on the reading level, and OLS models were used in all other cases. All functional forms include a linear spline and covariates defined in Table 4. Akaike's Information Criteria indicates a linear spline is optimal. All standard errors are clustered on the day of birth running variable, except for the conditional negative binomial and ordinal logit models which must be clustered on the teacher-by-year cell. +indicates p<0.10, *p<0.05, **p<0.01

Table 7: ITT RD estimates of kindergarten and first grade CELDT outcomes by subgroup

| Dependent Variable: Overall Score | Kindergarten | | First Grade | |
|---|---|---|---|---|
| | (1) | N | (2) | N |
| All Limited English Proficient (LEP) | 0.183* | 3,344 | 0.221** | 2,697 |
| | (0.078) | | (0.073) | |
| Male | 0.143 | 1,690 | 0.198 | 1,382 |
| | (0.118) | | (0.120) | |
| Female | 0.241* | 1,654 | 0.195+ | 1,315 |
| | (0.111) | | (0.104) | |
| Asian | 0.110 | 1,533 | 0.268** | 1,301 |
| | (0.116) | | (0.098) | |
| Hispanic | 0.365** | 1,179 | 0.162 | 970 |
| | (0.136) | | (0.136) | |

*Note:* Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the overall CELDT scale score. Row headers indicate the subsample. All functional forms include a linear spline and covariates defined in Table 4. Akaike's Information Criteria indicates a linear spline is optimal. All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01

Table 8: Robustness check: Placebo estimates of fall and midyear literacy outcomes

| | (1) | (2) | (3) | **(4)** | (5) | (6) | (7) | |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Kindergarten Outcomes* | $B_{ict}$-50 | $B_{ict}$-40 | $B_{ict}$-30 | **$B_{ict}$** | $B_{ict}$+30 | $B_{ict}$+40 | $B_{ict}$+50 | N |
| Total Items Missed | -0.075 | -0.084 | -0.136+ | **-0.182**** | 0.034 | 0.039 | 0.061* | 6,773 |
| | (0.112) | (0.083) | (0.071) | **(0.041)** | (0.033) | (0.032) | (0.031) | |
| Overall CELDT Score | -0.248 | -0.100 | 0.141 | **0.183*** | 0.040 | -0.095 | -0.053 | 3,344 |
| | (0.251) | (0.123) | (0.118) | **(0.078)** | (0.074) | (0.073) | (0.069) | |
| *Panel B: First Grade Outcomes* | | | | | | | | |
| Overall CELDT Score | -0.006 | 0.156 | 0.132 | **0.221**** | -0.007 | -0.086 | -0.034 | 2,697 |
| | (0.212) | (0.132) | (0.137) | **(0.073)** | (0.074) | (0.076) | (0.077) | |
| Covariates | √ | √ | √ | √ | √ | √ | √ | |
| Fixed Effects | √ | √ | √ | √ | √ | √ | √ | |

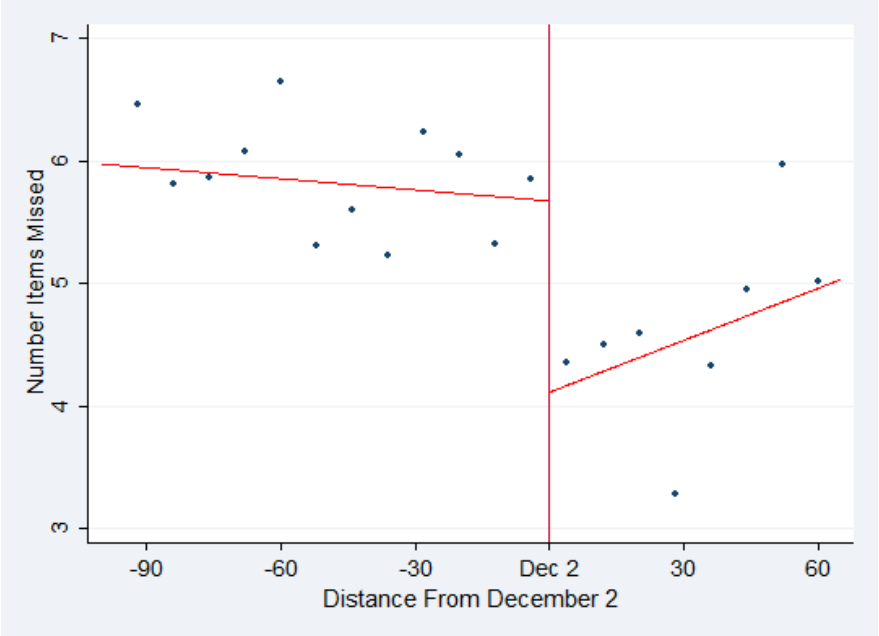
*Note:* Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated literacy outcome. Row headers indicate the dependent variable. Column 4 contains estimates from the regression discontinuity found in Table 4, Columns 2 and 4. All other columns contain estimates from placebo RDs. Covariates are the same as those in Table 4. The functional form of all regressions is a linear spline. All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01

Table 9: Robustness check: Estimates after eliminating heaps

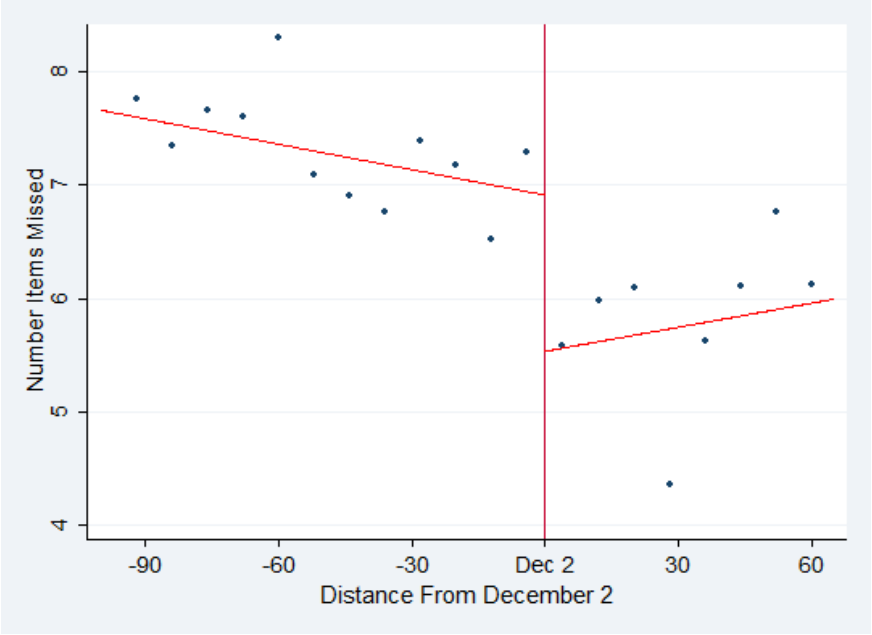
| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Panel A: Fall Outcomes* | **Full Sample** | $H_B \leq 25$ | $H_B \leq 20$ | $H_B \leq 18$ | $H_B \leq 15$ |
| Total Items Missed | **-0.182**** | -0.254** | -0.295** | -0.371** | -0.332** |
| | **(0.041)** | (0.050) | (0.068) | (0.078) | (0.115) |
| N | **6,773** | 5,682 | 3,417 | 2,519 | 1,287 |
| Overall CELDT Score | **0.183*** | 0.225* | 0.185 | 0.279+ | 0.398+ |
| | **(0.078)** | (0.091) | (0.119) | (0.146) | (0.211) |
| N | **3,344** | 2,804 | 1,691 | 1,247 | 653 |
| *Panel B: Midyear Outcomes* | | | | | |
| Total Items Missed | **0.221**** | 0.251** | 0.168 | 0.308* | 0.302 |
| | **(0.073)** | (0.091) | (0.133) | (0.143) | (0.217) |
| N | **2,697** | 2,271 | 1,363 | 1,005 | 541 |

*Note:* Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated literacy outcome. Row headers indicate the dependent variable. Column 1 contains estimates from regression discontinuity found in Table 4, Columns 2 and 4. All other columns contain estimates from samples obtained from by eliminating heaps of varying sizes. $H_B$ represents heaps at values of the running variable, $B_{ict}$ . Heaps greater than the value in the column headers were eliminated from the sample. Covariates include those used in Table 4. The functional form of all regressions is a linear spline. +indicates p<0.10, *p<0.05, **p<0.01

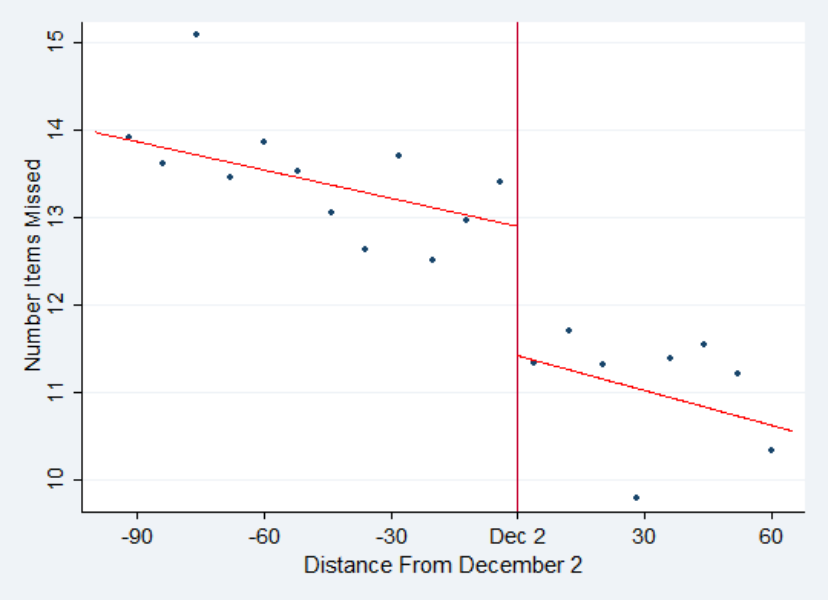
## Appendix



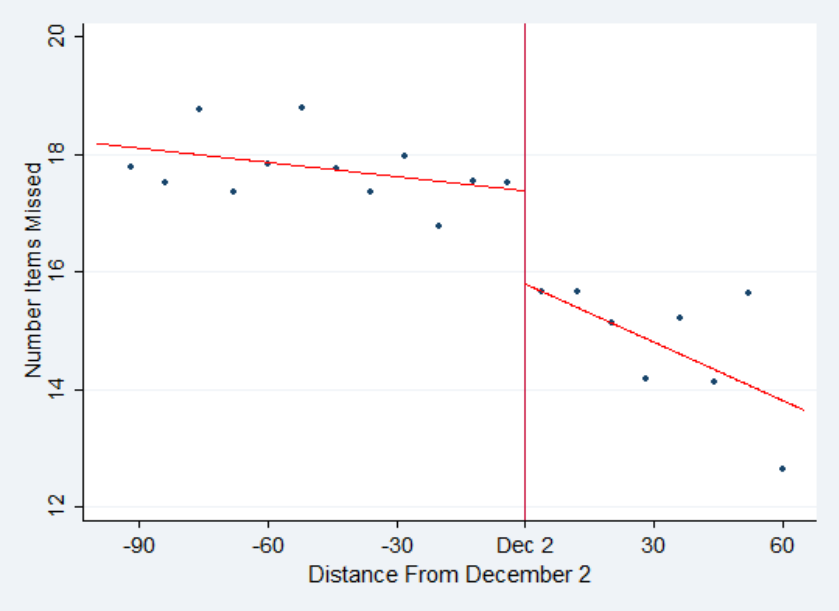

**(a) Upper Case Letter Recognition**



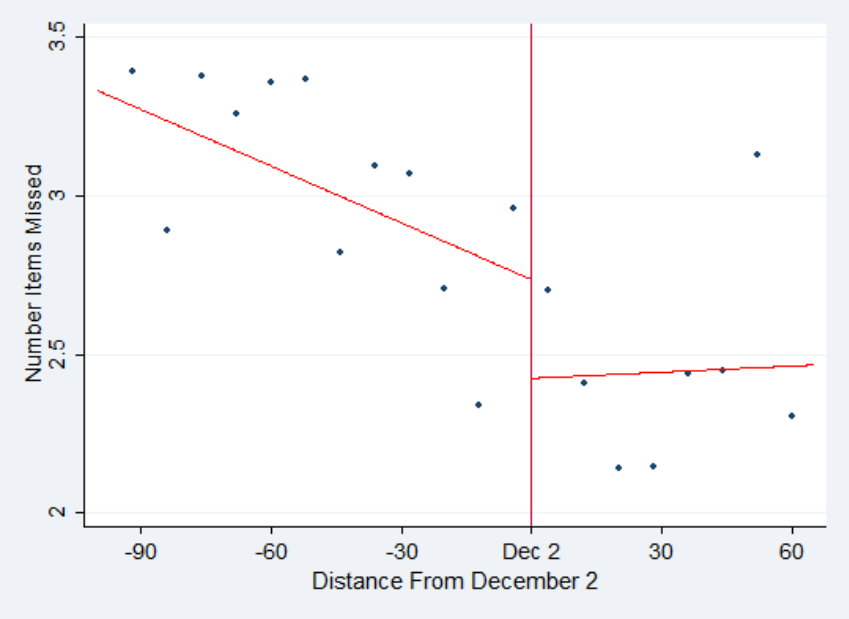

**(b) Lower Case Letter Recognition**



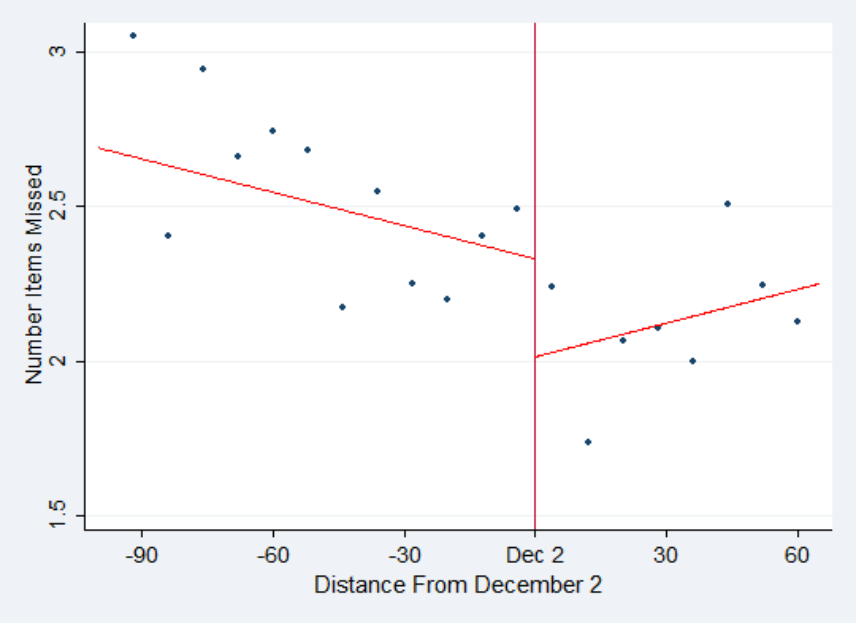

**(c) Letter Sounds**



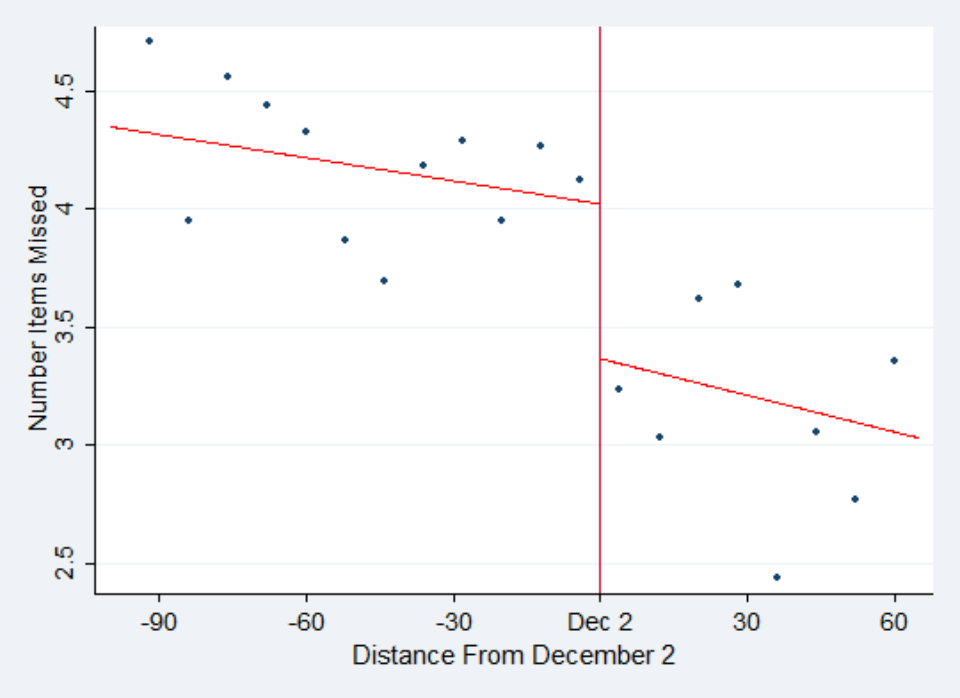

**(d) High Frequency Word Recognition**




**(e) Early Literacy Behaviors**

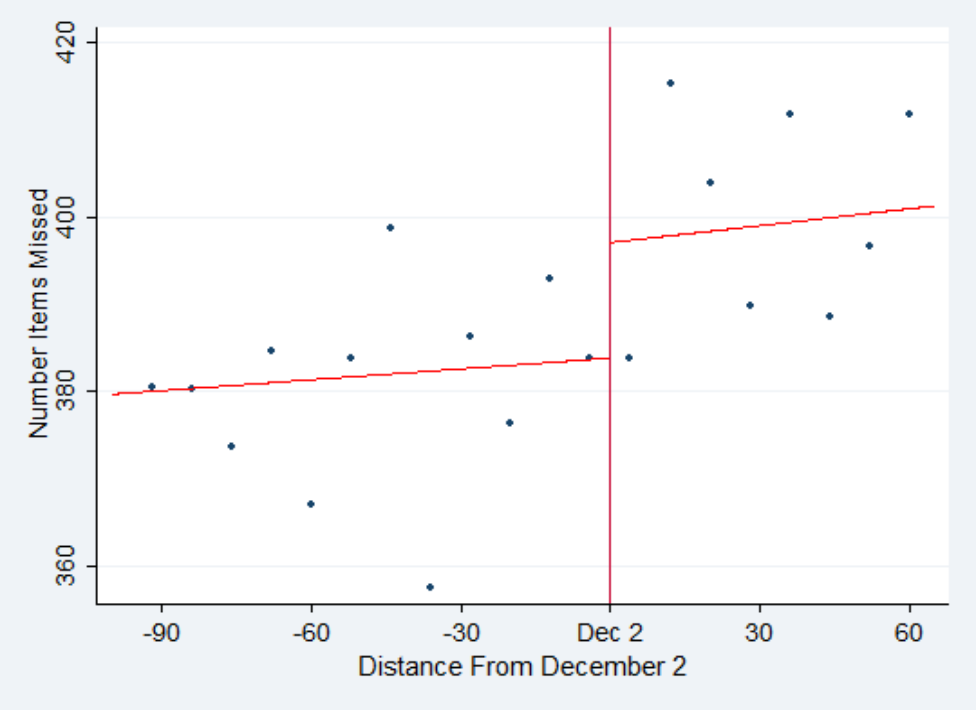



**(f) Initial Word Sounds**

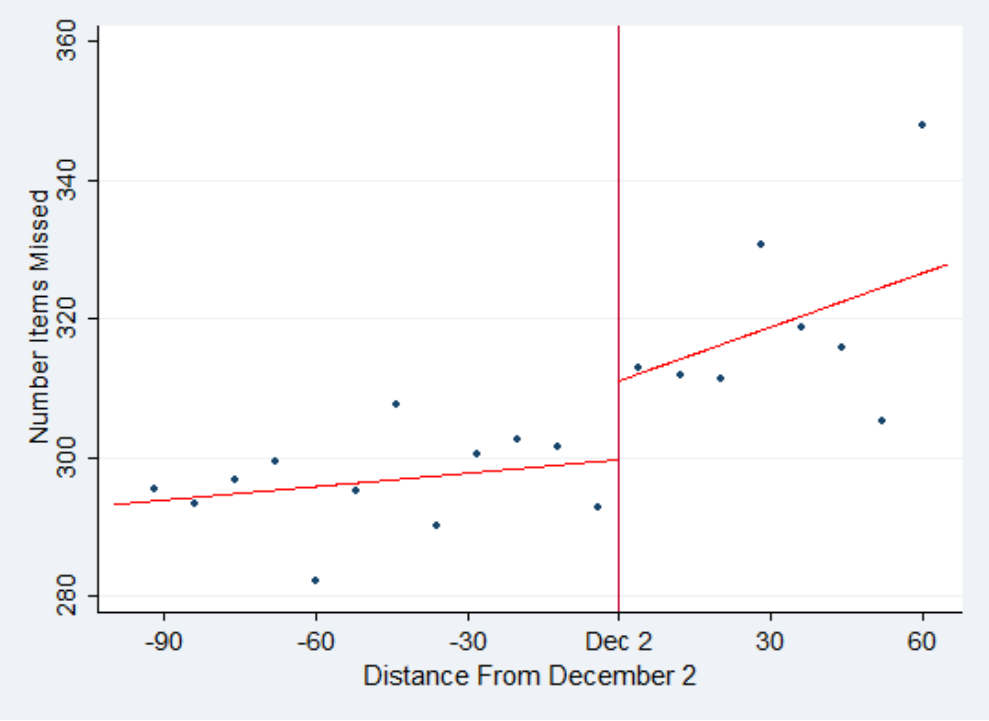



**(g) Rhyming**

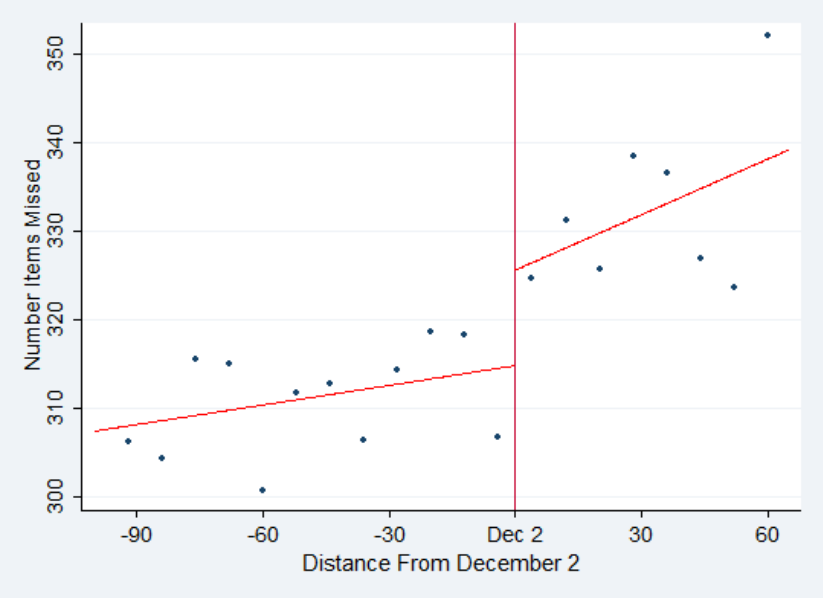
**Figure A1:** Fall kindergarten Fountas and Pinnell foundational literacy outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. Birthdays are centered at December 2.



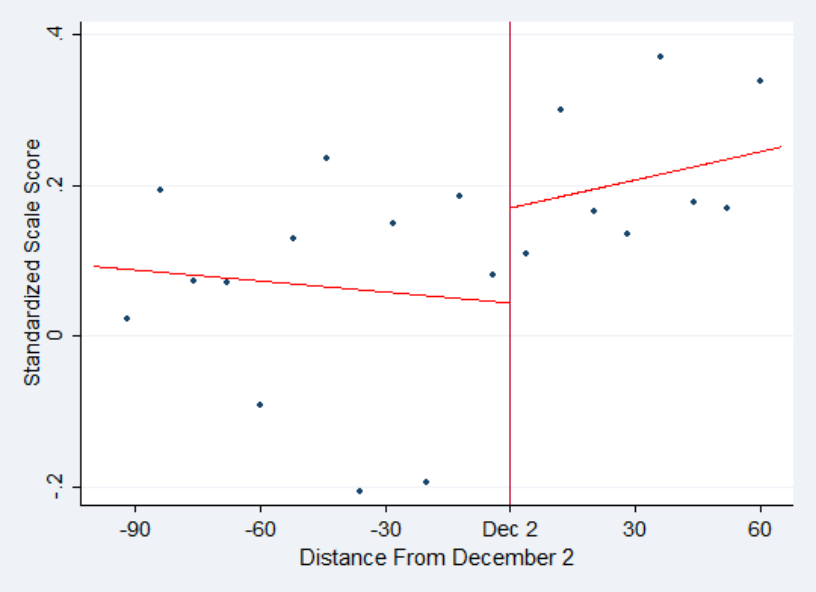

**(a) Listening**




**(b) Reading**

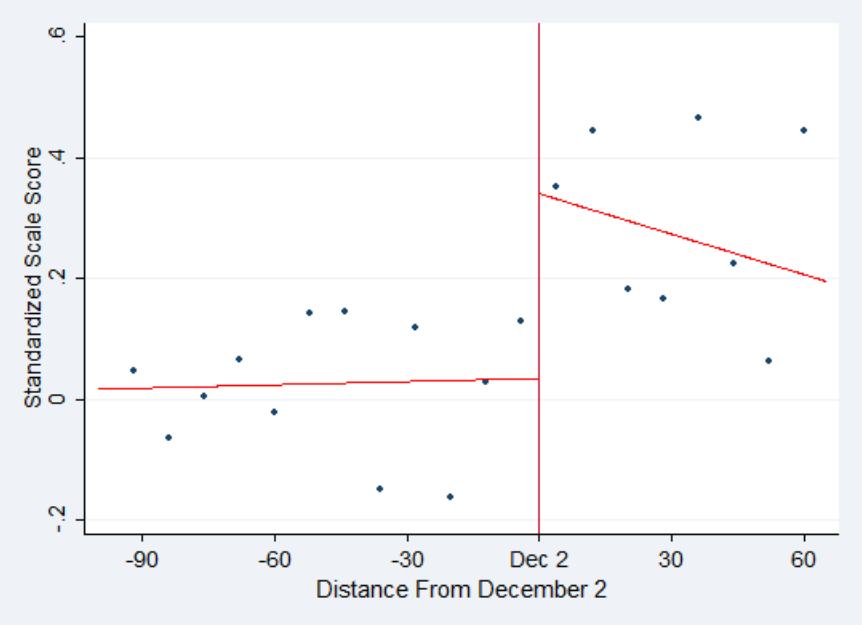



**(c) Writing**

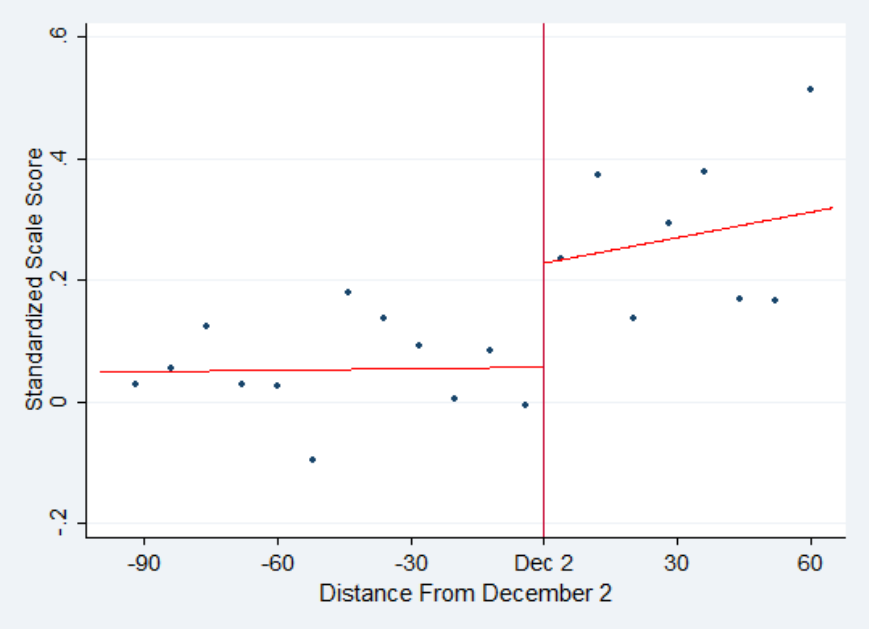



**(d) Speaking**

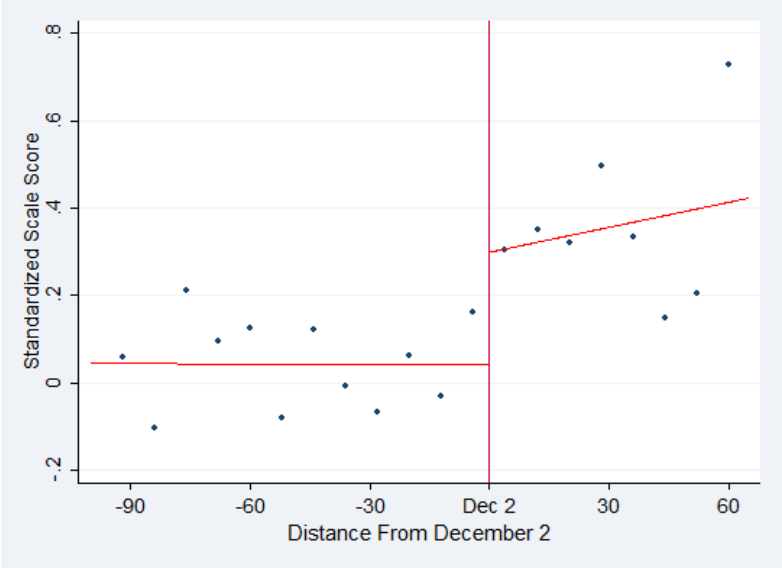
**Figure A2:** Fall kindergarten CELDT subtest outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. Birthdays are centered at December 2. CELDT stands for the California English Language Development Test.



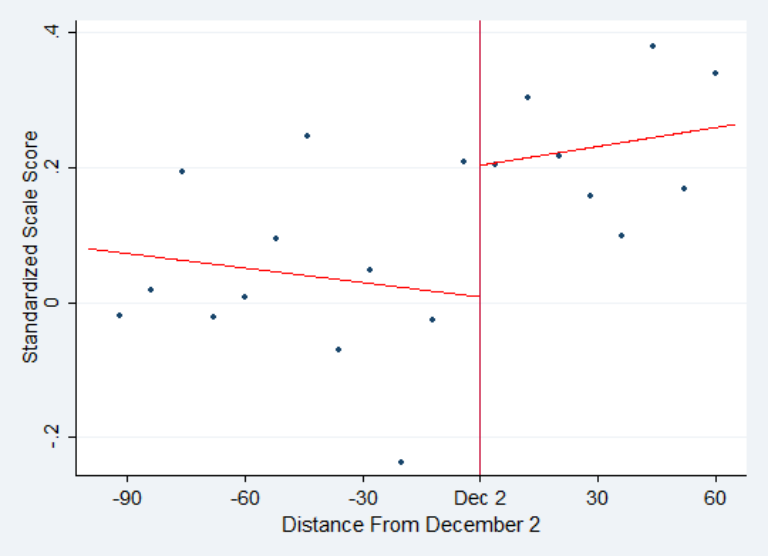

**(a) Listening**




**(b) Reading**

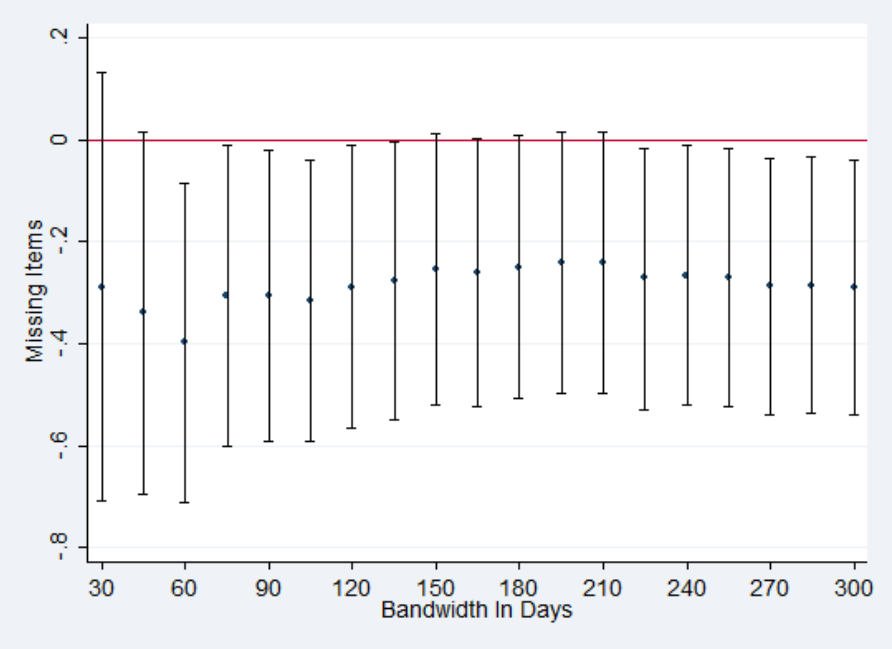



**(c) Writing**

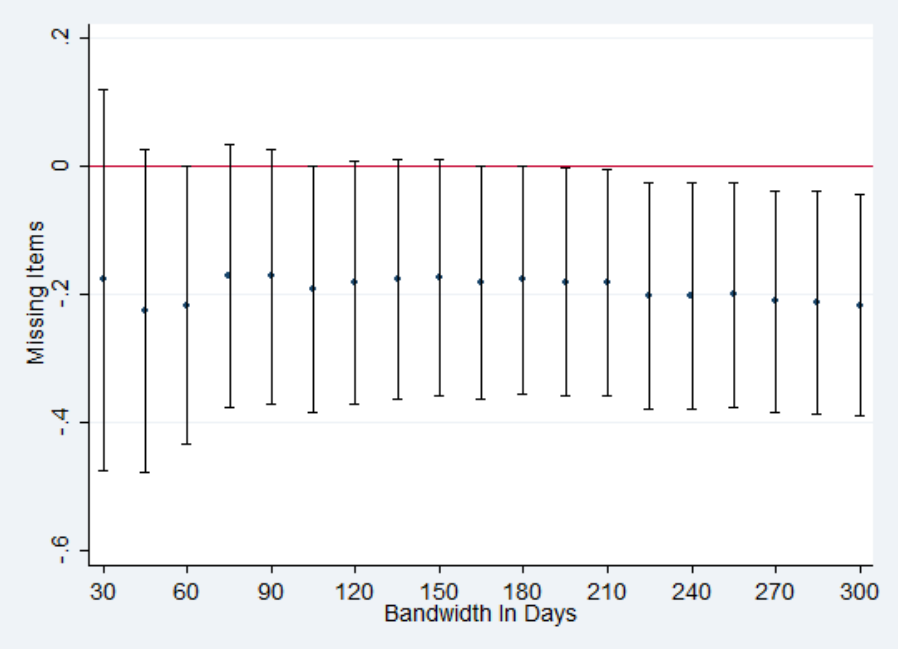



**(d) Speaking**

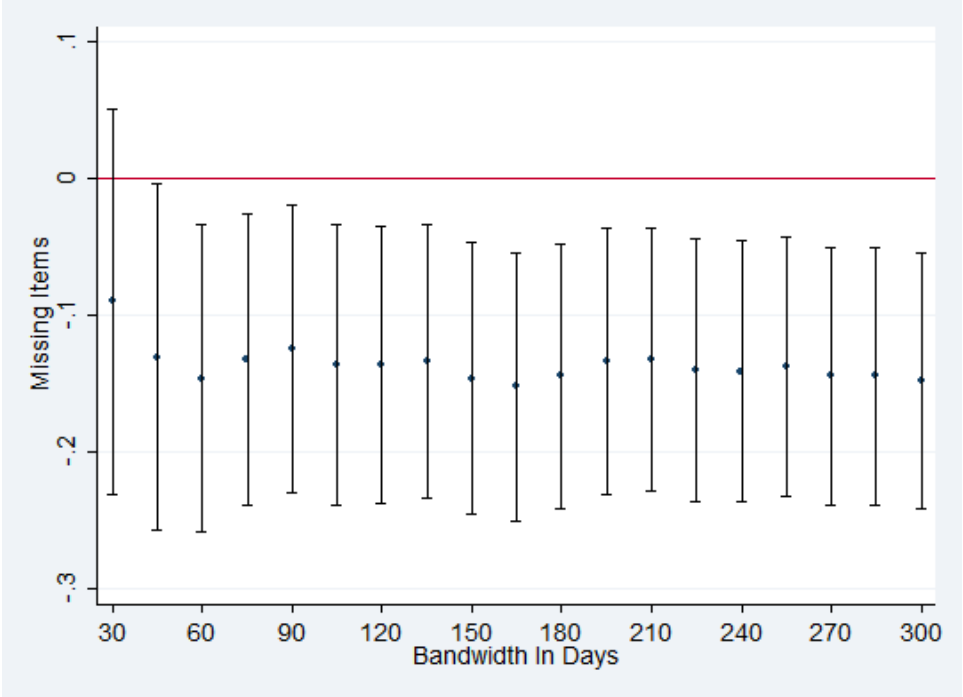
**Figure A3:** Fall first grade CELDT subtest outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. Birthdays are centered at December 2. CELDT stands for the California English Language Development Test.



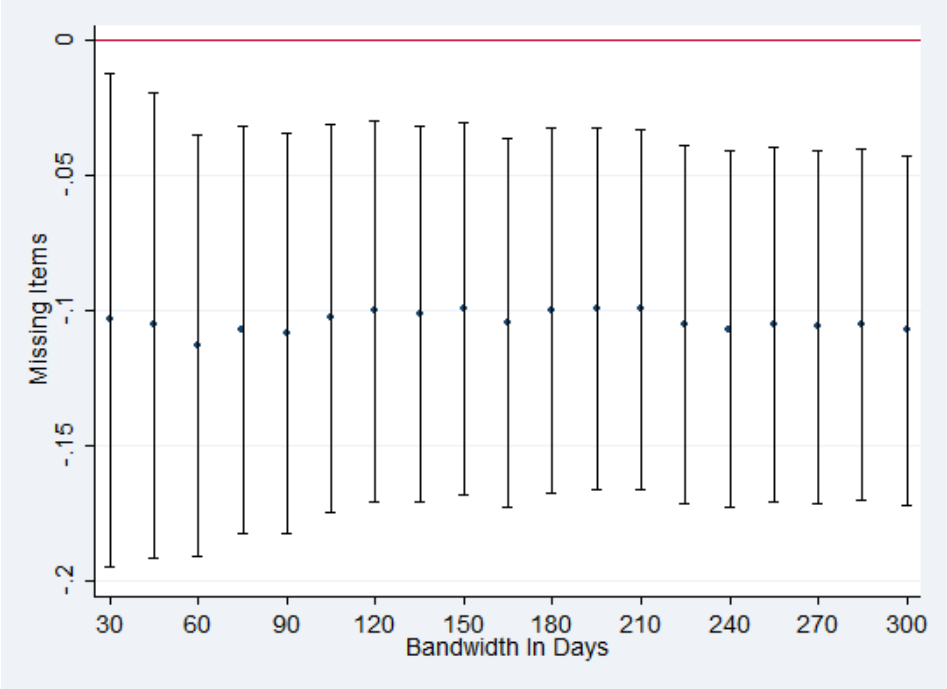

**(a) Upper Case Letter Recognition**



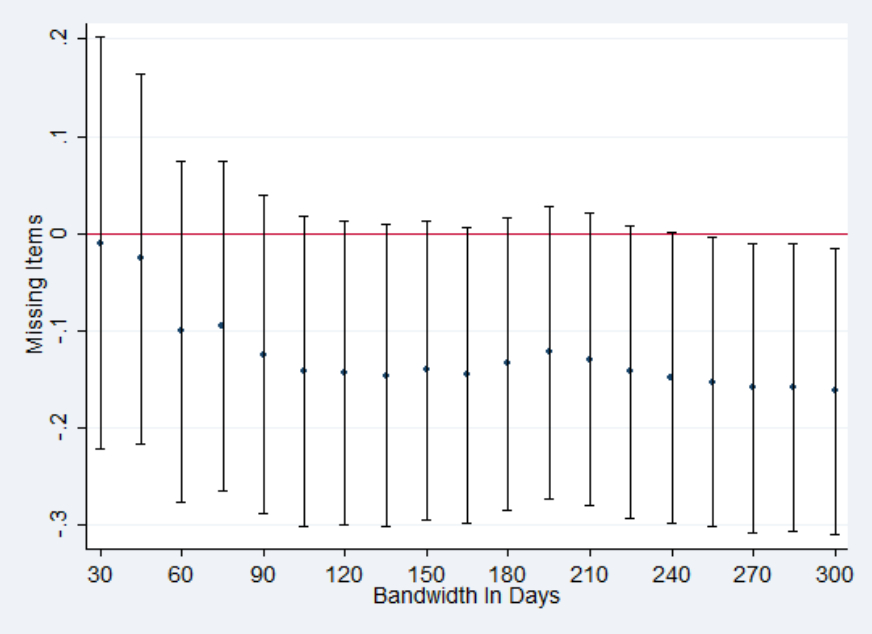

**(b) Lower Case Letter Recognition**



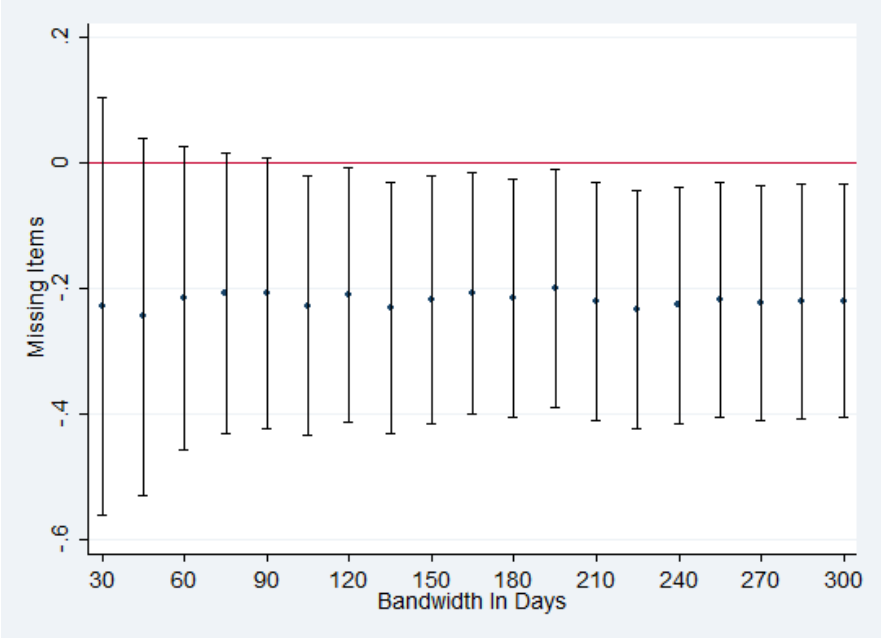

**(c) Letter Sounds**



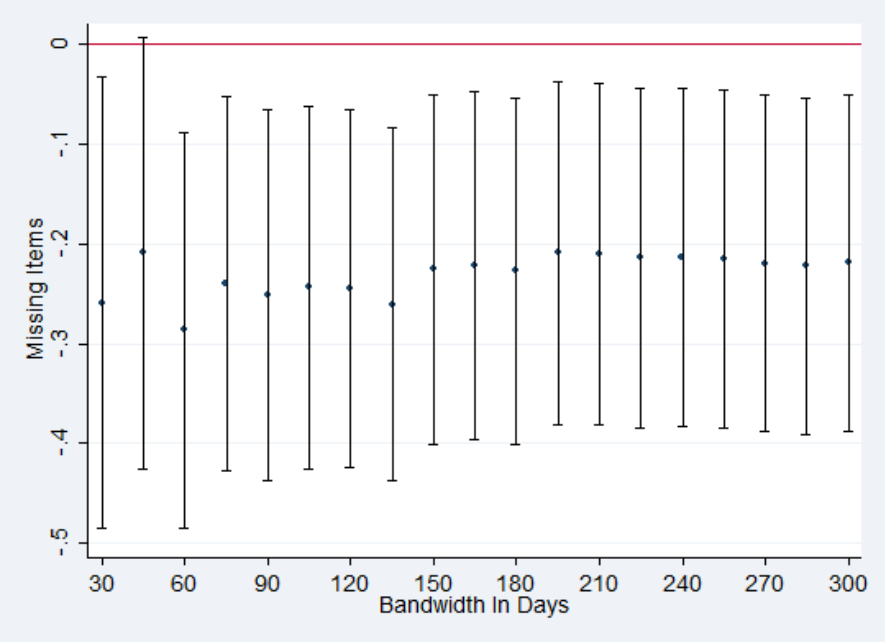

**(d) High Frequency Word Recognition**




**(e) Early Literacy Behaviors**

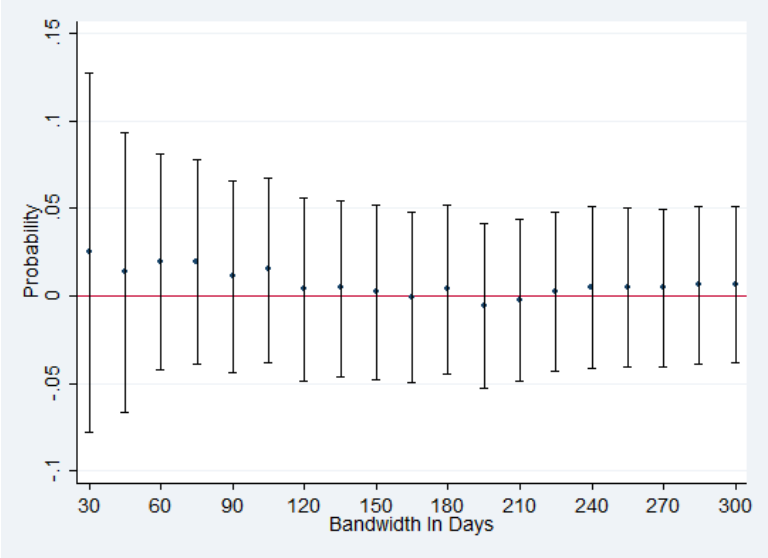



**(f) Initial Word Sounds**

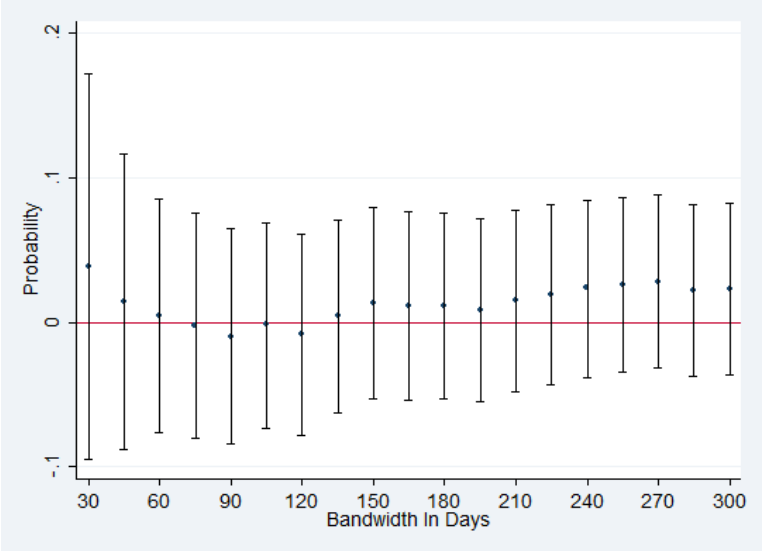



**(g) Rhyming**

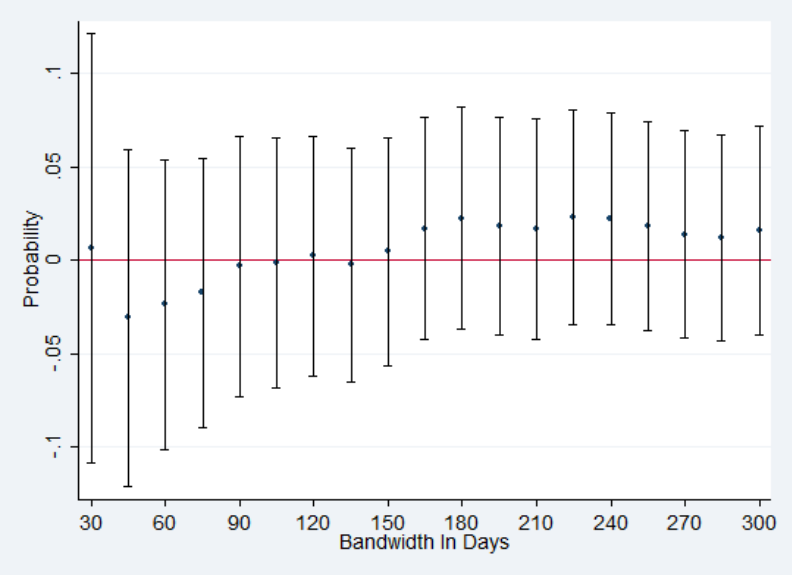
**Figure A4:** Auxiliary robustness checks of fall kindergarten Fountas and Pinnell foundational literacy outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relavant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All figures employ a negative binomial regression. Teacher-by-year fixed effects are not included because models would not converge for all banwidths. All regressions employ a linear spline functional form with covariates detailed in Table 4. Standard errors are clustered at the teacher-by-year cell.




**(a) Pr(Reading at Level C or Above) In Fall First Grade**

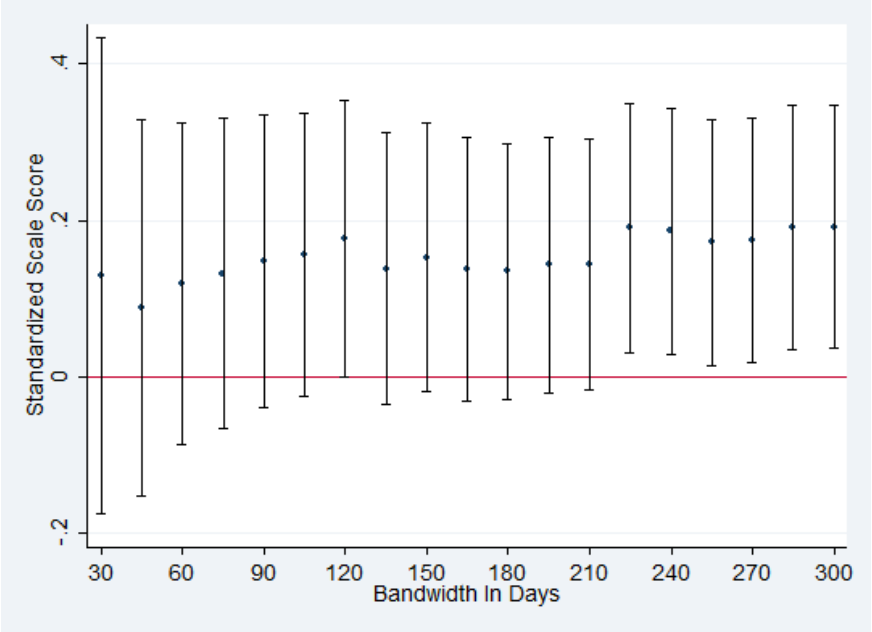



**(b) Pr(Reading at Level E or Above) In Fall First Grade**

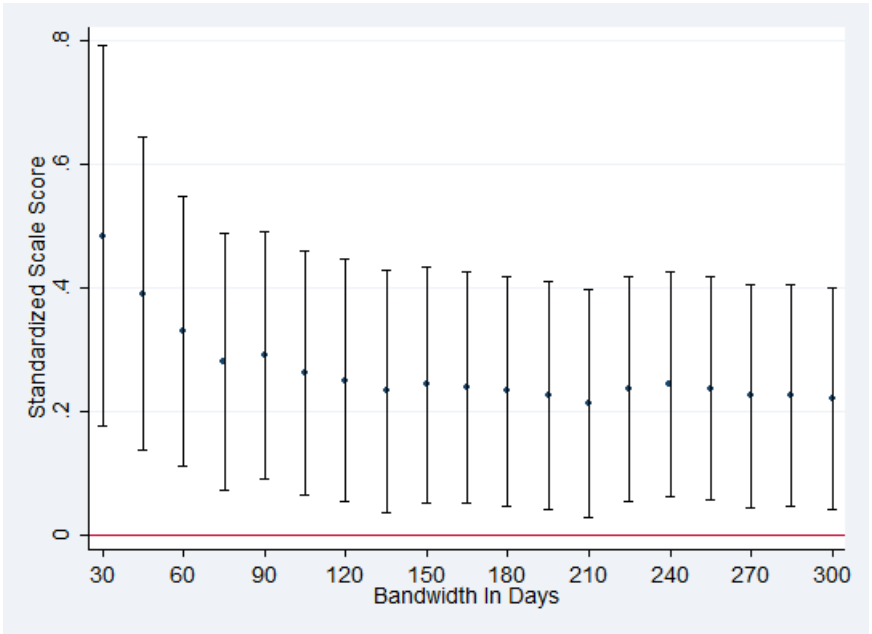



**(c) Pr(Reading at Level I or Above) In Fall First Grade**

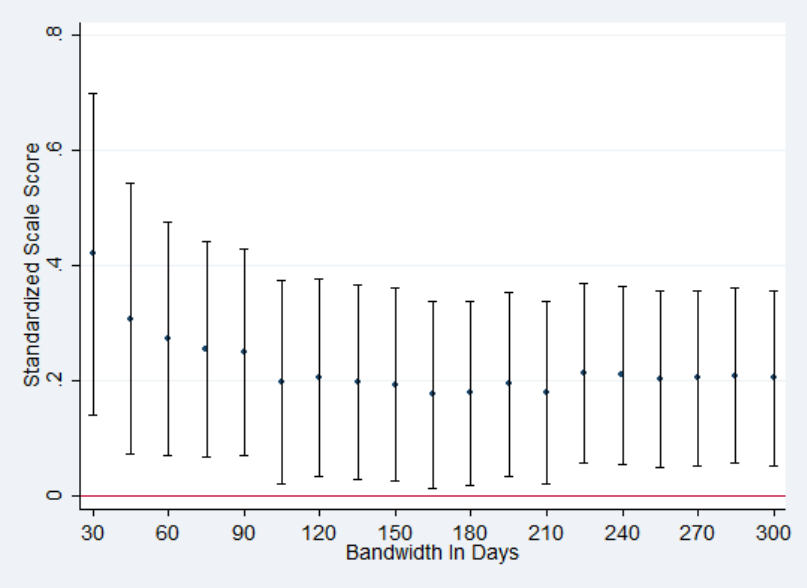
**Figure A5:** Robustness checks of fall first grade Fountas and Pinnell foundational literacy outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relavant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 4. Standard errors are clustered on the day of birth rating variable.



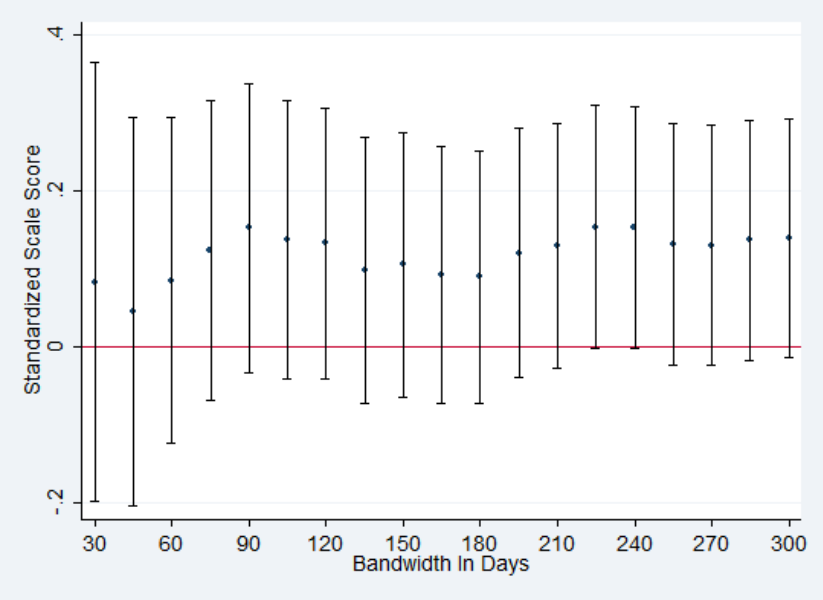

**(a) Listening**




**(b) Reading**

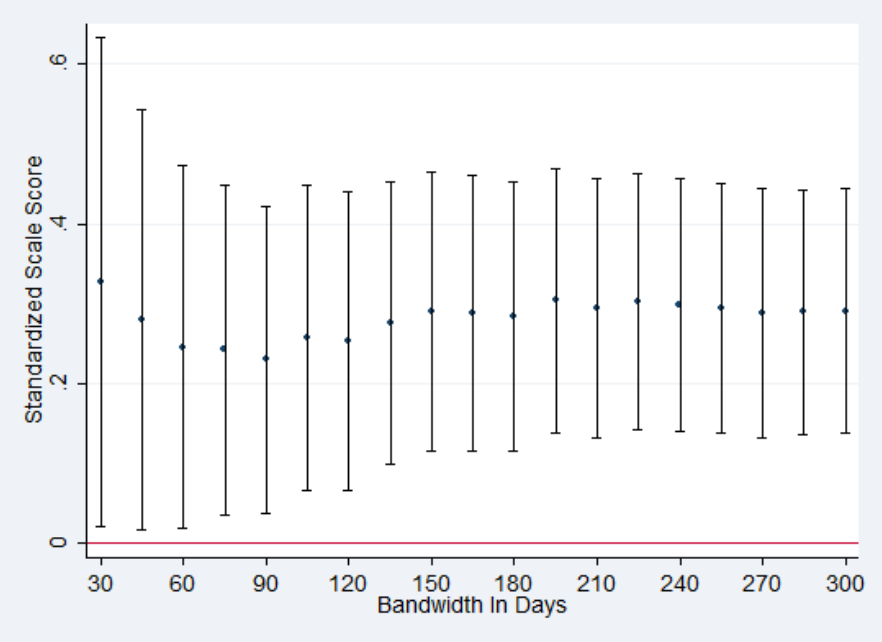



**(c) Writing**

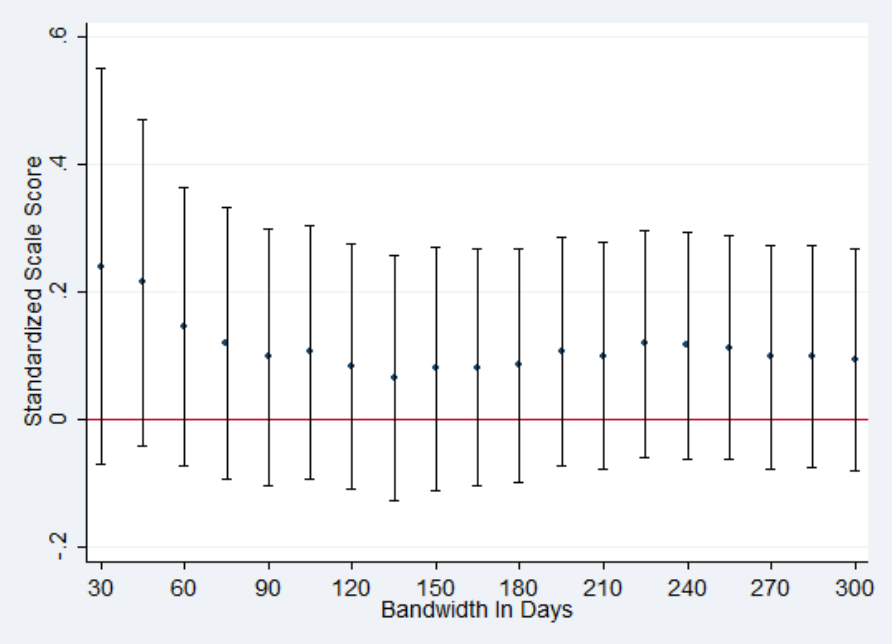



**(d) Speaking**

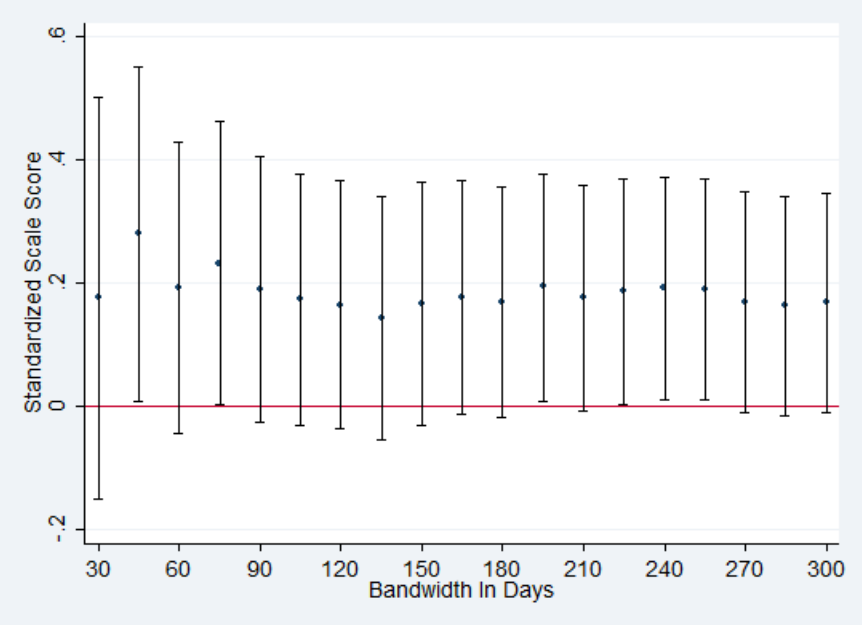
**Figure A6:** Auxiliary robustness checks of fall kindergarten CELDT subtest outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relavant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 4. Standard errors are clustered on the day of birth rating variable.



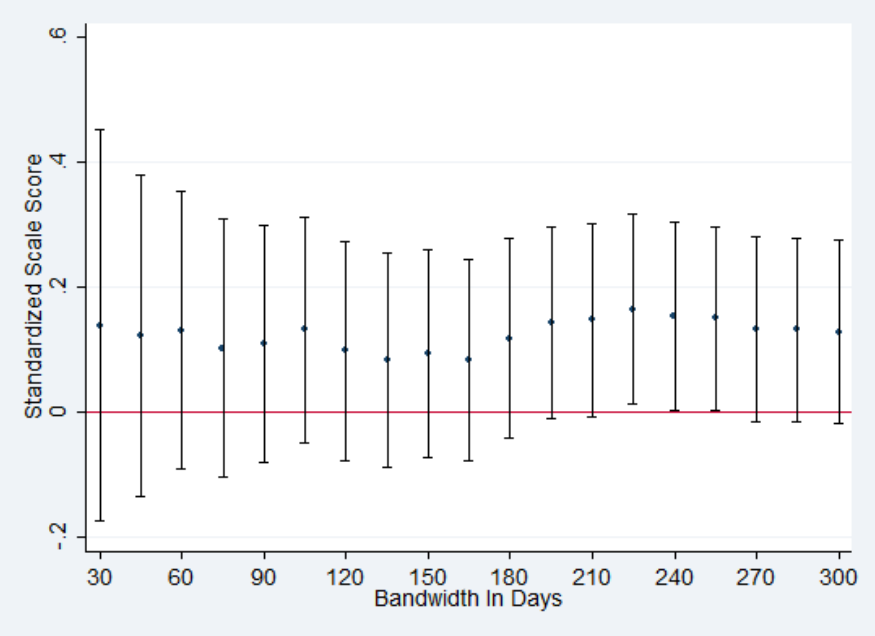

**(a) Listening**




**(b) Reading**




**(c) Writing**




**(d) Speaking**

**Figure A7:** Auxiliary robustness checks of fall first grade CELDT subtest outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relavant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 4. Standard errors are clustered on the day of birth rating variable.

Table A1: RD regressions of balance In sample restrictions

| | (1) | (3) | (5) | (5) |
|---|---|---|---|---|
| | Full Sample | $\|B_{ict}\| \leq 60$ | $\|B_{ict}\| \leq 30$ | $\|B_{ict}\| \leq 15$ |
| Missing Kindergarten Blending | 0.012 | 0.003 | 0.013 | 0.054 |
| | (0.017) | (0.020) | (0.028) | (0.037) |
| Missing Kindergarten Rhyming | -0.032 | -0.023 | 0.016 | -0.001 |
| | (0.023) | (0.028) | (0.033) | (0.044) |
| Missing First Grade Fountas and Pinnell | 0.023 | 0.035 | 0.071* | 0.033 |
| | (0.018) | (0.022) | (0.029) | (0.035) |
| Missing Kindergarten CELDT | 0.030 | 0.059 | 0.082 | -0.016 |
| | (0.038) | (0.046) | (0.064) | (0.087) |
| Missing First Grade CELDT | -0.009 | 0.021 | 0.036 | -0.036 |
| | (0.040) | (0.049) | (0.072) | (0.104) |
| N | 6,773 | 2,191 | 1,278 | 666 |

*Note:* Each cell represents the results of a separate regression discontinuity estimate on an indicator for not being in the sample defined in the row headers. Column headers indicate the bandwidth restriction. The functional form in all regressions is a linear spline. All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01

Table A2: RD regressions of initial CELDT test taking

| | (1) | (3) | (5) | (5) |
|---|---|---|---|---|
| | Full Sample | $\|B_{ict}\| \leq 60$ | $\|B_{ict}\| \leq 30$ | $\|B_{ict}\| \leq 15$ |
| Initial CELDT Examination In Kindergarten | -0.290** | -0.282** | -0.305** | -0.285** |
| | (0.031) | (0.030) | (0.033) | (0.043) |
| N | 3,334 | 1,112 | 648 | 331 |
| Initial CELDT Examination In First Grade | N/A | N/A | N/A | N/A |
| | N/A | N/A | N/A | N/A |
| N | 2,697 | 899 | 523 | 267 |

*Note:* Each cell represents the results of a separate regression discontinuity estimate on an indicator for not being in the sample defined in the row headers. Column headers indicate the bandwidth restriction. The functional form in all regressions is a linear spline. All standard errors are clustered on the day of birth running variable. In first grade, "N/A" indicates that no student in the sample took CELDT for the first time and there is therefore no variation in initial CELDT status. +indicates p<0.10, *p<0.05, **p<0.01