Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 16-07

How Much Regulation? A Fuzzy Regression Discontinuity Analysis of Student Literacy Skills in Prekindergarten vs. Transitional Kindergarten

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 the San Francisco Unified School District. TK is a highly regulated, state funded, early education program. Importantly, universal prekindergarten was already established in San Francisco, making this study a comparison of prekindergarten opportunities. This study tests whether a more highly regulated pre-K program, situated solely in schools, can provide benefits to young five year olds over a modern, robust universal pre-K market. I find that students who attended TK outperform their peers on a variety of foundational literacy skills, with some evidence that the gains are larger for minority children. TK, however, had little effect on the rate of absences in kindergarten and first grade.

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 2017

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

How Much Regulation? A Fuzzy Regression Discontinuity Analysis of Student Literacy Skills in Prekindergarten vs. Transitional Kindergarten

Christopher Doss[i]

Stanford Graduate School of Education

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 the San Francisco Unified School District. TK is a highly regulated, state funded, early education program. Importantly, universal prekindergarten was already established in San Francisco, making this study a comparison of prekindergarten opportunities. This study tests whether a more highly regulated pre-K program, situated solely in schools, can provide benefits to young five year olds over a modern, robust universal pre-K market. I find that students who attended TK outperform their peers on a variety of foundational literacy skills, with some evidence that the gains are larger for minority children. TK, however, had little effect on the rate of absences in kindergarten and first grade.

[i] 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 high quality early childhood education to young children 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 improve long term outcomes (Currie and Thomas 1995, 2000; Garces, Thomas, and Currie 2002; Gormley et al. 2005; Belfield et al. 2006; Deming 2009; Heckman et al. 2010; Puma et al. 2010; Campbell et al. 2012). The results of these and other studies have spurred states and localities to invest in prekindergarten (pre-K) programs.

With the proliferation of pre-K services available to families, the conversation has now shifted to identifying the types of programs and pedagogical approaches that are most effective for our youngest students. From a programmatic standpoint, the pre-K sector is currently marked with a dramatic variation in the quality of programs and in the qualifications, compensation, and stability of the teaching staff (Bassok et al. 2013). Low-income and minority families often enroll in less effective programs, or fewer hours of instruction, leading to weaker academic outcomes (Magnuson et al. 2004; Phillips and Lowenstein 2011). Pedagogically, researchers and practitioners are debating what level of academic instruction is appropriate for young children, with many pushing back at the increasing academic nature of early childhood education (Elkind and Whitehurst 2001; Stipek 2006; Zigler and Bishop 2006; Bassok, Latham, and Rorem 2016).

The institution of a state-mandated pre-K program in California provides an opportunity to evaluate a large early childhood education policy while speaking to these pressing issues surrounding modern pre-K programs and markets. In 2010, Governor Schwarzenegger signed the Kindergarten Readiness Act into law in California. Previously, all children who turned five on or before December 2 were eligible for kindergarten. Stakeholders were concerned that the youngest of these children were not ready for kindergarten (Governor's State Advisory Council 2013). Beginning in 2012-2013, the law gradually moved the cutoff date to September 2 and established

Transitional Kindergarten (TK) for students who turn five between September 2 and December 2. The state considers TK to be the first in a two-year kindergarten sequence whose goal is to prepare children for kindergarten (Governor's Advisory Council 2013). TK is therefore a state-mandated pre-K program for age-eligible children, though it is voluntary for families to participate.

TK distinguishes itself from other pre-K programs in that it is funded and governed in the same manner as the K-12 system, is situated solely within schools, and is completely free to families. TK is more highly regulated than typical prekindergarten programs and provides a relatively highly educated and compensated teaching force compared to pre-K programs. Further, the San Francisco Unified School District (SFUSD) created a curriculum that is a middle ground between pre-K and kindergarten, in keeping with the trend of increasing the academic focus of early childhood programs. Statewide, TK was projected to cost $675 million a year (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 SFUSD. The San Francisco context provides an opportunity to compare the more regulated and academic TK program to traditional programs in a robust pre-K market because in 2004 San Francisco established universal pre-K. A child turning five years old on December 2 can enroll in TK (or choose from any pre-K program in San Francisco), while a child turning five years old on December 3 can only enroll in pre-K programs offered in the city. Both sets of children enter kindergarten the following year. Figure 1(a) illustrates this assignment mechanism for the second cohort.

The unique eligibility requirements detailed in Figure 1(a) also provide the opportunity to address weaknesses in previous birthday RD studies of early childhood programs. Lipsey et al. (2014) argue that these weaknesses stem from the fact that previous birthday RD studies compare children from different cohorts. This cross cohort comparison may not be capturing an accurate counterfactual and may result in biased estimates if children are subject to different assessment rules.

A within cohort comparison is ideal because all children are assessed in the same way and the efficacy of a specific program can be compared with other educational opportunities available to children in the same cohort in the same year. The TK program eligibility requirements allow me to make this type of comparison. The robust nature of the San Francisco universal pre-K market also means that the alternate experiences available to children are of relatively high quality. Program effectiveness can vary significantly based on the quality of the counterfactual early childhood experiences (Shager et al. 2012; Zhai, Brooks-Gunn, and Waldfogel 2014; Feller et al. 2015), making this study especially relevant and timely.

I analyze 6,739 kindergarteners enrolled in SFUSD in the 2013-2014 and 2014-2015 school years. These classes contain the first two TK cohorts. Of the students in the sample, 946 were eligible for TK in the previous year and 335 enrolled. The primary outcomes are the fall kindergarten and fall first grade administrations of the Fountas and Pinnell Benchmark Assessment System (BAS), the California English Language Development Test (CELDT), and attendance records in these grades. The BAS measures student pre-literacy skills and reading level. The CELDT is given to all students whose families do not speak English at home and measures reading, listening, speaking, and writing. I find that, in the fall of kindergarten, former-TK students outperform their peers on both assessments. Fall first grade results show that the advantages in CELDT remain, but the advantages for students on the BAS are no longer evident. There is some evidence that the effects are largest for minority children, consistent with the notion that TK reduced the sorting of children to less effective programs. TK did not have an effect on absences, except for Asian students (about one-third of the sample) in kindergarten, who were absent 1.3 fewer days.

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 large, short to medium term effects on IQ, reading, and math scores, as well as large positive effects on other outcomes such as incarceration (Ramey and Campbell 1984; Belfield et al. 2006; Heckman et al. 2010; Campbell et al. 2012). Head Start is a quintessential example of a large, federally funded program meant to provide services to economically disadvantaged children. Though less intensive than the Perry-Preschool and Abecedarian programs, Head Start has positive effects on language, literacy, and math (Currie and Thomas 1995; Deming 2009; Puma et al. 2010).

The establishment of TK fits into a larger trend of states and localities investing in pre-K programs as a response to this encouraging evidence. Researchers often evaluate these programs by exploiting enrollment cutoff dates and a regression discontinuity design (RD) to compare children who just finished pre-K and entered kindergarten with children who just entered pre-K. Some programs, such in Oklahoma (Gormley et al. 2005) and Boston (Weiland and Yoshikawa 2013), have positive effects on a variety of cognitive and non-cognitive outcomes. Other studies, such as Wong et al.'s evaluation of pre-K programs in five states (2008) show mixed results, with some programs providing advantages and others providing no measurable advantage, depending on the outcome. A recent evaluation of Tennessee's voluntary pre-K program is similarly mixed. Lipsey et al. (2013) use oversubscription lotteries and find robust evidence of positive effects on cognitive and non-cognitive outcomes at the end of the pre-K year. These results, however, are largely gone by the end of kindergarten. In contrast, Ladd, Muschkin, and Dodge (2015) use a difference-in difference strategy to evaluate two pre-K programs in North Carolina and find more persistent positive benefits in the form of increased reading and math scores in third grade.

Recent scholarship has posited that this variation in effectiveness can be explained, in part, by variation in the counterfactual. The counterfactual can change across geographic regions and over time because of differences in the strength of early childhood education markets and their programs.

As pre-K markets expand over time, for example, more families enroll their children in center-based early childhood education programs, which tend to be of higher quality than informal care. Programs such as Head Start may seem less effective in some instances than in others if the control group is receiving more services. In support of this hypothesis, studies have found that the benefits of Head Start are concentrated on students who, in the counterfactual, do not experience center care (Shager et al. 2012; Zhai, Brooks-Gunn, and Waldfogel 2014; Feller et al. 2015). The counterfactual for many program evaluations is not clear and thus it is difficult to determine whether the differences in results are driven by differences in the quality of the target program or by differences in the experiences of the control group. In San Francisco, the comparison group to TK is clearer than in many studies because all four year olds have access to universal pre-K in the city and the vast majority of these children make use of this access.

A second source of variation in the counterfactual comes from the different methodologies used across studies. The RD evaluations of pre-K programs usually use cross-cohort comparisons. Lipsey et al.'s (2013) use of oversubscription lotteries and Ladd, Muschkin, and Dodge's (2015) difference-in-differences strategy avoid such cross-cohort comparisons. Lipsey et al. (2014) argue that the cross-cohort counterfactual contains significant weaknesses. Students in pre-K in year T (cohort 1) are compared to students who are ineligible for pre-K in year T (cohort 2). In year T+1 cohort 1 will advance to kindergarten while cohort 2 will begin pre-K. The aim of these evaluations is to estimate the effect of pre-K over the alternative child care arrangements parents would make for the same cohort. Parents of children in cohort 2 are not an accurate counterfactual because they are likely to make different arrangements knowing that their children are eligible for pre-K the next year. Furthermore, a change in the supply of pre-K programs in year T can change pre-K enrollment patterns in cohort 2 in year T+1. This change would affect the types of students observed and assessed in the control group. Cohort differences can even complicate the assessment process. Many

assessments have different start rules based on age or grade. If the two cohorts start at different points in the assessment the results may be biased.

The unique enrollment criteria of TK allows this study to address the major weakness inherent in previous RD evaluations because the TK eligibility requirements allows for comparisons of students in the same cohort. As Figure 1(a) illustrates, in year T students born on December 3 must attend pre-K while students born on December 2 have the same exact pre-K opportunities in San Francisco, but also have the option to attend TK. In year T+1 both sets of children attend kindergarten. The children are in the same cohort and enter kindergarten at the same time. All children are concurrently assessed with the same rules, in the same classrooms.

Lipsey et al. (2014) point to a second issue with the counterfactual in RD studies that continues to be a challenge for this study. Only children who enroll in SFUSD are observed and assessed. If the availability of TK affected enrollment then the comparison between TK eligible and ineligible students could be biased. Ideally one would identify the sample in the previous year and follow the students so as to ensure that attrition from, or entrance into, the sample does not bias the results. While I cannot take this approach, I have the universe of students in public kindergarten in San Francisco and compare those eligible and ineligible for TK. I leverage an extensive set of RD checks to ensure the internal validity of the study is not compromised.

While the counterfactual may drive some differences in the estimated effects of programs, the quality of the programs themselves are also likely to be a determining factor in their relative success. The school-based nature of TK, for example, may provide benefits because TK falls under the same regulations as the broader K-12 system. Salaries of teachers in TK are, as a result, meaningfully higher than the salaries of pre-K teachers, on average, as are their education requirements. Typically, pre-K programs can vary meaningfully in the stability, education, and compensation of the teachers (Bassok et al. 2013). Moreover, the TK curriculum is consistent across schools, while the curriculum across pre-K sites also can vary.

Low-income and minority families may gain the most benefits from the consistent quality of TK, at least in part because they are typically less likely to opt into formal early childhood programs and more likely to enroll in less effective programs (Magnuson et al. 2004; Phillips and Lowenstein 2011). These sorting patterns are related to academic outcomes (Lee, Loeb, and Lubeck 1998; Loeb et al. 2004; Bassok et al. 2016). Some research shows 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 an increase in the uptake of center care. Meanwhile, pre-K programs in more highly regulated markets are associated with better outcomes (Fuller et al. 2004; Rigby, Ryan, and Brooks-Gunn. 2007; Hotz and Xiao 2011).

The free nature and consistent curriculum of TK, along with the high compensation and education of the TK labor force represent a new level of regulation of a pre-K program. If the universal pre-K market provides variable quality options, some of which are lower than TK, then TK may benefit the children who enroll. If, despite the universal pre-K market in San Francisco, low-income and minority children still attend prekindergarten programs of relatively lower quality, combatting these selection effects can result in greater outcomes for these children.

The academic underpinnings of TK are also relevant to a current debate in the literature as to what an appropriate curriculum looks like for young children. Recent studies have shown that kindergarten is becoming increasingly focused on academic instruction in subjects such as reading and math (Bassok, Latham, and Rorem 2016). This trend has caused parents, researchers, and practitioners to debate whether we are asking too much of children too soon (Elkind and Whitehurst 2001; Stipek 2006; Zigler and Bishop 2006). The effects of TK, with its greater academic focus relative to more typical pre-K programs, provide further evidence on the relative merits of this focus, though other, aforementioned factors differ between these programs as well.

Transitional Kindergarten is reminiscent of past efforts to institute two-year kindergarten programs such as developmental kindergarten and transitional first grade. These programs were

often targeted to at-risk children. Meta-analyses generally conclude that they were ineffective (Ferguson 1991; Karweit and Wasik 1992). This study provides evidence on the efficacy of a modern version of this type of program. TK may yield different results given the academicization of the earlier grades and the availability of the program to the all students, not just at-risk students.

Finally, this study is similar in design and focus to an independent study that was concurrently fielded by a contractor and that looked at TK statewide (Manship, K. et al. 2015). The results of their unpublished report are broadly similar to the ones here. This study distinguishes itself from their report in a few ways. The authors sampled districts throughout the state while I use population data for a single diverse urban area. This area, SFUSD, was not included in the report sample. By focusing on the population of students, I have one, geographically consistent counterfactual pre-K condition. Given the great variation in counterfactual pre-K experiences seen in the literature, and their effects on estimates, this makes interpretation of results cleaner. The counterfactual is especially relevant when looking at subgroups because subgroups are likely sorted to different geographical areas with different TK programs and counterfactual pre-K experiences. Having a defined population off which to judge heterogeneity will greatly help in determining if results are larger for minority students, which is consistent with notion that TK mitigated the sorting of low income and minority students to less effective pre-K programs. Further, the report does not include heterogeneity analysis.

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. The pre-K market, thus, is regulated to an extent that is not typical in the country. There is evidence that San Francisco's efforts have created a robust pre-K market that offers high quality programs. Applied Survey Research (2013) leveraged a regression discontinuity design to evaluate the umbrella

organization's programs. They found that the program produced a three-month gain in letter and word recognition, a three- to four-month gain in problem solving and gains in self-regulation.

This type of regulation is likely to establish a floor with regard to the quality of services provided to children in the city. Even in this regime the opportunity for sorting of children to settings remains. City providers must be licensed by the state; however, providers range from school-based programs, to Head Start, to home-based care. The teachers they employ must have 24 early childhood or child development credits and 16 general education credits, but providers can employ more highly educated teachers. Additionally, there is no minimum compensation for teachers. Programs can attract teachers of varying quality, partially through compensation.

Between 2013 and 2015, 142 of the current 147 programs in the universal pre-K market volunteered to be rated with the Quality Rating and Improvement System (QRIS). QRIS is an increasingly common tool used to measure the quality of pre-K services. Table 1 presents the average QRIS scores for SFUSD pre-K centers, Head Start centers, other center-based care, and home-based care.[1] Though, on the whole, programs are rated relatively highly, there are differences in quality across the pre-K sector with the overall rating ranging from 3.35 to 4.1 stars (of 5 stars). This variation may be smaller than expected. Home-based programs, which typically produce weaker outcomes, were rated an average of 3.69 stars.

Despite the strength of the pre-K programs, variation remains among programs within a sector and in the components of care provided among sectors. Head Start has a comparative advantage in providing health screenings, teacher qualifications, and child interactions. SFUSD centers have a comparative advantage in director qualifications, child/teacher ratios, and program environment. The remaining variation in the market leaves the door open to the sorting of families to programs. The city also provides funding for only 612.5 hours of instruction spread through 175 to

[1] Averages were calculated by the author. Source data is from First Five, 2015.

245 days. This amounts to 3.5 to 2.5 hour school days. The organization does not subsidize more time, meaning disadvantaged families may select into fewer hours of instruction.

The highly regulated nature of TK can mitigate many of these lingering selection effects. TK is strictly school-based, eliminating the variation in types of programs offered to families. The state requires teachers to hold a bachelor's degree and the same credentials as other elementary school teachers. The district also compensates TK teachers at the same rate as other teachers. This approach raises the floor of, and reduces the variation in, provider qualifications, education, and compensation. TK is also open to all residents of the city and is a completely free, full day program. The quality of TK classrooms across the city likely still varies and selection to these classrooms may be correlated with demographic and economic variables. However, on the balance, these selection effects are likely muted in comparison to the larger pre-K market.

TK further distinguishes itself from pre-K by 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 *prekindergarten* program, which is part the universal pre-K system, and SFUSD's TK program. Table 1 shows that in comparison to other center-based care, SFUSD performs about as well on almost all dimensions of QRIS. SFUSD's pre-K curriculum is therefore likely to approximate of the types of instruction the vast majority of students receive in the universal pre-K system.

Figure 1(b) 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 six hours. The district uses a homegrown TK curriculum designed to be the middle ground between their pre-K and kindergarten curriculums. District officials emphasized literacy skills and socio-emotional skills and began to emphasize math skills. In many

ways, pre-K represents a student centered and play-based approach while TK represents an academic and structured approach. In pre-K, students are allowed to guide the activities and instruction, no curriculum map or timeline exists, and students are given ample naptime and outdoor time. In TK, naptime is eliminated, outdoor time is limited, and teachers, who stay on a curriculum map and timeline, guide the activities. In both programs each session of whole group instruction lasts no more than 10 minutes, but TK utilizes it more often.

TK also differs from pre-K in the composition of the classroom. TK classrooms contain students of a relatively small age range, which may make it easier for teachers to target their instruction to children at a similar developmental level. This advantage is moderated by the fact that there are fewer adults in the room. Qualified pre-K programs must have a maximum class size of 24 and a child-adult ratio of 8:1. In contrast 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.

3. Data

This study examines the first two cohorts of TK students in SFUSD. 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 between October 2 and December 2 were eligible. Enrollment into TK was not mandatory, and families also had all other pre-K opportunities in San Francisco available to them. Children born after December 2 were eligible for the same pre-K opportunities in San Francisco, less TK. Children born before November 2 (or October 2 in year two) enrolled in kindergarten and are not in the study.[2] The structure of the program means that a plausibly exogenous cut point, based solely on birthdate, dictates different

[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. From a policy standpoint this contrast is less relevant because TK is meant not meant to replace kindergarten, but to better prepare students for kindergarten. From a methodological standpoint I found significant covariate imbalance across this threshold, undermining the causal warrant of this approach.

educational experiences for children. Children born around the cutoff should, on average, be similar except for the probability of enrolling in TK. A FRD design can leverage this cut point to estimate the effect of TK on outcomes. 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 repeat the process with pre-K rosters to identify students who attended pre-K in the district.

The district uses the Fountas and Pinnell Benchmark Assessment System (BAS) to measure literacy skills of every student in TK to third grade. In the fall, all teachers are required to assess their children on foundational skills. In 2013-2014, these skills were: upper- and lower-case letters, 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 the ten skills they read books. Students started with the easiest books (level A) and after reading with enough accuracy and comprehension they progressed to harder books (levels B-Z).

In 2014-2015, the district 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 enough skills to move on to the leveled reading assessment, and the probability of reading at least at level A. The test could be administered in either English or Spanish. My main specification includes controls for test language. By first grade almost all children (98 percent) were assessed on their ability to read. The fall first grade results are whether TK students are reading more advanced books.

The BAS has been shown to be a valid assessment of literacy development in children (Fountas and Pinnell 2012). In addition, many of the foundational skills are common in early childhood assessments and are predictive of future literacy skills. For example, letter knowledge and

phonological awareness have been linked to later decoding skills and reading comprehension, while letter sounds and sight word knowledge have been identified as critical to making the transition to reading (National Literacy Panel 2008; Kjeldsen et al. 2014; Ehri 2015).

Because almost half the students in the district are English Language Learners (ELLs), I assess the effects of TK on the performance of ELL students on the CELDT. Students are identified as ELL if the family speaks a language other than English in the home. Any student identified as ELL is required to take the CELDT the first year they enter the district and every year until they are reclassified as English proficient. The results of the CELDT are consequential for these students because reclassification as English proficient depends, in part, on their test scores.

The CELDT was created and validated by the California Department of Education in conjunction with testing experts and is designed to measure the English development of students whose first language is not English (California Department of Education 2014). Students are assessed in listening, speaking, reading, and writing. The listening section tests students' ability to follow directions and comprehend oral stories. The speaking section tests students on oral vocabulary, speech, the ability to construct stories from pictures, and the ability to communicate reasoning skills. The reading section tests similar skills as the BAS including identifying letter sounds, pictures associated with words, and parts of a book. In the writing section, students 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 outcomes are expressed in traditional scale scores, which lends itself to a traditional interpretation of the estimates. Finally, because 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 kindergarten) 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 ELL 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 were exposed to the assessments, thereby eliminating any practice effects.

Finally, I analyze the number of absences in kindergarten and first grade.[3] Evaluations of state-funded prekindergarten programs have found a positive association between enrollment in pre-K programs and attendance in kindergarten (Gilliam and Zigler 2004; Huang et al. 2012). This effect of more formal care on attendance may be especially salient in this context because folding pre-K programs into the school and modelling them after kindergarten programs may help parents and students better acclimate to the school environment and an academic schedule. For example, in TK, parents and students have the experience of arriving to school on time every morning and students are expected to perform for an entire school day. Thus, attending TK might increase student attendance in kindergarten and first grade since the students are more used to school. However, if students react negatively to the more structured TK environment, their engagement in school might suffer, reducing attendance in kindergarten and first grade.

Across the two years 8,717 kindergarten students matched to the fall kindergarten administrations of the BAS. Teachers varied in the extent to which they followed district assessment guidelines in administering the BAS. Many students were missing individual skills scores and some teachers assessed the child's reading level if they were close to mastering the required number of skills. The final analytical sample consists of 6,739 out of the original 8,717 students. These students had scores for all skills except rhyming and blending. The missing data was largest for those two

[3] I also analyze the effect of the program on retention. Very few students are retained in kindergarten and first grade. There is no effect on the program on retention for the entire sample and all subgroups. For brevity I do not present these results, though they are available on request.

domains and the sample sizes are smaller. If the missing data is not the same for students around the birthday threshold, comparisons of outcomes may be biased. Table A1 shows that missing scores are not related to the birthday threshold, making bias unlikely.[4]

Of the 6,739 students in the analytical sample, 3,310 are ELLs and were tested with the CELDT in the fall of kindergarten, 6,219 continued to first grade and were assessed in the fall with the BAS, and 2,663 ELL students progressed to first grade and were assessed. Again the results for the ELL and first grade samples would be biased if the probability of being in those samples is discontinuous across the threshold. Table A1 indicates that this is not the case.

Table 2 presents the descriptive statistics for the analytical sample, former TK students, and students who did not attend TK. The students are mostly Asian (31.1 percent) and Hispanic (25.0 percent), with fewer whites (16.5 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 7.6 percent of the sample, while 49.1 percent has been classified as ELL. Compared to the former pre-K students, former TK students differ in important ways. Due to the eligibility criteria, they are older. TK students were also more likely to be minority and ELL and less likely to be special education. Overall TK students, on average, significantly outperformed non-TK students on all assessments, but there is no significant difference in absences.

Twenty two percent of the sample was enrolled in the district in the prior year, 16.9 percent attended SFUSD pre-K, and 5 percent attended TK. Most other students attended another universal pre-K program. Table 1 indicates that the vast majority of programs in the pre-K market are center-based. SFUSD centers compose 22 percent of that sample (containing 142 of 147 programs), Head Start centers compose 12 percent, and the remaining 57 percent are other center-based care. With only 9 percent of programs in the home, the vast majority the students in who did not attend pre-K or TK in SFUSD likely experienced some sort of center care.

[4] Furthermore, results are robust to including all students in the sample.

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. For example, children develop quickly in this age range and TK students may have higher academic outcomes and better attendance simply because they are older. A FRD eliminates this bias by estimating differences in outcomes between TK-eligible and ineligible students near the December 2 cutoff. Near the cutoff students are of similar age and, in aggregate, there should be no differences in the distribution of background characteristics among students. Any differences in outcomes can then be attributed to differences in TK eligibility.

One challenge in working with the BAS foundational skills and attendance data is the left skewed nature of the distributions. In the fall kindergarten assessment 6.5 percent to 48.5 percent of the sample achieved the highest score on the foundational skills. The distribution of attendance is similarly skewed with about 7 percent of students having zero absences. The non-normal distribution of the outcomes make OLS inappropriate.[5] I therefore recode each skill so that I have the number of items a student missed or how many days a student was absent, and treat each variable as a count variable. I can then 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 negative binomial models.[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 levels C, E, and I or above. I choose these levels because they

[5] All inferences are consistent when using OLS models. Results available on request.

[6] In choosing from among the models I follow 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 models.

represent approximately the 20th, 50th, and 80th percentiles of the 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 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, δ_{at}. The rating variable, B_{ict}, is the distance, in days, a child is born from December 2. Following Lee and Lemieux (2010), I cluster standard errors on the rating variable because it may be considered a coarse rating variable. The coefficient of interest is β_1, the TK eligibility requirement compliance rate.

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. γ_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 3 and an indicator for kindergarten year. For the BAS outcome, the assessor-by-year fixed effect accounts for differences among teachers in how they assess their students in a given year. I cannot identify CELDT assessors, but one to three assessors were deployed to a school depending on its size. δ_{at} in these cases are school-by-year fixed effects. Finally, I use Akaike's Information Criterion (AIC) to determine the optimal functional form of *f* (Schochet et al. 2010). The test indicates a linear spline, which allows the slope to differ across the discontinuity, is optimal. As a robustness check I present results from many bandwidths and results are robust to quadratic specifications.

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, students near the threshold are, on average, similar. Any attempt to sort children to either side of the threshold undermines this identification strategy. The first two cohorts of TK students were born 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. Figures 2(a) and (b) present visual depictions of the distribution of observations around the threshold. Figure 2(a) shows that there could be a drop in observations in crossing the threshold, however, fluctuations exist throughout the range of the rating variable. I follow McCrary (2008) and test whether a change in the density of observations around the threshold is significant. Figure 2(b) presents the graphical results. I cannot reject the null hypothesis that there is no change in density at the threshold. The point estimate and standard error of the density discontinuity is 0.110 (0.089).[7]

These natural fluctuations are indicative of regular heaping often found in birthday rating variables. Recent work by Barreca, Lindo, and Waddell (2015) shows that heaping can cause bias in RD estimates if observations in the heaps are different from other observations. To test for bias they recommend estimating the effects on heaped and non-heaped data separately. As shown in the histogram in Figure 2(a), 15 to 32 students are concentrated on some values of the rating variable. In Section 7 I test for bias by eliminating observations in values of the rating variable that contain 15 or more students. The results are robust to eliminating 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

[7] To further ensure that the density of observations is continuous across the threshold, I perform the McCrary density test on each baseline covariate. Table A2 shows that the density of observations is continuous for virtually all covariates. Only one is marginally significant, which may occur by chance.

partially test this assumption by running RD regressions to see if the covariates are discontinuous around the threshold. Table 3 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 balanced across the threshold. No covariate is consistently unbalanced across all the bandwidths tested.

To be a valid FRD the December 2 threshold must predict a strong treatment contrast. Figure 3 presents the first stage results graphically. Virtually nobody who was TK-ineligible enrolled in TK. Only one child, born on December 3, enrolled into the program in the two years of the study. For those children born before December 2, the probability of enrollment increases considerably. Table 4 presents estimates of the compliance rate for the full sample, and the sample in bandwidths of 60 and 30 days. 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 4 graphically presents the main fall kindergarten BAS results. After aggregating all foundational skills together, the number of items missed drops as one crosses the December 2 threshold. Figure 4(a) indicates that TK-eligible students missed about 8 items less than their peers, or a 14 percent decrease from a base of about 56 items missed by TK-ineligible students at the threshold. For the individual skills, improvements are evident for upper- and lower-case letters, letter sounds, high frequency words, early literacy behaviors, and rhyming. Figure A1 in the appendix illustrates these results. The probability of mastering enough skills to be assessed in reading and the probability of reading at level A or above also jumps at the threshold. For ELL students, Figure 4(d) shows a jump in the overall CELDT performance. Similar jumps are evident for each subtest of CELDT – listening, reading, and writing – as shown in Figure A2. Finally, Figure 4(e) shows no significant discontinuity in the number of days absent when crossing the threshold.

The picture changes somewhat by the fall of first grade. Figure 5 shows the advantage seen in foundational skills does not translate to the ability to read more advanced books. There are small,

but insignificant, jumps in the probability of reading at levels C, E, and I or above. However, the advantages in CELDT remain and former-TK students still outperform their peers. Similarly, there is no significant discontinuity in absences in first grade.

Table 5 presents the results from the statistical models. For brevity I report the effects on the main outcomes. Table A3 contains the estimates for the subsections of the BAS and CELDT assessments. I report the coefficients for the unconditional FRD results, as well as results from my preferred specification that includes 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.[8] Columns 1 and 2 of panel A show that there is a significant effect on the number of items missed in the fall kindergarten administration of the BAS, with TK-eligible students getting fewer items incorrect. Table A3 shows that this improvement was seen in all foundational skills. TK-eligible students, however, were equally as likely to move on to the leveled reading portion of the assessment, and equally as likely to read at level A or beyond.

The coefficients on the negative binomial models may be difficult to interpret. Table A4 presents incidence rate ratios versions of the coefficients for the overall number of items missed and for the number of items missed in each foundational skill. These estimates are obtained by taking the inverse natural log of the coefficient (e^{γ_1}). Incidence rate ratios indicate the rate at which TK-eligible students, on average, miss an outcome compared to TK-ineligible students. TK-eligible students were less likely to miss foundational skills by factors of about 0.91 to 0.72. This translates to a nine percent to 28 percent decrease in items missed, respectively. To make these results more meaningful I calculate the 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

[8] 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.

average TK students missed nine fewer items, knew about two more upper-case letters and letter sounds, and knew one more lower-case letter. They could also recognize about two more words out of 25. TK students performed better by about half of a point out of ten on the remaining skills. With a 33 percent compliance rate, the treatment-on-the-treated estimates will be about three times as big.

Turning our attention to the performance of ELL students in kindergarten, columns 1 and 2 of panel A in Table 5 indicate that overall students performed 0.176 standard deviations (SD) better on the CELDT exam ($p<0.05$). Table A3 indicates that all subtests, except speaking, were significantly better and estimates range from 0.132 SD to 0.221 SD. Overall, the CELDT results reinforce the BAS results, with TK students outperforming their peers on literacy outcomes.

Because TK students entered the district a year earlier and were exposed to the tests, some of the gains could be from practice instead of from a more effective learning environment. The first grade CELDT outcomes seen in columns 3 and 4 of panel B in Table 5 indicate that practice is not likely biasing the results. At this point all ELL students have been assessed at least once and the results remain similar. ELL students still outperform their peers by 0.231SD ($p<0.01$) Table A3 shows that estimate for the listening section is significant at the one percent level. The speaking and writing estimates are significant at the ten percent level.

The results differ for the first grade results of the BAS. Column 3 and 4 of panel B of Table 5 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. There is robust evidence that TK students scored higher on pre-literacy skills in kindergarten than they would of if they had not attended TK, but there is no evidence that TK increased children's reading ability as measured by the BAS.

Turning our attention to the non-academic outcome, Table 5 indicates that, in the full sample, TK did not affect kindergarten or first grade attendance. In each case the point estimates are

quantitatively small and insignificant. For all students, there are no measurable attendance benefits to exposing parents and children to a full-day academic program in the prior year.

6. Heterogeneity of Results

Aggregate results can be hiding heterogeneity based on gender, ethnicity and English proficiency status. Despite the regulation of the universal pre-K market, sorting of families to programs of varying quality may remain. TK can mitigate these trends because it is free and decreases variation in credentials, compensation, and the curriculum offered. In this regime low-income and traditionally underserved minority students may 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, ELL, and English proficient subgroups. Looking at the total items missed, all subgroups of TK-eligible children, except for the white and other subgroups, score higher in the kindergarten administration of the BAS. There is some indication that the Asian subgroup of TK-eligible students benefitted the most, with the most negative coefficient on the negative binomial model of -0.381 (or missing 32 percent less items). However I cannot reject the null hypothesis that all coefficients on the four racial subgroups are equal ($\chi_3^2 = 5.54$, p<0.1364). Looking at the probability of mastering the requisite number of foundational skills, only male and Asian TK-eligible students were more likely to move onto the leveled reading assessments in kindergarten. Males were 4.7 percentage points more likely move onto the leveled reading assessment if they attended TK (p<0.10) and Asian students were 12.6 percentage points more likely to do so (p<0.01). TK-eligible white students were actually less likely to move onto the leveled reading assessments by 11.6 percentage points. 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). Little heterogeneity is found in the fall first grade BAS results. Here, no subgroup is reading at a higher level.[9]

[9] Table A5 shows that in the fall of kindergarten, males were also more likely to read at levels A or above. In the fall of first grade the linear probability models show little heterogeneity in reading at levels C, E, or I and above.

TK also had no robust effect on absences in all cases except for the Asian subgroup in kindergarten. In this case, former TK-eligible students were significantly less absent than their non-eligible counterparts. The coefficient on the negative binomial model in column (1) translates to an intent-to-treat estimate of 1.3 days fewer days absent. In first grade, however, the coefficients become half as large and insignificant. This result is consistent with the notion that TK may have been particularly helpful in acclimating these students to a full-day, academic environment. By first grade, however, this advantage would disappear after all students were exposed to a similar environment throughout kindergarten.

Table 7 presents subgroup results for the CELDT assessment. The white and other subgroup results are not reported due to small sample sizes. Column 1 presents the kindergarten results where Hispanic TK-eligible students particularly benefit by 0.356SD ($p<0.05$) and female TK-eligible students outperform their female peers by 0.241SD ($p<0.05$). The point estimates on the male and Asian subgroups are also positive and large, but the smaller sample makes it harder to detect a significant effect. I cannot reject the null hypothesis that the male and female effects are equal ($\chi_1^2 = 0.42, p < 0.5181$), nor that the effects on the Asian and Hispanic subgroups are equal ($\chi_1^2 = 1.81, p < 0.1780$). Column 2 of Table 8 indicates that in the fall of first grade the TK advantage for females remains at 0.199SD, though the slightly smaller point estimate results in a 10 percent significance level. The TK effect for Hispanics is now half as large and insignificant, and TK-eligible students in the Asian subgroup now have a 0.279SD ($p<0.01$) advantage. TK point estimates for the male and Hispanic subgroups are again relatively large, but imprecisely estimated due to sample sizes. I cannot reject the null hypothesis that the male and female TK effects are equal ($\chi_1^2 = 0.16\ p < 0.6903$), nor that the Asian and Hispanic TK effects are equal ($\chi_1^2 = 0.39, p < 0.5340$).

Taken together the data indicate that TK increased the pre-literacy skills of most subgroups, though this increase did not translate to a higher observed reading level in first grade. There is some evidence that the Asian subgroup benefitted the most on the BAS and kindergarten attendance, while

the white subgroup benefitted the least on the BAS. The CELDT and BAS results reinforce each other with the Hispanic and Asian subgroups experiencing advantages on both assessments. In SFUSD the Asian subgroup is a socio-economically diverse community with many immigrants and first generation Americans. These results are consistent with the notion that the regulation associated with TK attenuates selection effects that disadvantage traditionally underserved students.[10]

7. Robustness Checks

The results thus far employ the full set of data. While utilizing the full data maximizes precision, it relies 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 6 presents these robustness checks for the main outcomes. Figures A4 through A7 in the appendix present robustness checks for all other results. Each plot presents ITT estimates and their 95 percent confidence intervals for bandwidths from 30 days to 300 days. Figure 6 presents results of the total number of items missed in kindergarten as well as the overall CELDT scores in kindergarten and first grade. The point estimates are largely stable for all bandwidths, though the significance tends to decrease as the bandwidths get shorter and sample sizes decrease.

As a second robustness check, I run a series of placebo regression discontinuities. The effects previously seen should occur uniquely at the December 2 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 December 2. Table A6 presents the results of this exercise for the total items missed in kindergarten and the overall CELDT results in both grades. The results from the original estimates, found in column 4, disappear in these placebo specifications.

[10] The larger estimates for minority subgroups could occur if those subgroups were more likely to take up the program. Table A8 presents first stage estimates for each subgroup. The Hispanic and white populations enrolled in TK at rates almost identical to the full sample. The ELL and Asian subgroups enrolled at slightly higher rates. The 4 to 5 percentage point increase in the first stage, however, does not completely account for the larger effects.

The last robustness check builds off by recent work by Barreca, Lindo, and Waddell (2015) who find that heaping can cause biased 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 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 larger than the sample average of 18.5 students born in a day. I re-estimate my main results on portions of the data that exclude successively smaller heaps. In Table A7 I present estimates 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. Point estimates are noticeably larger after heaps containing more than 18 or 15 students are eliminated, but less than half the sample remains. Even in these most restrictive situations the study's inferences remain: there are significant gains for TK-eligible students.

8. Discussion and Policy Implications

This paper presents evidence that Transitional Kindergarten produces large gains in pre-literacy skills as measured by the BAS and CELDT in kindergarten in students when compared to pre-K programs available to families as part of the San Francisco's universal pre-K program. The positive effects on CELDT performance are evident in first grade as well, though the literacy measure for the full population does not show differential performance in first grade.

Despite the causal nature of the study, one issue complicates the inference. The district uses the BAS as a formative assessment tool in TK. If other pre-K programs in the city did not use the assessment, TK students were exposed to the BAS up to three times more in the year prior to kindergarten than their comparison group. Similarly, TK ELL students were exposed to the CELDT a year before non-TK ELL students. The differential fall results, then, may be the result of practice

with the test in addition to improved educational opportunities. The first grade CELDT results indicate that this practice effect is not likely an issue, at least not for ELLs. When taking the CELDT in first grade all ELL students had practice with the assessment in the prior year, kindergarten, yet the TK CELDT advantage remains evident. Nonetheless, for the broader population, the pre-literacy advantages for TK students in kindergarten were no longer evident on the reading assessment given in first grade. This lack of effect could be due to unsustained gains for participating students or to the nature of the first grade assessment.

TK differs in a number of ways from the pre-K offerings available to the control group. This study cannot separate out the contribution of each of these differences to the gains made by TK students. Nonetheless, the research literature suggests a set of possible mechanisms that could be in play. First, the greater regulation that resulted from folding TK into the larger K-12 system could account for some of these gains. This regulation likely increased the compensation and educational qualifications of teachers and decreased variation in the quality of experiences for students. The differences in the workforce may have increased the quality overall, while the reduced variation likely benefited children more likely to be in lower-quality care had TK not been available. Prior literature has shown that minority and economically disadvantaged families often enroll in less formal pre-K or lower quality pre-K experiences (Magnuson et al. 2004; Magnuson and Waldfogel 2005; Phillips and Lowenstein 2011). If TK provides these families with larger amounts of higher quality instruction, we would expect them to particularly benefit from this program. This study presents evidence that the Asian subgroup saw the greatest benefits in the BAS, while the white subgroup saw the least benefits. Further, the Asian and Hispanic subgroups saw benefits on both the BAS and CELDT. Overall, these results support studies such as Hotz and Xiao (2011) and Rigby, Ryan, and Brooks-Gunn (2007) who find that regulated markets lead to improved student outcomes.

Second, the more academic curricular and instructional focus of TK could account for the increases in child performance on the assessments. Aligning the curriculum to the development of

children in this age range may also have provided academic benefits. The district structured their TK classrooms and school days to be similar to those of kindergarteners and the curriculum contained less student-directed learning and playtime than other pre-K programs. At the same time, TK was less structured and academic than kindergarten. The positive findings in this study could be because a more academically oriented curriculum led to increased student learning.

The increased focus on academic skills could, in theory disadvantage students if it reduced children's engagement in school and other non-academic outcomes that have long-term benefits for students (Elkind and Whitehurst 2001; Stipek 2006; Zigler and Bishop 2006; Bassok, Latham, and Rorem, 2016). One limitation of this study is that I am unable to measure the effects of the program on social-emotional development directly. However, negative social-emotional effects might be reflected in negative effects on academic performance, which we do not see. Moreover, TK did not have negative effects on school attendance. Overall there was no detectable effect on the number of absences, except for students in the Asian subgroup who were, on average, absent 1.3 *fewer* days in kindergarten, though that advantage faded out by first grade. This result for Asian children is consistent with the notion that folding services into the school and modeling the school day after kindergarten helped students and parents acclimate to a full day academic environment. In this case the advantage likely dissipated by first grade as all students acclimated to this process throughout the kindergarten year. The results more broadly indicate that the socio-emotional health of the child was not likely impacted to such an extent that it affected the propensity of the child to attend school. Of course, this does not rule out more subtle effects on a child's socio-emotional health.

The estimates from this study are somewhat smaller than those from evaluations of pre-K programs in other urban areas. Weiland and Yoshikawa (2013) find literacy effect of 0.45 SD – 0.62 SD in their evaluation of Boston's program and Gormley et al. (2005) find literacy effects of 0.64 SD - 0.79 SD in their evaluation of Tulsa's program. In this study, CELDT estimates and BAS

estimates from OLS models are on the order of 0.15 SD – 0.30 SD. These differences could result from differences in the programs investigated or from methodological differences.

As Lipsey et al. (2014) point out, a shortcoming of previous studies is that students in the control group are part of a younger cohort and have yet to attend pre-K. The "treated students" consists of children who attended pre-K in the previous year and are starting their kindergarten year (cohort 1). The "control" students are those that are starting their pre-K year (cohort 2). This sampling strategy results in "treatment-on-the treated" estimate because it excludes any child who did not attend pre-K. In contrast, this study is a within-cohort comparison that includes all children, regardless of their pre-K experience. With a 33 percent take up the TK program, these intent-to-treat estimates will naturally be smaller. Two-stage least squares estimates from OLS models in this study vary from 0.45 SD – 0.60 SD. This order of magnitude is on par with Weiland and Yoshikawa's Boston study and but is still less than Gormley's Tulsa study. They are also on par with the treatment-on-the-treated estimates from Manship et al.'s study of TK programs in California, which detected an advantage of 0.30 SD - 0.50 SD for TK students on comparable pre-literacy skills.

Even accounting for this methodological difference, estimates from Gormley's study are higher. This difference may be because the alternative pre-K experiences available to TK-ineligible four year olds in San Francisco are of higher quality than the alternative pre-K experiences available to children the year before they enter Tulsa's universal pre-K program. Though the data I use do not contain information on the pre-K experience of each child who did not attend SFUSD's TK or pre-K program, at least 83 percent of 4-year olds attend pre-K in San Francisco where about 91 percent of programs are center-based. The control group received services not typically seen in other studies. This study estimates the benefits of TK above the benefits of a robust pre-K market of prekindergarten programs. From this perspective, smaller estimates should not be surprising.

TK, like many other high quality educational programs, is not inexpensive. Nonetheless, a back-of-the-envelope calculation estimates that the TK literacy benefits may not come at a

substantially greater cost than San Francisco's current spending on pre-K. In 2012-2013 San Francisco spent $17.24 million on preschool subsidies, building early childhood education capacity, wages, and curriculum. The program served 3,225 students at a cost of $5,346 per student. The program provides 612.5 hours of instruction for a total cost of $8.73 per student per hour. 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. In 2012-2013 the district spent $9,479 per pupil (California Department of Education, 2012). TK costs SFUSD $8.78 per student per hour, just 5 cents per student per hour more. These calculations do not represent the complete costs of each program because they only include costs associated with the district or universal pre-K program. They do not include opportunity costs that parents may regain by sending their child to a free, full day TK program. The calculations also likely understate the cost of providing pre-K services in San Francisco because the program provides subsidies only for families in financial need. Nevertheless, these calculations indicate the academic gains do not have to come at a significantly higher cost.

The TK program has recently been expanded with the introduction of Extended TK. Starting in 2015-2016, children who turn five after December 2, 2015 and before the end of school year can either enter TK at the time they turn 5, or start TK at the beginning of the school year (Torlakson 2015). This study cannot speak to whether extending TK to all four year olds, making it a form of universal pre-K, will benefit children. Offering free pre-K services to all four year olds would likely benefit families. However, more scrutiny is needed to determine if the TK curricula are appropriate for younger children. Like all RD studies, the results are valid only for children near the cutoff. This limitation is especially pertinent in this case because children of this age develop rapidly in a small amount of time. This study indicates that for students near the December 2 threshold SFUSD's efforts to implement TK has led to achievement gains, especially for English Language Learners.

References

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

Barreca, A. I., Jason M. Lindo, and Glen R. Waddell. (2015). Heaping induced bias in regression discontinuity designs. *Economic Inquiry, 54*(1), pp. 268-293.

Bassok, D., Maria Fitzpatrick, Erica Greenberg, and Susanna Loeb. (2016). Within-and between sector quality differences in early childhood education and care. *Child Development, 87*(5), pp. 1627-1645.

Bassok, D., Maria Fitzpatrick, Susanna Loeb, and Agustina S. Paglayan. (2013). The early childhood care and education workforce in the United States: Understanding changes from 1990 through 2010. *Education Finance and Policy, 8*(4), pp. 581–601.

Bassok, D., Scott Latham, and Anna Rorem. (2016). Is kindergarten the new first grade? *AREA Open, 2*(1), pp 1-31.

Belfield, C., Milagros Nores, Steve Barnett, and Lawrence Schweinhart. (2006). High/Scope Perry Preschool: Cost benefit analysis at age 40. *Journal of Human Resources, 41*(1), pp. 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

California Department of Education. (2014). California English Language Development Test technical report. Retrieve, June 27, 2015 from www.cde.ca.gov/ta/tg/el/techreport.asp

Campbell, F. A., Elizabeth P. Pungello, Margaret Burchinal, Kirsten Kainz, Yi Pan, Barbara H. Wasik, Joseph J. Sparling, Oscar A. Barbarin, and Craig T. Ramey. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology, 48*(4), pp. 1033-1043

Currie, J. and Duncan Thomas. (1995), Does Head Start make a difference? *The American Economic Review, 85*(3), pp. 341-364.

Currie, J. and Duncan Thomas. (2000), School quality and the longer-term effects of Head Start. *The Journal of Human Resources, 35*(4), pp. 755-774.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Econometrics, 1*(3), pp. 111-134.

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

Ehri, L. C. (2015). How children learn to read words. In A. Pollatsek and R. Treiman (Eds.), *The Oxford Handbook of Reading* (pp. 293–310). Oxford, U.K.: Oxford University Press.

Elkind, D. and Grover J. Whitehurst (2001). Young Einsteins: Should Head Start emphasize academics? Education Next, 1. Retrieved, Sept. 14, 2015 from: educationnext.org

Feller, A., Todd Grindal, Luke Miratrix, and Lindsay Page. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care-type settings, Working Paper, Retrieved June 17, 2016 from: papers.ssrn.com/sol3/papers.cfm?abstract_id=2534811

First Five (2015). Quality ratings of early care and education programs in San Francisco through December 7, 2015. Retrieved June 17, 2016 from: first5sf.org/programs/quality ratings- of early-care-and-education-programs-in-san-francisco/

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

Fuller, B., Sharon L. Kagan, Susanna Loeb, and Yueh-Wen Chang. (2004). Child care quality: Centers and home settings that serve poor families. *Early Childhood Research Quarterly, 4*(19), pp. 505-527.

Ferguson, P.C. (1991). Longitudinal outcome differences among promoted and transitional at risk kindergarten students, *Psychology in the Schools*, *28*(2), pp. 139-146.

Garces, E., Duncan Thomas, and Janet Currie. (2002). Longer-term effects of Head Start. *The American Economic Review*, *92*(4), pp. 999-1012.

Gilliam, W. and Edward Zigler. (2004). State efforts to evaluate the effects of prekindergarten: 1977 to 2003. New Haven, CT: Yale University Child Study Center.

Gormley, W.T., Ted Gayer, Deborah Phillips, and Brittany Dawson. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), pp. 872-884.

Governor's State Advisory Council on Early Learning and Care. (2013). Transitional Kindergarten implementation guide. Retrieved, April 3, 2015 from http://www.cde.ca.gov/ci/gs/em/documents/tkguide.pdf

Heckman, J.J., Seong H. Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the High Scope Perry Preschool Program. *Quantitative Economics, 1*(1), pp. 1-46.

Hotz, V. J. and Mo Xiao. (2011). The impact of regulations on the supply and quality of care in child care markets. *American Economic Review, 101*(5), pp. 1775–1805.

Huang, F.L., Marcia A. Ivernizzi, and E. Allison Drake. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly, 27*(1), pp. 33-45.

Imbens, G.W. and Karthik Kalyanaraman. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies, 79*(3), pp. 933-959.

Karweit, N.L. and Barbara A. Wasik. (1992). A review of extra-year kindergarten programs and transitional first grades, Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students. Retrieved from http://eric.ed.gov/?id=ED357894

Kjeldsen, A., Antti Karna, Pekka Niemi, Ake Olofsson, and Katarina Witting. (2014). Gains from training in phonological awareness in kindergarten predict reading comprehension in grade 9, *Scientific Studies of Reading, 18*(6), pp. 452-467.

Ladd, H.F., Clara G. Muschkin, and Kenneth A. Dodge. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina, *Journal of Policy Analysis and Management, 33*(1), pp. 162-187.

Lee, D.S. and Thomas Lemieux. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), pp. 281-355.

Lee, V. E., Susanna Loeb, and Sally Lubeck. (1998). Contextual effects of pre-K classrooms for disadvantaged children on cognitive development. *Child Development, 69*(2), pp. 479-494.

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

Lipsey, M.W., Kerry G. Hofer, Nianbo Dong, Dale C. Farran, and Carol Bilbrey. (2013). Evaluation of the Tennessee Voluntary Prekindergarten Program. Research Report. Retrieved June 17, 2016 from: my.vanderbilt.edu/tnprekevaluation/files/2013/10/ August2013_PRI_Kand1stFollowup_TN-VPK_RCT_ProjectResults_FullReport1.pdf

Lipsey, M.W., Christina Weiland, Hirokazu Yoshikawa, Sandra J. Wilson, and Kerry G. Hofer. (2014). Kindergarten age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis, 31*(3), pp. 1-18.

Loeb, S., Bruce Fuller, Sharon L. Kagan, Carrol Bidemi, Carroll Judith, and Jan Mc Carthy. (2004). Child care in poor communities: Early learning effects of type, quality and stability. *Child Development, 75*(1), pp. 47-65.

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

Manship, K., Heather Quick, Aleksandra Holod, Nicholas Mills, Burhan Ogut, Jodi J. Chernoff, Jarah Blum, Alison Hauser, Jennifer Anthony, and Raquel Gonzalez. (2015). Impact of California's Transitional Kindergarten program, 2013 2014. AIR. Retrieved, December 29, 2015 from: air.org /sites/default/files/ downloads/report/Impact-of-Californias-Transitional Kindergarten-Program-Dec-15.pdf

Magnuson, K.A., Marcia K. Meyers, Christopher J. Ruhm, and Jane Waldfogel. (2004). Inequality in preschool education and school readiness. *American Education Research Journal, 41*(1), pp. 115-157.

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

National Early Literacy Panel. (2008). Developing early literacy. *National Institute for Literacy*.

Phillips, D.A. and Amy E. Lowenstein. (2011). Early care, education, and child development. *Annual Review of Psychology, 62*, pp. 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

Ramey, C.T. and Frances A. Campbell. (1984). Preventative education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88*(5), pp. 515–523.

Rigby, E., Rebecca M. Ryan, and Jeanne Brooks-Gunn. (2007). Child care quality in different state policy contexts. *Journal of Policy Analysis and Management, 26*(4), pp. 887-907.

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

Shager, H.M., Holly S. Schindler, Katherine A. Magnuson, Greg J. Duncan, Hirokazu Yoshikawa, Cassandra M.D. Hart. (2012). Can research design explain variation in Head Start research results? A meta-analysis of cognitive and achievement outcomes. *Educational Evaluation and Policy Analysis, 35*(1), pp. 76-95.

Stipek, D. (2006). NCLB comes to preschool. *Elementary School Journal, 106*(5) pp. 455-465.

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

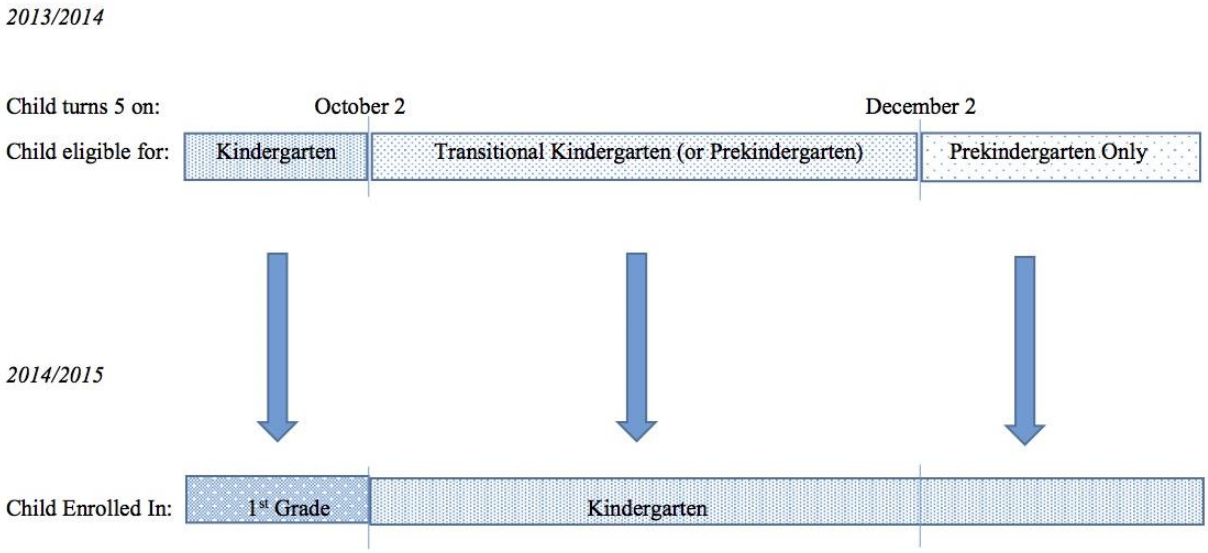
Vuong, Q.H (1989). Likelihood ratio tests for model selection. *Econometrica, 57(2)*, pp. 307-333.

Weiland, C., and Hirokazu Yoshikawa. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112 2130.

Wong, V., Thomas Cook, William Barnett, and Kwanghee Jung. (2008). An evaluation of five state prekindergarten programs. *Journal of Policy Analysis and Management, 27*(1), pp. 122-154.

Zhai, F., Jeanne Brooks-Gunn, and Jane Waldfogel. (2014). Head Start's impact is contingent on alternative type of care in comparison group. *APA, 50*(12), pp. 2572-2586.

Zigler, E. F. and Sandra J. Bishop-Josef. (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, pp. 15–35.

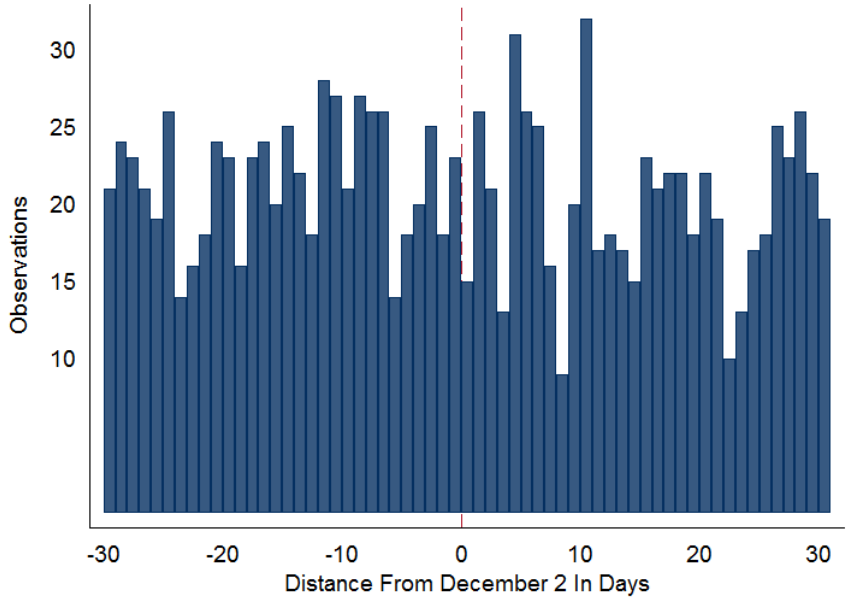


(a) Early childhood education experience based on birthdate cut point for cohort 2

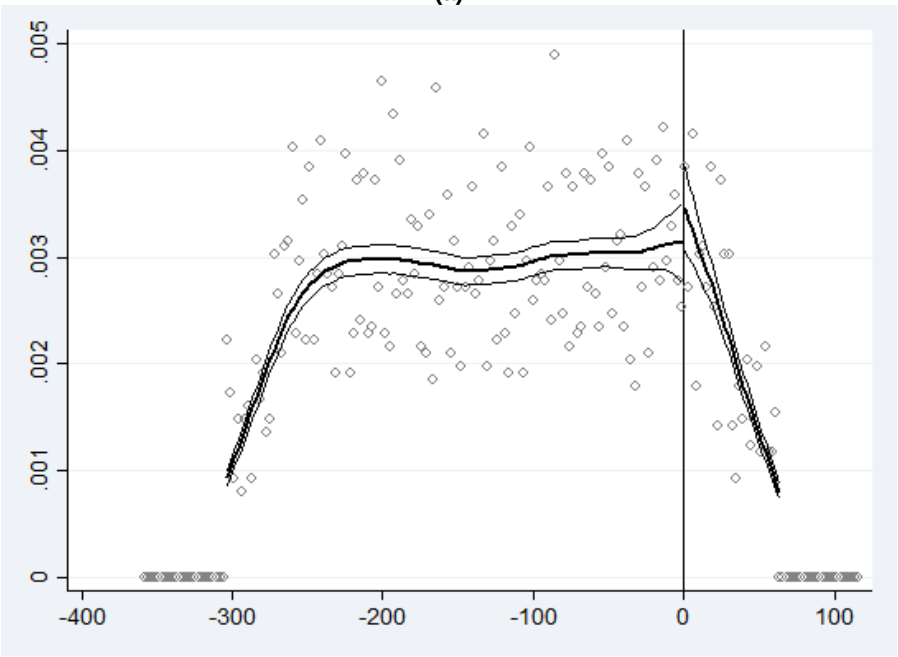
SFUSD Prekindergarten	SFUSD Transitional Kindergarten
Structure of Day	
Children start at different times based on contract	Academic day starts at same time for all children
Families select hours of instruction	6 hour program
Breakfast provided	No breakfast but may have morning snack
Nap time	No nap 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

(b) Differences in SFUSD Transitional Kindergarten and SFUSD prekindergarten programs

Figure 1: Transitional Kindergarten enrollment criteria and differences between SFUSD Transitional Kindergarten and SFUSD prekindergarten programs



(a)



(b)

Figure 2: 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 (0.089). Vertical lines indicate the December 2 threshold.

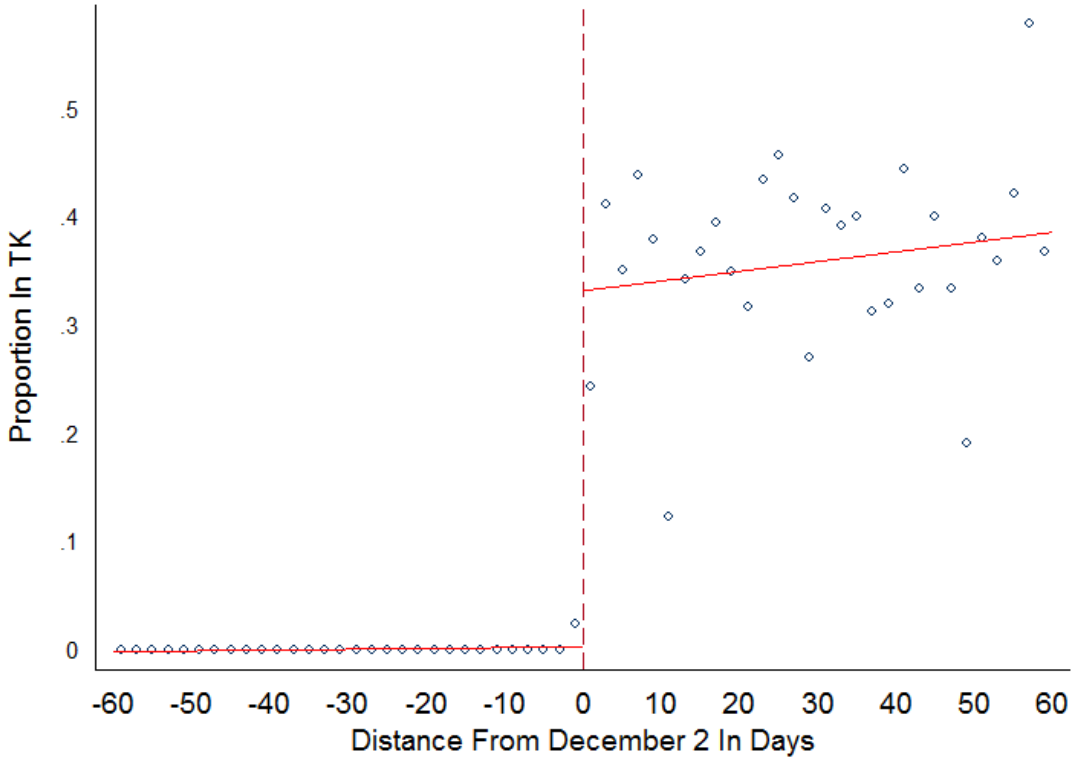
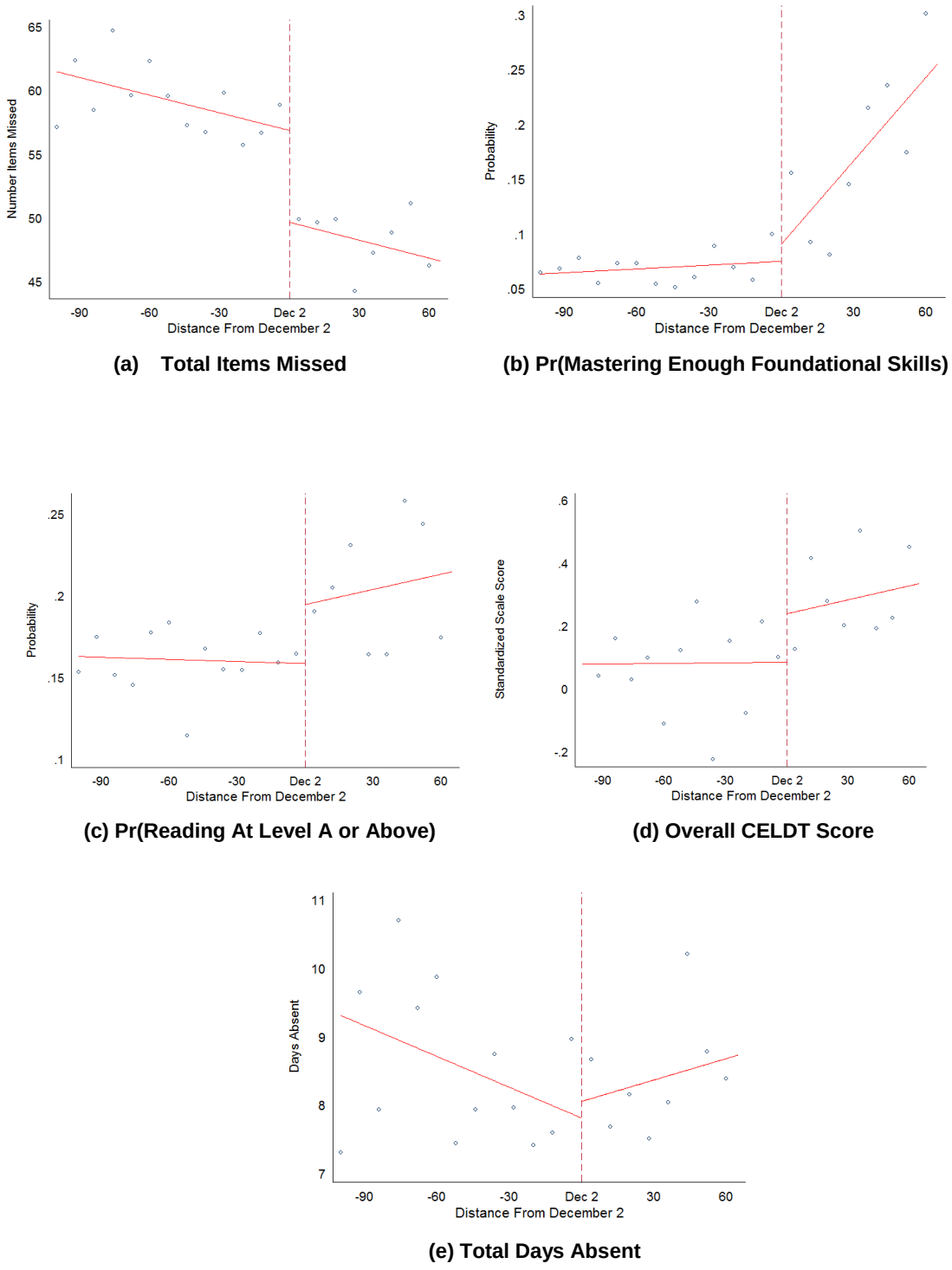


Figure 3: First Stage: Enrollment in TK in prior year by birthday. Each dot represents the proportion of students that enrolled in TK in the previous year within a bin of 2 days. The vertical line represents the December 2 threshold. Regression lines are estimated using local linear regression with a rectangular kernel on a bandwith of 60 days.



(a) Total Items Missed

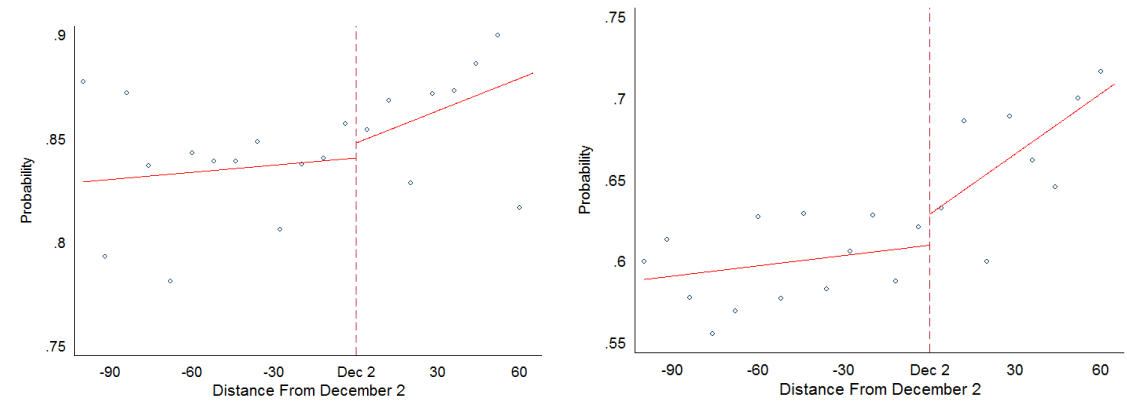
(b) Pr(Mastering Enough Foundational Skills)

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

(d) Overall CELDT Score

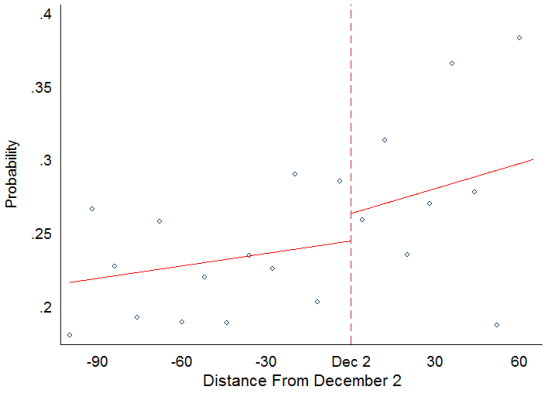
(e) Total Days Absent

Figure 4: Fall kindergarten 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.

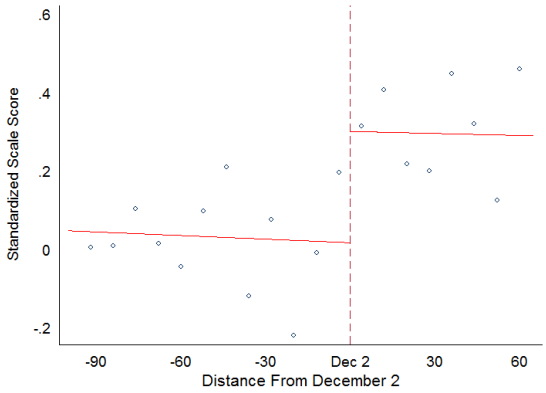


(a) Pr(Reading at Level C or Above)

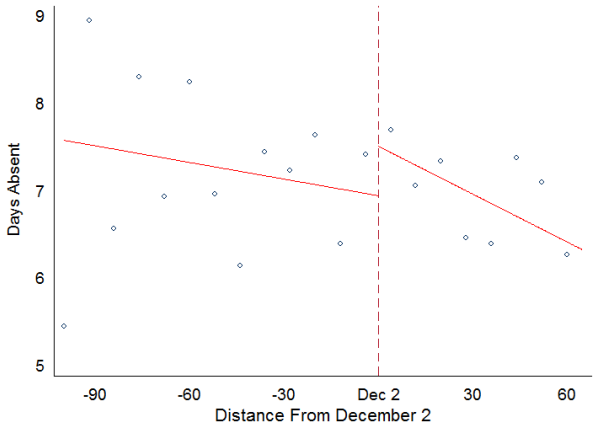
(b) Pr(Reading At Level E or Above)



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

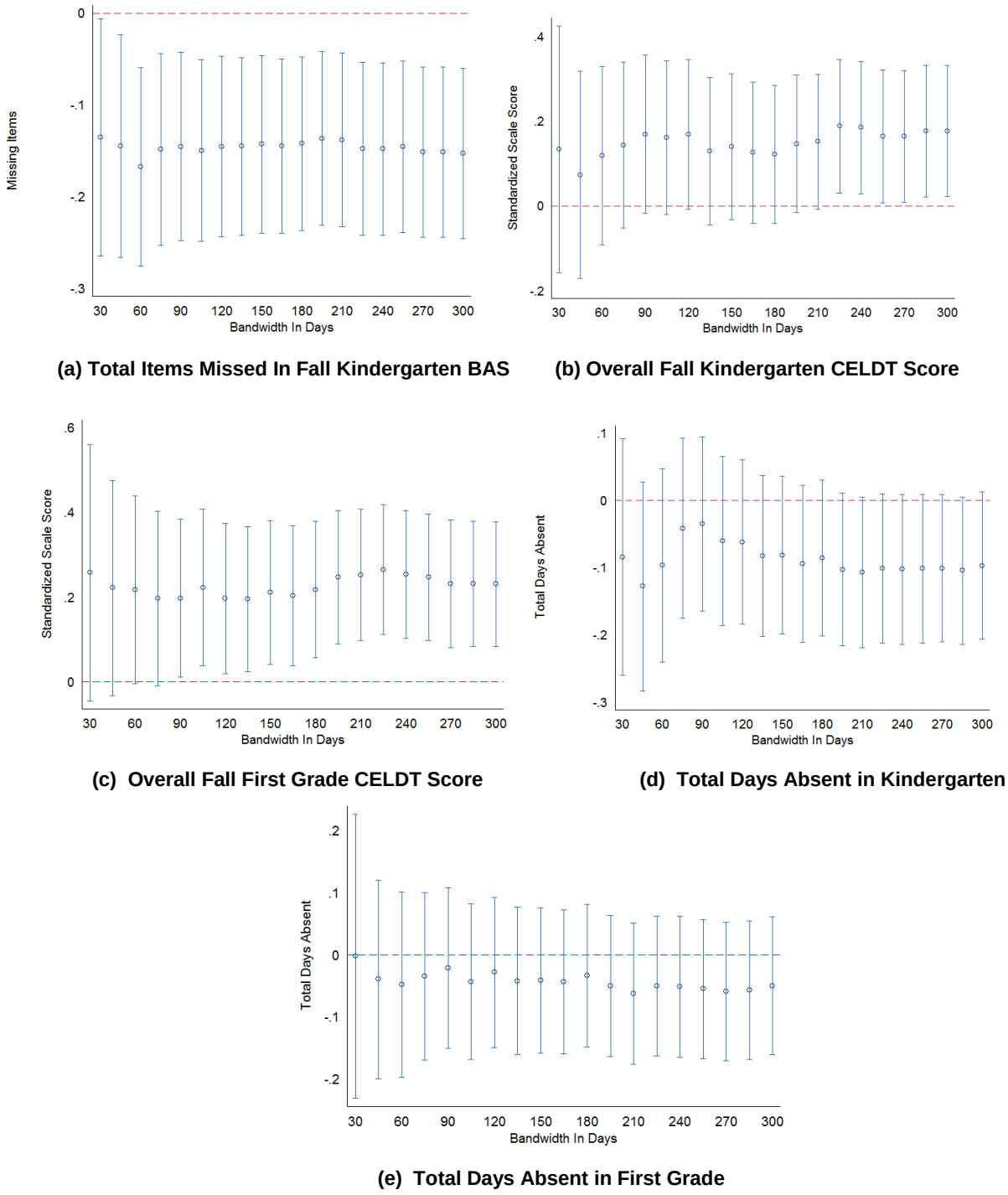


(d) Overall CELDT Score



(e) Total Days Absent

Figure 5: Fall first grade 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.



(a) Total Items Missed In Fall Kindergarten BAS

(b) Overall Fall Kindergarten CELDT Score

(c) Overall Fall First Grade CELDT Score

(d) Total Days Absent in Kindergarten

(e) Total Days Absent in First Grade

Figure 6: Robustness checks of outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relevant outcome for observations in bandwidths between 30 and 300 days. Figures (a), (d), and (e) employ negative binomial models Figures (b) and (c) employ OLS models. 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 3. Standard errors are clustered on the birthday rating variable except in negative binomial models where it must be clustered at the teacher-by-year cell.

Table 1: San Francisco universal pre-K Quality Rating and Improvement System results by sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	N(Centers)
	Child Observation	Developmental & Health Screening	Minimum Qualifications of Lead Teacher	Child Interactions as Measured by CLASS	Ratio and Group Size	Program Environment Rating Scale	Director Qualifications	Total Points	Star Level	
SFUSD School-Based Centers	3.32	0.42	4.03	3.29	4.45	4.45	4.90	24.87	3.35	31
Head Start Centers	4.06	5.00	4.35	3.94	4.29	3.88	3.82	29.35	4.12	17
Other Center Care	3.11	2.54	4.07	3.43	3.96	3.91	3.86	24.81	3.47	81
Home Based Care	2.69	2.85	4.69	3.38	N/A	4.46	N/A	18.08	3.69	13

Note: Each cell contains the average rating, calculated by the author, for programs in San Francisco's Universal Prekindergarten which opted to be evaluated on the Quality Rating and Improvement System (QRIS). This sample includes 142 of the 147 pre-K providers in the San Francisco universal pre-K market. These programs were evaluated between 2013 and 2015. Source data is from First Five, 2015.

Table 2: Descriptive Statistics

	Analytical Sample					Former TK		Former Non-TK		p-value
Variable	Mean	St. Dev.	Min	Max	N (Total)	Mean	N	Mean	N	(TK-Non TK)
Programmatic Characteristics										
TK Eligible	0.140	0.347	0	1	6739	0.997	335	0.096	6404	0.000
Attended TK In Year T-1	0.050	0.217	0	1	6739	1.000	335	0.000	6404	---
Attended District PreK in Year T-1	0.169	0.374	0	1	6739	0.000	335	0.177	6404	0.000
Birthday (days from December 2)	-120.143	98.367	-304	61	6739	26.188	335	-127.798	6404	0.000
Student Characteristics										
Female	0.492	0.500	0	1	6739	0.487	335	0.492	6404	0.837
Asian	0.311	0.463	0	1	6739	0.421	335	0.305	6404	0.000
Hispanic	0.250	0.433	0	1	6739	0.260	335	0.249	6404	0.666
White	0.165	0.371	0	1	6739	0.099	335	0.168	6404	0.001
Other	0.175	0.380	0	1	6739	0.179	335	0.175	6404	0.837
Declined To State Ethnicity	0.098	0.297	0	1	6739	0.042	335	0.101	6404	0.000
Special Education	0.076	0.265	0	1	6739	0.033	335	0.078	6404	0.002
Limited English Proficient (LEP)	0.491	0.500	0	1	6739	0.594	335	0.486	6404	0.000
Home Language:										
Chinese	0.171	0.376	0	1	6739	0.296	335	0.164	6404	0.000
Spanish	0.149	0.356	0	1	6739	0.173	335	0.148	6404	0.206
English	0.597	0.491	0	1	6739	0.457	335	0.604	6404	0.000
Other	0.084	0.277	0	1	6739	0.075	335	0.084	6404	0.539
Dominant Language:										
Chinese	0.206	0.404	0	1	6739	0.304	335	0.201	6404	0.000
Spanish	0.174	0.379	0	1	6739	0.182	335	0.173	6404	0.675
English	0.506	0.500	0	1	6739	0.418	335	0.511	6404	0.001
Other	0.114	0.318	0	1	6739	0.096	335	0.115	6404	0.267
Kindergarten Fountas and Pinnell Outcomes										
Upper Case Letters	20.410	8.355	0	29	6739	22.499	335	20.300	6404	0.000
Lower Case Letters	18.804	8.596	0	29	6739	21.857	335	18.645	6404	0.000
Letter Sounds	12.679	9.137	0	29	6739	17.552	335	12.424	6404	0.000
High Frequency Words	6.912	7.815	0	25	6739	13.663	335	6.559	6404	0.000
Initial Word Sounds	5.293	3.219	0	8	6739	6.421	335	5.234	6404	0.000
Early Literacy Behaviors	6.915	3.049	0	11	6739	8.400	335	6.837	6404	0.000
Blending	6.915	3.049	0	10	6427	5.792	317	3.700	6110	0.000
Rhyming	6.915	3.049	0	10	5997	7.260	292	5.642	5705	0.000
Mastered Required Found. Skills	6.915	3.049	0	1	6739	0.239	335	0.061	6404	0.000
Reading at Level A or Above	6.915	3.049	0	1	6739	0.224	335	0.164	6404	0.004
Test Given In Spanish	0.140	0.347	0	1	6739	0.131	335	0.141	6404	0.631
Kindergarten CELDT Outcomes										
Listening	374.863	86.019	220	570	3310	419.422	199	372.013	3111	0.000
Speaking	388.218	94.436	140	630	3310	428.211	199	385.659	3111	0.000
Reading	294.571	57.558	220	570	3310	343.297	199	291.455	3111	0.000
Writing	306.521	52.327	220	600	3310	352.688	199	303.567	3111	0.000
Overall	372.973	77.503	184	580	3310	415.759	199	370.236	3111	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	0.016
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	454.807	62.608	220	570	2663	485.439	180	452.586	2483	0.000
Speaking	457.292	65.408	140	630	2663	483.778	180	455.372	2483	0.000
Reading	396.753	76.247	220	570	2663	426.289	180	394.612	2483	0.000
Writing	400.983	57.135	220	600	2663	430.872	180	398.816	2483	0.000
Overall	449.836	56.290	184	594	2663	478.500	180	447.758	2483	0.000
Attendance										
Total Days Absent in Kindergarten	8.424	9.168	0	174	6739	8.376	335	8.426	6404	0.922
Total Days Absent in First Grade	7.095	7.846	0	177	6219	6.752	315	7.113	5904	0.426

Note: Former TK students are students in the analytical sample who enrolled in the district's TK program in the previous year. Former prekindergarten students are students who enrolled in the district's pre-kindergarten 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. Students who experienced district 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 Langauge Development Test.

Table 3: RD regressions of covariate balance

Variable	Full Sample	$\|B_{ict}\| \leq 60$	$\|B_{ict}\| \leq 30$
Student Characteristics			
Female	0.011	-0.017	-0.029
	(0.029)	(0.037)	(0.050)
Asian	-0.016	-0.044	-0.034
	(0.035)	(0.044)	(0.059)
Hispanic	0.016	0.017	-0.022
	(0.028)	(0.036)	(0.046)
White	-0.028	-0.032	-0.001
	(0.028)	(0.036)	(0.050)
Other	0.047+	0.036	0.034
	(0.025)	(0.035)	(0.055)
Declined To State Ethnicity	-0.019	0.021	0.018
	(0.019)	(0.024)	(0.030)
Special Education	-0.011	-0.013	-0.002
	(0.015)	(0.018)	(0.021)
Limited English Proficient (LEP)	-0.029	-0.057	-0.078
	(0.038)	(0.047)	(0.066)
Home Language:			
Chinese	-0.000	-0.018	-0.036
	(0.030)	(0.034)	(0.047)
Spanish	-0.005	-0.014	-0.024
	(0.020)	(0.028)	(0.041)
English	-0.011	-0.004	0.045
	(0.035)	(0.041)	(0.061)
Other	0.016	0.036+	0.015
	(0.015)	(0.020)	(0.026)
Dominant Language:			
Chinese	-0.019	-0.048	-0.066
	(0.028)	(0.034)	(0.046)
Spanish	-0.010	0.000	-0.002
	(0.021)	(0.027)	(0.038)
English	0.029	0.049	0.072
	(0.037)	(0.046)	(0.065)
Other	-0.000	-0.001	-0.004
	(0.018)	(0.024)	(0.032)
Test Characteristic			
Test Given In Spanish	-0.026	-0.012	0.027
	(0.026)	(0.033)	(0.045)
N	6,739	2,182	1,271

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 4: RD regressions of first stage

Dependent Variable: Enrolled In TK in Year T-1			
	(1)	(2)	N
Full Sample	0.335**	0.321**	6,739
	(0.032)	(0.027)	
$\|B_{ict}\| \leq 60$	0.329**	0.309**	2,182
	(0.032)	(0.031)	
$\|B_{ict}\| \leq 30$	0.312**	0.284**	1,271
	(0.042)	(0.044)	
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 3. 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 5: Reduced form estimates of fall kindergarten and first grade outcomes

	(1)	(2)			(3)	(4)	
Panel A: Kindergarten Outcomes				Panel B: First Grade Outcomes			
Fountas And Pinnell Outcomes			N	Fountas And Pinnell Outcomes			N
Total Items Missed	-0.141*	-0.181**	6,739	Reading Scale (Ordinal Logit)	-0.051	-0.036	6,219
	(0.059)	(0.042)			(0.120)	(0.120)	
Pr(Mastering Required Found. Skills)	0.012	0.033	6,739	Pr(Reading at Level C or Above)	0.007	0.008	6,219
	(0.022)	(0.021)			(0.027)	(0.023)	
Pr(Reading at Level A or Above)	0.020	0.014	6,739	Pr(Reading at Level E or Above)	0.013	0.021	6,219
	(0.028)	(0.016)			(0.038)	(0.030)	
				Pr(Reading at Level I or Above)	0.021	0.017	6,219
					(0.031)	(0.028)	
CELDT Outcomes			N	CELDT Outcomes			N
				Overall Score	0.250**	0.231**	2,663
Overall Score	0.118	0.176*	3,310		(0.092)	(0.075)	
	(0.110)	(0.079)					
Attendance Outcome			N	Attendance Outcome			N
Total Days Absent	-0.055	-0.050	6,739	Total Days Absent	0.031	0.022	6,219
	(0.072)	(0.051)			(0.067)	(0.053)	
Covariates		√				√	
Fixed Effects		√				√	

Note: Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated outcome. Row headers indicate the dependent variable. Covariates include an indicator for kindergarten year, teacher-by-year fixed effects, and all variables in Table 3. Negative binomial models are used to estimate the effect of Transitional Kindergarten on the total items missed on the Foutas and Pinnell assessment and the total number of days absent. Ordinal logit models are used to estimate the effect of Transitional Kindergarten on the Fountas and Pinnell reading scale. 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 teacher-by-year fixed effect. +indicates p<0.10, *p<0.05, **p<0.01

Table 6: Reduced form estimates of Fountas and Pinnell and attendance outcomes by subgroup

Kindergarten	(1)	1st Grade	(2)	Kindergarten	(3)	1st Grade	(4)
Panel A: Full Sample, N= 6,739		*N=6,219*		*Panel F: White N=1,111*		*N=1,001*	
Total Items Missed on BAS	-0.181**	BAS Reading Scale	-0.036	Total Items Missed on BAS	-0.039	BAS Reading Scale	-0.122
	(0.042)		(0.120)		(0.128)		(0.331)
Pr(Mastering Required Found. Skills)	0.033			Pr(Mastering Required Found. Skills)	-0.116*		
	(0.021)				(0.058)		
Total Days Absent	-0.050	Total Days Absent	0.022	Total Days Absent	0.007	Total Days Absent	0.191
	(0.051)		(0.053)		(0.129)		(0.133)
Panel B: Male, N=3,423		*N=3,144*		*Panel G: Other N=1,179*		*N=1,068*	
Total Items Missed on BAS	-0.210**	BAS Reading Scale	-0.136	Total Items Missed on BAS	0.018	BAS Reading Scale	-0.136
	(0.060)		(0.167)		(0.115)		(0.280)
Pr(Mastering Required Found. Skills)	0.047+			Pr(Mastering Required Found. Skills)	-0.038		
	(0.027)				(0.056)		
Total Days Absent	-0.087	Total Days Absent	0.004	Total Days Absent	0.006	Total Days Absent	-0.053
	(0.072)		(0.075)		(0.128)		(0.139)
Panel C: Female, N=3,316		*N=3,075*		*Panel H: Limited English Proficient (LEP), N=3,310*		*N=3,115*	
Total Items Missed on BAS	-0.164**	BAS Reading Scale	0.078	Total Items Missed on BAS	-0.166**	BAS Reading Scale	-0.084
	(0.061)		(0.177)		(0.056)		(0.173)
Pr(Mastering Required Found. Skills)	0.023			Pr(Mastering Required Found. Skills)	0.045		
	(0.031)				(0.029)		
Total Days Absent	-0.010	Total Days Absent	0.017	Total Days Absent	-0.069	Total Days Absent	0.013
	(0.076)		(0.080)		(0.081)		(0.083)
Panel D: Asian, N=2,095		*N=2,017*		*Panel I: English Proficient N=3,429*		*N=3,104*	
Total Items Missed on BAS	-0.381**	BAS Reading Scale	0.133	Total Items Missed on BAS	-0.227**	BAS Reading Scale	0.067
	(0.086)		(0.215)		(0.063)		(0.170)
Pr(Mastering Required Found. Skills)	0.126**			Pr(Mastering Required Found. Skills)	0.019		
	(0.048)				(0.030)		
Total Days Absent	-0.266*	Total Days Absent	-0.101	Total Days Absent	-0.057	Total Days Absent	-0.021
	(0.112)		(0.110)		(0.069)		(0.072)
Panel E: Hispanic, N=1,683		*N=1,546*					
Total Items Missed on BAS	-0.174**	BAS Reading Scale	-0.146				
	(0.067)		(0.241)				
Pr(Mastering Required Found. Skills)	0.028						
	(0.022)						
Total Days Absent	0.161	Total Days Absent	0.096				
	(0.097)		(0.104)				

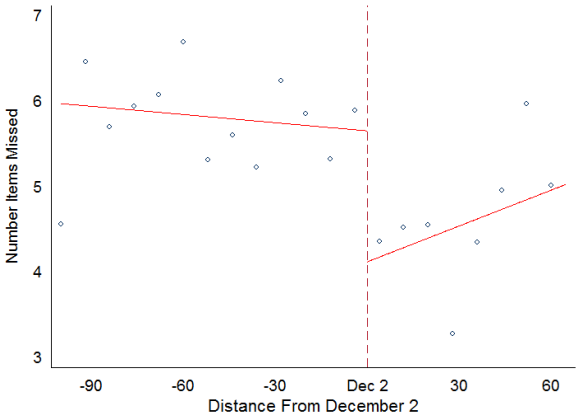
Note: Each cell represents the results of a separate regression discontinuity estimate of the effect of Transitional Kindergarten on the indicated outcome. Row headers indicate the dependent variable and panel headers indicate the subsample. Negative binomial models were used to estimate the effect of Transitional Kindergarten on the total items missed and total days absent and ordinal logit models were used to estimate the effect of Transitional Kindergarten on the Fountas and Pinnell reading scale. OLS was used in all other cases. All functional forms include a linear spline and covariates defined in Table 5. Akaike's Information Criteria indicates a linear spline is optimal. All standard errors are clustered on day of birth running variable except for conditional negative binomial and ordinal logit models which must be clustered on the teacher-by-year fixed effect. +indicates p<0.10, *p<0.05, **p<0.01

Table 7: Reduced form estimates of kindergarten and first grade CELDT outcomes by subgroup

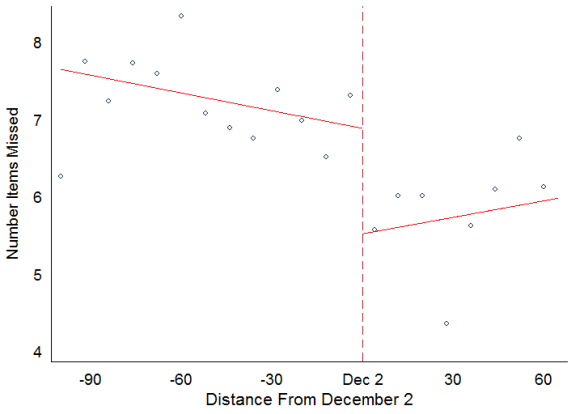
Dependent Variable: Overall Score	Kindergarten		First Grade	
	(1)	N	(2)	N
All English Language Learners (ELLs)	0.176*	3,310	0.231**	2,663
	(0.079)		(0.075)	
Male	0.135	1,662	0.212+	1,354
	(0.120)		(0.123)	
Female	0.241*	1,648	0.199+	1,309
	(0.111)		(0.106)	
Asian	0.117	1,523	0.279**	1,291
	(0.117)		(0.099)	
Hispanic	0.356*	1,159	0.159	950
	(0.138)		(0.139)	

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 5. 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

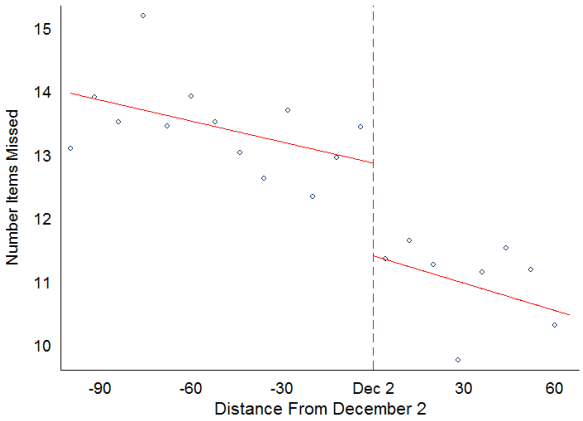
Appendix



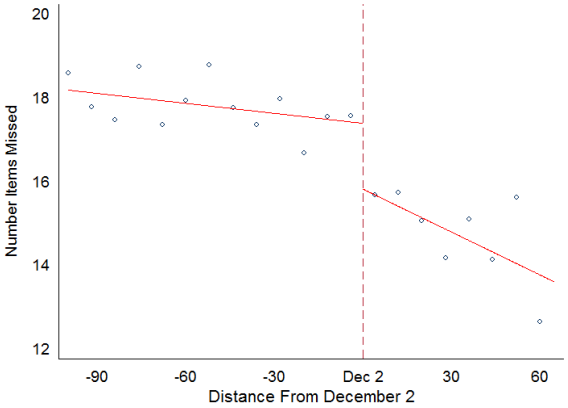
(a) Upper Case Letter Recognition



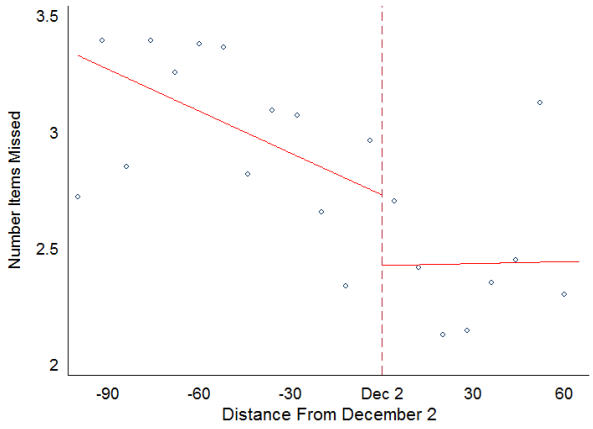
(b) Lower Case Letter Recognition



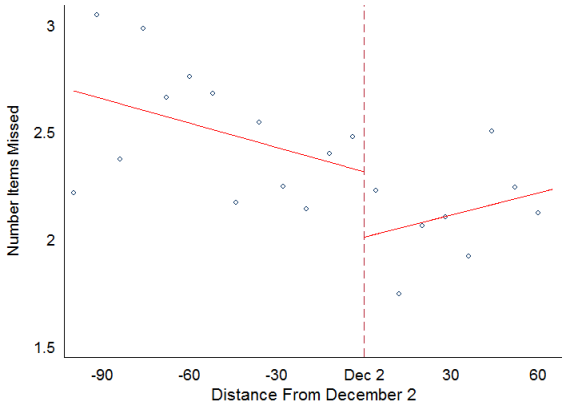
(c) Letter Sounds



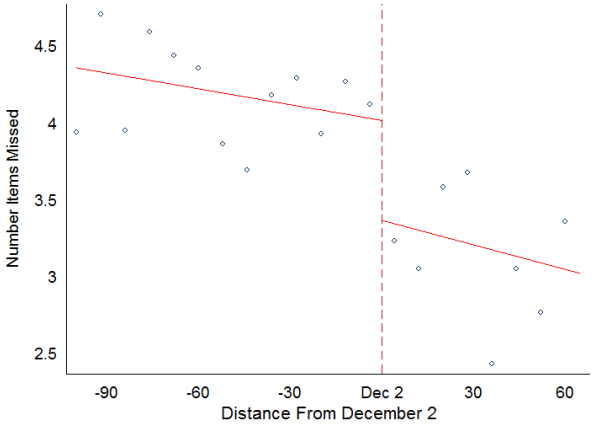
(d) High Frequency Word Recognition



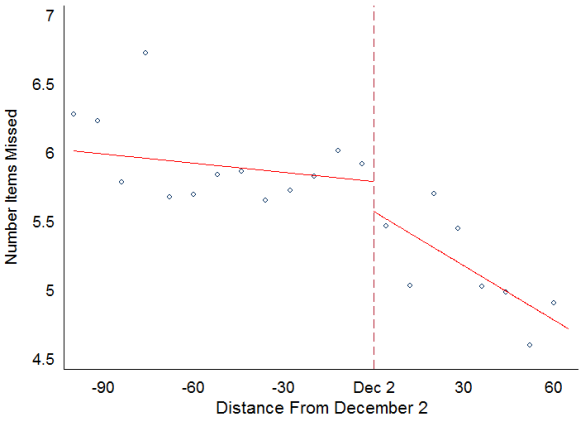
(e) Early Literacy Behaviors



(f) Initial Word Sounds

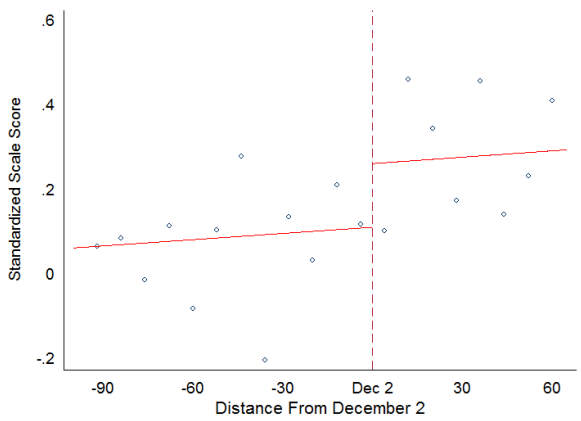


(g) Rhyming

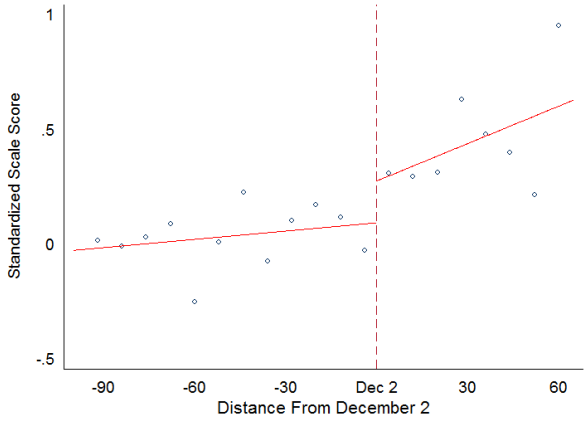


(h) Blending

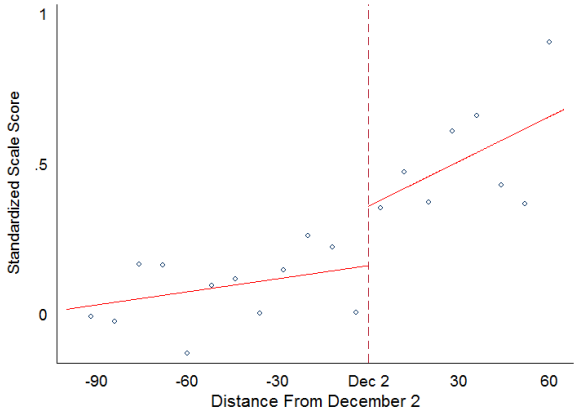
Figure A1: Fall kindergarten Fountas and Pinnell foundational literacy outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday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



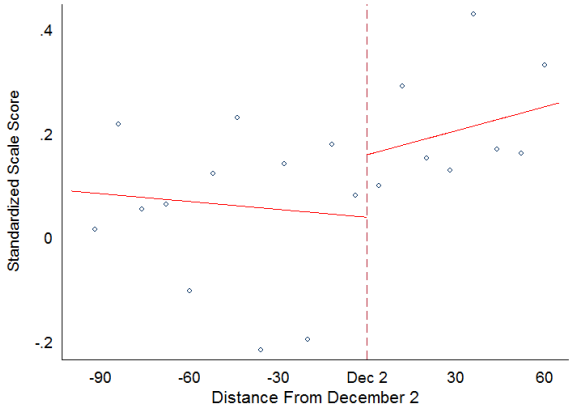
(a) Listening



(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. CELDT stands for the California English Language Development Test.

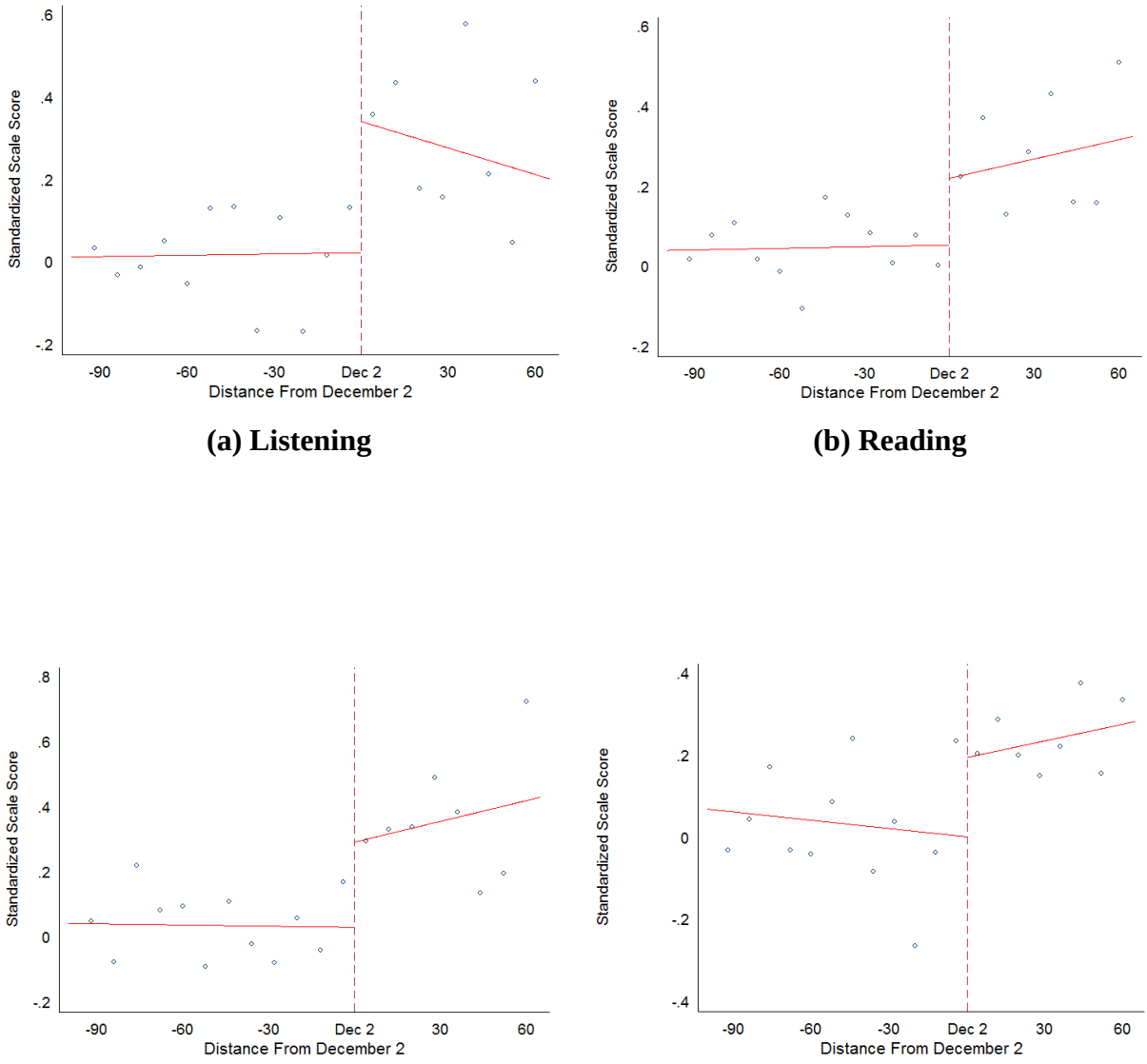
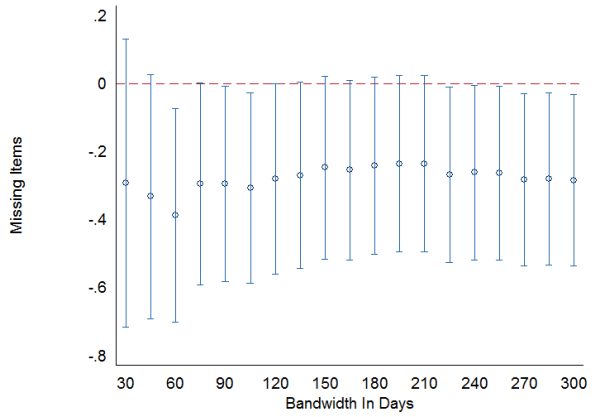
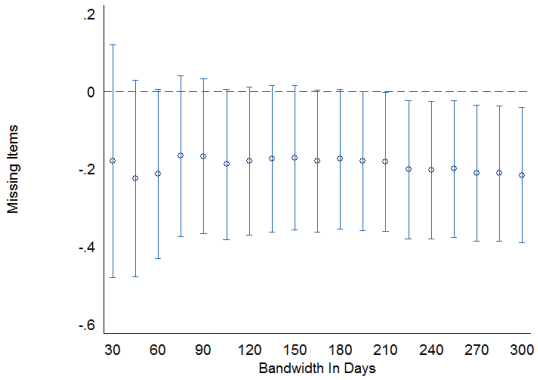


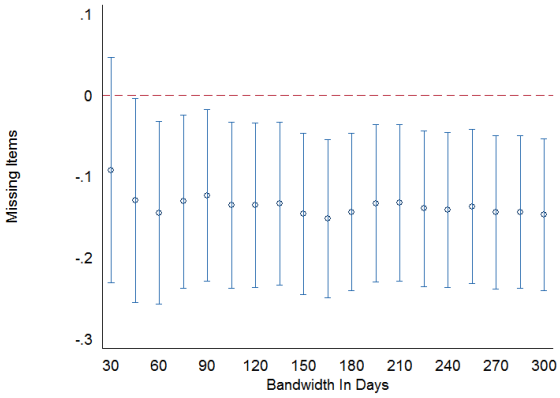
Figure A3: Fall first grade CELDT subtest outcomes. Each dot represents the average outcome in an 8 day bin width. TK eligible students are to the right of the vertical line and TK ineligible students are to the left of the line. The x-axis represents distance of birthday in days from December 2. CELDT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



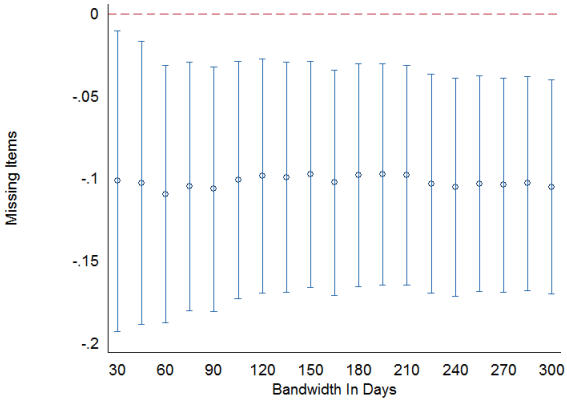
(a) Upper Case Letter Recognition



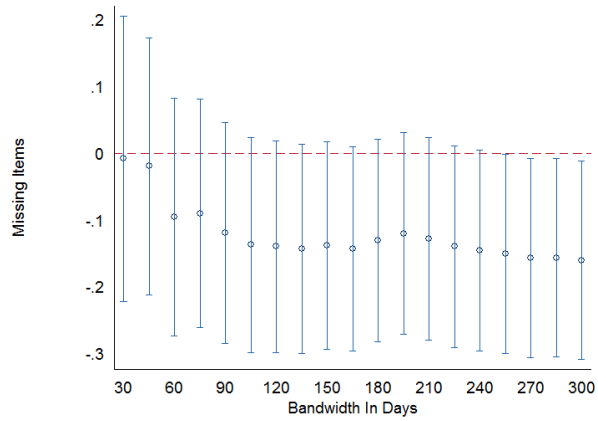
(b) Lower Case Letter Recognition



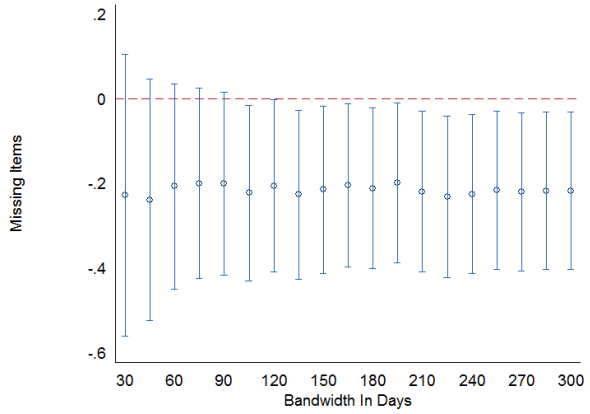
(c) Letter Sounds



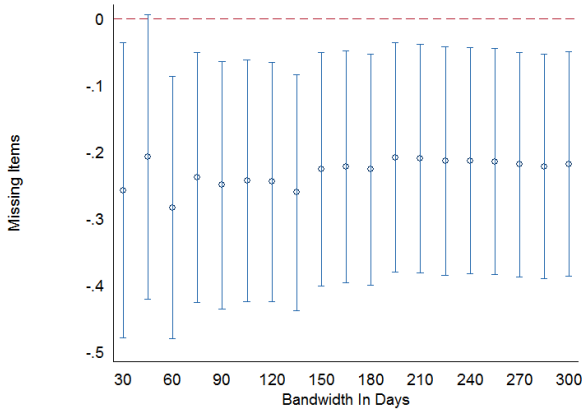
(d) High Frequency Word Recognition



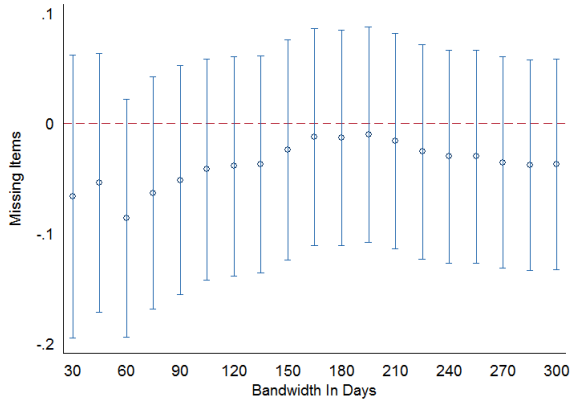
(e) Early Literacy Behaviors



(f) Initial Word Sounds

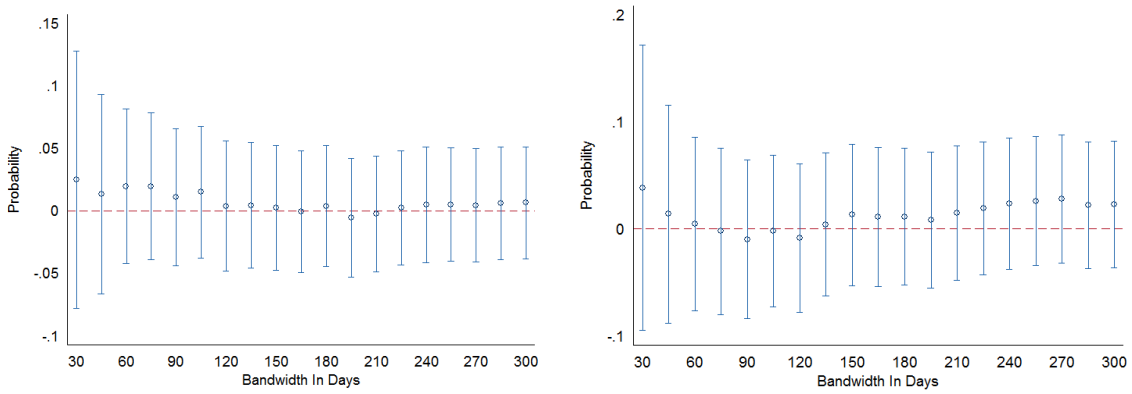


(g) Rhyming

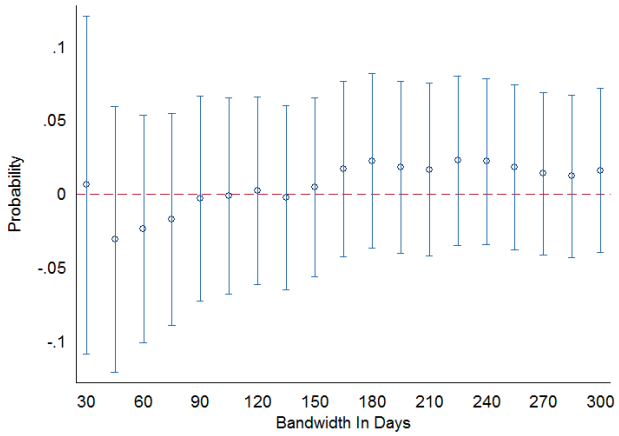


(h) Blending

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 relevant 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 bandwidths. All regressions employ a linear spline functional form with covariates detailed in Table 5. Standard errors are clustered at the teacher-by-year cell.

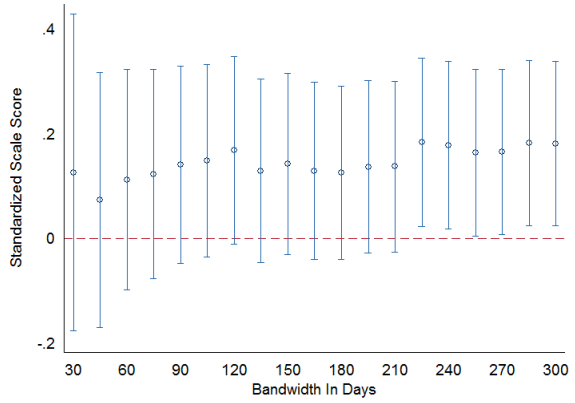


(a) Pr(Reading at Level C or Above) **(b) Pr(Reading at Level E or Above)**

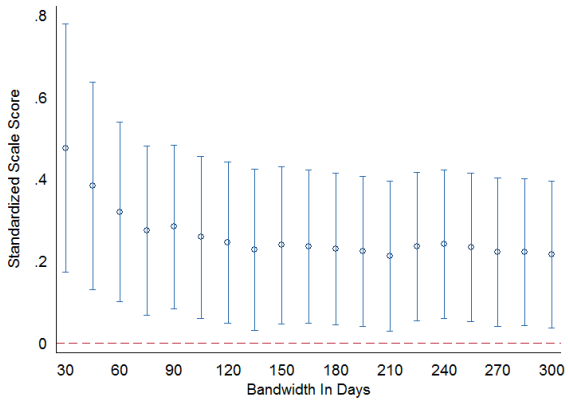


(c) Pr(Reading at Level I or Above)

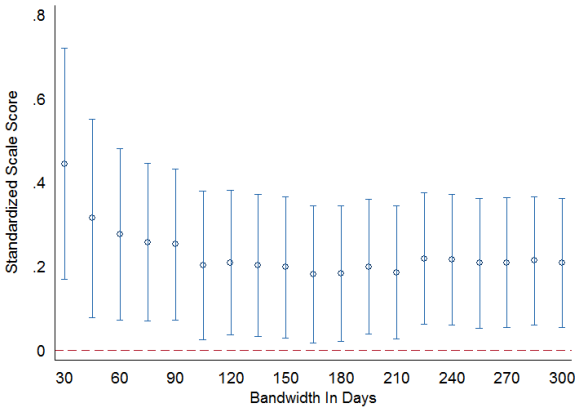
Figure A5: Robustness checks of fall first grade Fountas and Pinnell foundational literacy outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relevant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 5. Standard errors are clustered on the day of birth rating variable.



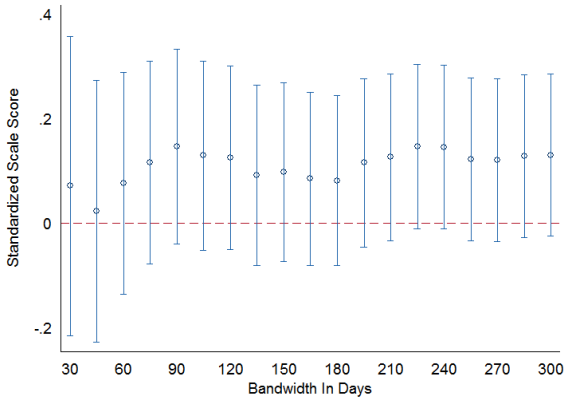
(a) Listening



(b) Reading

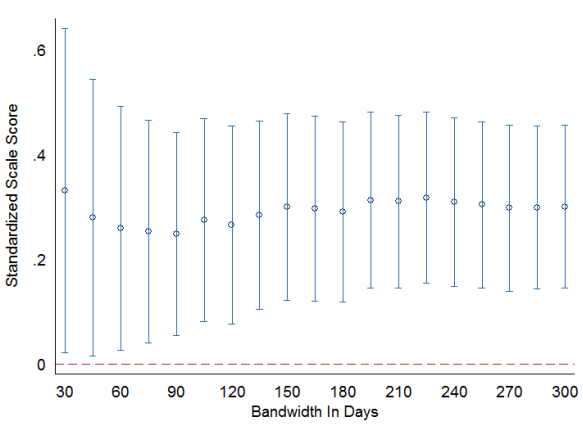


(c) Writing

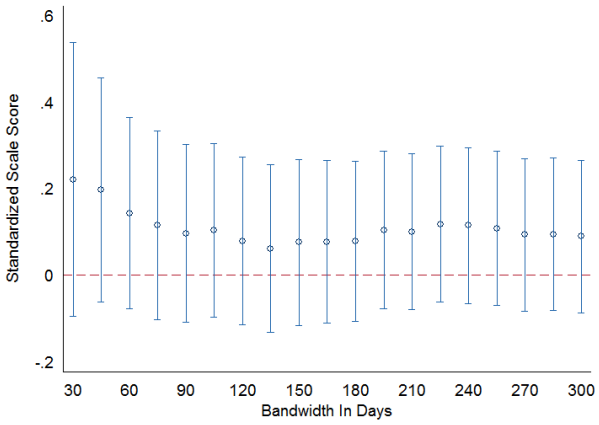


(d) Speaking

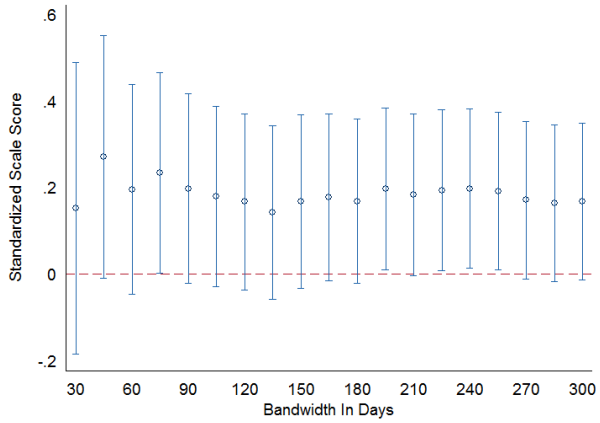
Figure A6: Auxiliary robustness checks of fall kindergarten CELDT subtest outcomes. Each dot represents a regression discontinuity estimate of the effect of Transitional Kindergarten on the relevant outcome for observations in bandwidths between 30 and 300 days. Dots represent point estimates and vertical lines represent the 95 percent confidence inteval. All regressions employ a linear spline functional form with covariates detailed in Table 5. Standard errors are clustered on the day of birth rating variable.



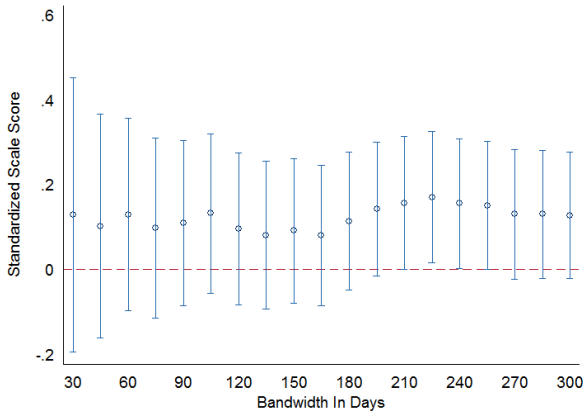
(a) Listening



(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 relevant 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 5. 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	0.010	0.001	0.011	0.056
	(0.017)	(0.019)	(0.028)	(0.037)
Missing Kindergarten Rhyming	-0.035	-0.026	0.011	-0.010
	(0.023)	(0.028)	(0.035)	(0.047)
Missing First Grade Fountas and Pinnell	0.019	0.035	0.070*	0.034
	(0.017)	(0.020)	(0.026)	(0.031)
Missing Kindergarten CELDT	0.032	0.059	0.083	-0.016
	(0.038)	(0.047)	(0.066)	(0.089)
Missing First Grade CELDT	-0.007	0.021	0.037	-0.037
	(0.040)	(0.050)	(0.074)	(0.105)
N	6,739	2,182	1,271	662

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: McCrary density test on baseline covariates

	Point Estimate (Standard Error)	N
Student Characteristics		
Female	0.048 (0.134)	3,316
Asian	0.026 (0.170)	2,095
Hispanic	0.119 (0.183)	1,683
White	-0.006 (0.211)	1,111
Other	0.254 (0.193)	1,179
Declined To State Ethnicity	0.218 (0.287)	660
Special Education	0.123 (0.339)	510
Limited English Proficient (LEP)	-0.019 (0.122)	3,310
Home Language:		
Chinese	0.000 (0184)	1,150
Spanish	0.049 (0.213)	1,005
English	0.148 (0.127)	4,020
Other	0.388 (0.268)	564
Dominant Language:		
Chinese	-0.075 (0.178)	1,387
Spanish	0.188 (0.227)	1,170
English	0.236+ (0.129)	3,412
Other	0.009 (0.248)	770
Test Characteristic		
Test Given In Spanish	0.147 (0.257)	945

Note: Each cell represents the results of a separate McCrary density test on the sample defined in the row headers. +indicates p<0.10, *p<0.05, **p<0.01

Table A3: Reduced form estimates of all 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	-0.141*	-0.181**	6,739	Overall Score	0.118	0.176*	3,310
	(0.059)	(0.042)			(0.110)	(0.079)	
Upper Case Letters	-0.289*	-0.332**	6,739	Listening	0.135	0.178*	3,310
	(0.133)	(0.087)			(0.105)	(0.080)	
Lower Case Letters	-0.229*	-0.163*	6,739	Speaking	0.067	0.132+	3,310
	(0.103)	(0.068)			(0.106)	(0.079)	
Letter Sounds	-0.130*	-0.184**	6,739	Reading	0.195*	0.216*	3,310
	(0.055)	(0.050)			(0.098)	(0.092)	
High Frequency Words	-0.099**	-0.141**	6,739	Writing	0.199+	0.210**	3,310
	(0.035)	(0.038)			(0.103)	(0.078)	
Early Literacy Behaviors	-0.161	-0.210**	6,739				
	(0.099)	(0.060)					
Initial Word Sounds	-0.157	-0.221*	6,739				
	(0.110)	(0.091)					
Rhyming	-0.164	-0.191*	5,997				
	(0.103)	(0.080)					
Blending	-0.033	-0.098*	6,427				
	(0.053)	(0.050)					
Pr(Mastering Required Found. Skills)	0.012	0.033	6,739				
	(0.022)	(0.021)					
Pr(Reading at Level A or Above)	0.020	0.014	6,739				
	(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	0.250**	0.231**	2,663
	(0.120)	(0.120)			(0.092)	(0.075)	
Pr(Reading at Level C or Above)	0.007	0.008	6,219	Listening	0.307**	0.301**	2,663
	(0.027)	(0.023)			(0.087)	(0.079)	
Pr(Reading at Level E or Above)	0.013	0.021	6,219	Speaking	0.145	0.128+	2,663
	(0.038)	(0.030)			(0.093)	(0.076)	
Pr(Reading at Level I or Above)	0.021	0.017	6,219	Reading	0.146	0.095	2,663
	(0.031)	(0.028)			(0.115)	(0.090)	
				Writing	0.234*	0.172+	2,663
					(0.110)	(0.092)	
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 3. 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 readin scale, 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 teacher-by-year fixed effect. +indicates p<0.10, *p<0.05, **p<0.01

Table A4: Reduced form incidence rate ratio estimates of fall kindergarten literacy outcomes

	(1)	(2)	(3)
Literacy Outcome	Incidence Rate Ratio	Avg Number of Items Missed by Control Group	Fewer Items Missed By TK-Eligible Students
Total Items Missed	0.835**	57.311	9.456
Upper Case Letters	0.718**	5.792	1.633
Lower Case Letters	0.850*	7.023	1.053
Letter Sounds	0.832**	12.92	2.171
High Frequency Words	0.869**	17.337	2.271
Early Literacy Behaviors	0.811**	2.705	0.511
Initial Word Sounds	0.802*	2.311	0.458
Rhyming	0.826*	4.120	0.717
Blending	0.907*	5.844	0.543
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 A3. 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 3. +indicates p<0.10, *p<0.05, **p<0.01

Table A5: Reduced form estimates of additional kindergarten and first grade Fountas and Pinnell outcomes by subgroup

Kindergarten		1st Grade		Kindergarten		1st Grade	
	(1)		(2)		(3)		(4)
Panel A: Full Sample, N= 6,739		*N=6,219*		*Panel F: White N=1,111*		*N=1,001*	
Pr(Reading at Level A or Above)	0.014	Pr(Level C or Above)	0.008	Pr(Reading at Level A or Above)	-0.033	Pr(Level C or Above)	0.031
	(0.016)		(0.023)		(0.056)		(0.052)
		Pr(Level E or Above)	0.021			Pr(Level E or Above)	0.039
			(0.030)				(0.089)
		Pr(Level I or Above)	0.017			Pr(Level I or Above)	0.151
			(0.028)				(0.097)
Panel B: Male, N=3,423		*N=3,144*		*Panel G: Other N=1,179*		*N=1,068*	
Pr(Reading at Level A or Above)	0.046*	Pr(Level C or Above)	0.018	Pr(Reading at Level A or Above)	-0.023	Pr(Level C or Above)	0.055
	(0.021)		(0.034)		(0.044)		(0.072)
		Pr(Level E or Above)	-0.021			Pr(Level E or Above)	-0.016
			(0.043)				(0.090)
		Pr(Level I or Above)	-0.010			Pr(Level I or Above)	-0.145+
			(0.041)				(0.075)
Panel C: Female, N=3,316		*N=3,075*		*Panel H: Limited English Proficient (LEP), N=3,310*		*N=3,115*	
Pr(Reading at Level A or Above)	-0.021	Pr(Level C or Above)	-0.017	Pr(Reading at Level A or Above)	0.016	Pr(Level C or Above)	-0.011
	(0.024)		(0.034)		(0.019)		(0.036)
		Pr(Level E or Above)	0.064			Pr(Level E or Above)	-0.057
			(0.047)				(0.045)
		Pr(Level I or Above)	0.039			Pr(Level I or Above)	-0.026
			(0.042)				(0.039)
Panel D: Asian, N=2,095		*N=2,017*		*Panel I: English Proficient N=3,429*		*N=3,104*	
Pr(Reading at Level A or Above)	0.023	Pr(Level C or Above)	0.049	Pr(Reading at Level A or Above)	0.012	Pr(Level C or Above)	0.027
	(0.028)		(0.035)		(0.026)		(0.032)
		Pr(Level E or Above)	0.004			Pr(Level E or Above)	0.093*
			(0.054)				(0.043)
		Pr(Level I or Above)	0.028			Pr(Level I or Above)	0.056
			(0.054)				(0.041)
Panel E: Hispanic, N=1,683		*N=1,546*					
Pr(Reading at Level A or Above)	0.024	Pr(Level C or Above)	-0.091				
	(0.024)		(0.065)				
		Pr(Level E or Above)	-0.022				
			(0.070)				
		Pr(Level I or Above)	0.018				
			(0.045)				

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. Linear probability models were used in all cases. All functional forms include a linear spline and covariates defined in Table 5. 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 A6: Robustness check: Placebo estimates of fall and midyear literacy outcomes

	(1)	(2)	(3)	**(4)**	(5)	(6)	(7)	
Panel A: Kindergarten Outcomes	c-50	B_{ict}-40	B_{ict}-30	**B_{ict}**	B_{ict}+30	B_{ict}+40	B_{ict}+50	N
Total Items Missed	-0.075	-0.085	-0.138+	**-0.181****	0.033	0.037	0.060+	6,739
	(0.112)	(0.083)	(0.071)	**(0.042)**	(0.033)	(0.032)	(0.031)	
Overall CELDT Score	-0.248	-0.100	0.157	**0.176***	0.042	-0.094	-0.055	3,310
	(0.253)	(0.123)	(0.118)	**(0.079)**	(0.075)	(0.073)	(0.069)	
Panel B: First Grade Outcomes								
Overall CELDT Score	-0.034	0.151	0.194	**0.231****	-0.006	-0.089	-0.031	2,663
	(0.225)	(0.137)	(0.122)	**(0.075)**	(0.077)	(0.077)	(0.078)	
Covariates	√	√	√	√	√	√	√	
Fixed Effects	√	√	√	√	√	√	√	

Note: Row headers indicate the outcome. Column headers indicate the number of days the original rating variable, B_{ict}, was translated. 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 A7: Robustness check: Estimates after eliminating heaps

	(1)	(2)	(3)	(4)	(5)
Panel A: Fall Kindergarten Outcomes	**Full Sample**	$H_B \leq 25$	$H_B \leq 20$	$H_B \leq 18$	$H_B \leq 15$
Total Items Missed	**-0.181****	-0.253**	-0.298**	-0.364**	-0.337**
	(0.042)	(0.050)	(0.068)	(0.076)	(0.114)
N	**6,739**	5,663	3,417	2,536	1,248
Overall CELDT Score	**0.176***	0.220*	0.179	0.287*	0.381+
	(0.079)	(0.092)	(0.120)	(0.134)	(0.212)
N	**3,310**	2,794	1,703	1,263	661
Panel B: Fall First Grade Outcomes					
Overall CELDT Score	**0.231****	0.268**	0.191	0.400**	0.296
	(0.075)	(0.093)	(0.136)	(0.137)	(0.219)
N	**2,663**	2,251	1,360	1,017	547

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 5, 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 5. The functional form of all regressions is a linear spline. +indicates p<0.10, *p<0.05, **p<0.01

Table A8: RD regressions of first stage by subgroup

Dependent Variable: Enrolled In TK in Year T-1		
	(1)	N
Full Sample	0.321**	6,739
	(0.027)	
ELL Sample	0.371**	3,310
	(0.041)	
Asian Sample	0.384**	2,095
	(0.054)	
Hispanic Sample	0.320**	1,683
	(0.058)	
White Sample	0.334**	1,111
	(0.072)	
Covariates	√	
Fixed Effects	√	

Note: All standard errors are clustered on the day of birth running variable. +indicates p<0.10, *p<0.05, **p<0.01