AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Impact of California's Transitional Kindergarten Program, 2013–14

DECEMBER 2015

Impact of California's Transitional Kindergarten Program, 2013–14

December 1, 2015

Karen Manship
Heather Quick
Aleksandra Holod
Nicholas Mills
Burhan Ogut
Jodi Jacobson Chernoff
Jarah Blum
Alison Hauser
Jennifer Anthony
Raquel González

2800 Campus Drive, Suite 200
San Mateo, CA 94403-2555
650.843.8100 | TTY 650.493.2209

www.air.org

4479_12/15

Contents

Acknowledgments

The authors would like to thank Dr. Deanna Gomby and Justine Choy from the Heising-Simons Foundation; Dr. Bernadette Sangalang, Dr. Meera Mani, and Dr. Jeff Sunshine from the David and Lucile Packard Foundation; and Dr. David Dodds and Camille Maben from First 5 California for their support of this study and their substantive input on the design and implementation of the study. We also are grateful for the advice and support of our Technical Advisory Group—Dr. Margaret Burchinal (University of North Carolina), Dr. Megan Franke (University of California at Los Angeles), Dr. Eugene Garcia (Arizona State University), Dr. Tim Shanahan (University of Illinois at Chicago), and Dr. Ross Thompson (University of California at Davis)—whose questions, suggestions, and feedback helped to strengthen the design and methodological approach to the study.

We also would like to extend our thanks to the district and school administrators, staff, and teachers who worked with us to make the data collection possible, and to the children and families who participated in the study. Without the contributions of all these individuals, the study would not have been possible. We also appreciate the support of Hope Street Family Center in Los Angeles, which opened their facilities to the study team for our field staff trainings in southern California.

We also want to acknowledge the many contributions of the other members of the study team, including staff at Survey Research Management (SRM)—Linda Kuhn, Julie Berry, Ashley Bronzan, Daniel Mackin, Celina Montorfano —who managed the on-site data collection activities, and the many field staff members who worked one-on-one with the students at their school sites. Thanks go to the members of the AIR team who worked tirelessly to recruit districts, schools, and families into the study and provide ongoing support throughout the data collection: Emily Anderson, Michele Cadigan, Connie Chandra, Linda Choi, Connie Conroy, Marie Dalldorf, Matthew Keuter, Shannon Keuter, Alejandra Martin, and Jennell McHugh. We appreciate the careful eyes of Elena Rein and CoCo Massengale in assisting with final review and proofreading of this report. Finally, we are grateful for the technical and substantive reviews of Johannes Bos, Angela Minnici, Susan Muenchow, Deborah Parrish, and Sara Wraight.

Funding for this study was provided by the Heising-Simons Foundation, the David and Lucile Packard Foundation, and First 5 California.

Executive Summary

Transitional kindergarten (TK)—the first year of a two-year kindergarten program for California children who turn 5 between September 2 and December 2—is intended to better prepare young five-year-olds for kindergarten and ensure a strong start to their educational career. To determine whether this goal is being achieved, American Institutes for Research (AIR) is conducting an evaluation of the impact of TK in California. The goal of this study is to measure the success of the program by determining the impact of TK on students' readiness for kindergarten in several areas. Using a rigorous regression discontinuity (RD) research design,[1] we compared language, literacy, mathematics, executive function, and social-emotional skills at kindergarten entry for students who attended TK and for students who did not attend TK. Overall, we found that TK had a positive impact on students' kindergarten readiness in several domains, controlling for students' age differences. These effects are over and above the experiences children in the comparison group had the year before kindergarten, which for more than 80 percent was some type of preschool program.

TK Improves Preliteracy and Literacy Skills

TK had a notable impact on students' literacy and preliteracy skills (Exhibit E-1). For example, children who attended TK were significantly better able to identify letters and words in kindergarten than their peers who did not attend TK (effect size = .502).[2] This advantage was equivalent to approximately five months of learning. Students who attended TK also had greater phonological awareness (an understanding of the sounds of letters and syllables that make up words) in kindergarten than did students who did not attend TK (effect size = .307). The advantage shown by students who attended TK on these skills, which are fundamental for learning to read, places them approximately three months ahead of their peers who did not attend TK. The effect of TK on expressive vocabulary was smaller and only marginally significant (effect size = .157; not shown), which is not unexpected; very few early literacy interventions have been successful in increasing children's vocabulary (Wasik, 2010).

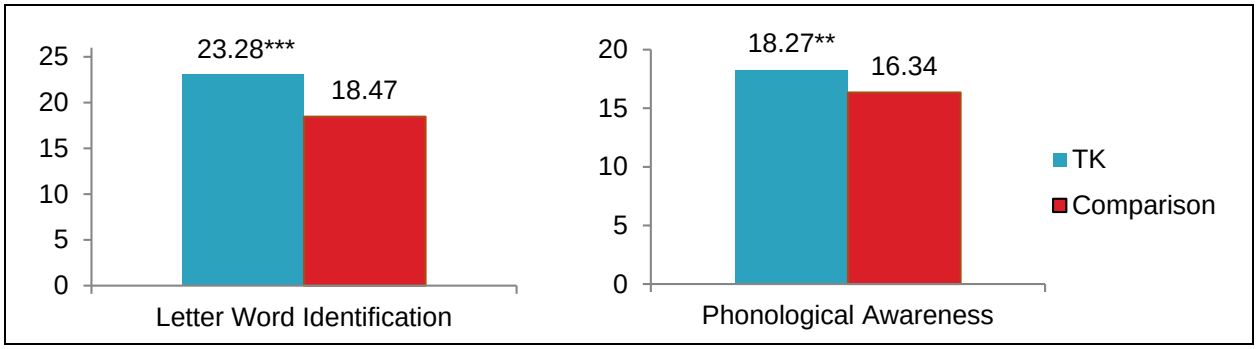
TK Improves Students' Mathematical Knowledge and Problem-Solving Skills

TK graduates also outperformed their peers who did not attend TK on measures of mathematics knowledge and skills (Exhibit E-2). In particular, TK participation improved students' knowledge of basic mathematical concepts and symbols (such as the equals sign) in kindergarten (Quantitative Concepts assessment, effect size = .356). Students who had attended TK also exhibited stronger mathematics problem-solving skills in kindergarten, such as counting objects,

[1] This study used an RD design to compare the outcomes of students with birthdates on either side of the December 2 cutoff date for TK eligibility. Students born on December 2 or earlier, who were eligible for TK, serve as the treatment group. Students who were too young to have qualified for TK (i.e., those born on December 3 or later) were the comparison group. These similarly aged children entered kindergarten at the same time as the TK students but without the TK experience. Because children's access to TK is determined by a specific birthdate cutoff (December 2), student and family characteristics that might otherwise influence participation in an education intervention, and thus bias the results (e.g., student learning needs, parent income or education, motivation to participate), did not drive eligibility. Birthdates cannot be manipulated by parents wanting to enroll their child. Thus, this analytical approach is a very strong research design, second only to a randomized controlled trial in which students are randomly assigned to participate in the TK program or not.

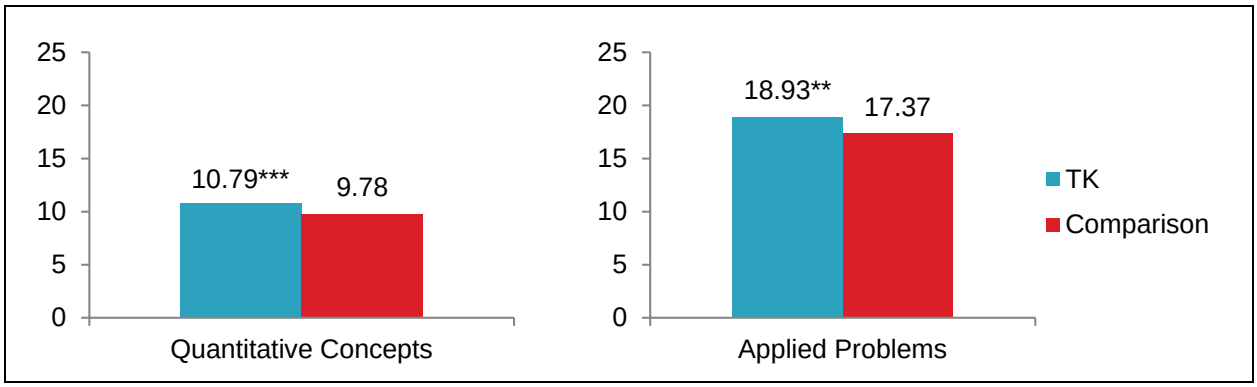
understanding measurement, conducting basic mathematical operations (such as addition and subtraction), and solving mathematical word problems, although the effect is somewhat smaller than for mathematical concepts and symbols (Applied Problems subtest, effect size = .260); this gave TK graduates a three-month advantage in learning over students who did not attend TK.

Exhibit E-1. Mean Scores for TK and Non-TK Students on Literacy and Preliteracy Measures[2]



** $p < .01$, *** $p < .001$
Note: Effect sizes: .502 for Letter-Word Identification and .307 for Phonological Awareness.
Source: Authors' analysis of student scores on the Woodcock-Johnson Letter-Word Identification test and the Clinical Evaluation of Language Fundamentals Phonological Awareness test.

Exhibit E-2. Mean Scores for TK and Non-TK Students on Mathematics Measures



** p < .01, *** p < .001
Note: Effect sizes: .356 for Quantitative Concepts and .260 for Applied Problems.
Source: Authors' analysis of student scores on the Woodcock-Johnson Applied Problems and Quantitative Concepts tests.

TK Supports Children's Behavioral Self-Regulation; No Detectable Impact on Social-Emotional Skills

Participation in TK gave students a relative advantage on executive function (effect size = .197) as well, meaning that TK graduates outperformed their peers on their ability to regulate their behavior, remember rules, and think flexibly—skills that support a solid foundation for school achievement (Schmitt, Pratt, & McClelland, 2014). The study did not find evidence that TK

[2] All means reported are adjusted for age, race/ethnicity, gender, English learner status, family income, students' eligibility for free and reduced-price lunch, parents' education, and students' participation in early education programs during the year before TK.

improved other aspects of students' social-emotional skills, however, such as increasing cooperation or engagement or decreasing problem behaviors (as reported by their teachers).

Conclusions and Next Steps

This study demonstrates that students who attended TK were better prepared for kindergarten than were similar students who did not attend TK, independent of age. We found that TK broadly benefited enrolled students, improving their reading and mathematics outcomes as well as their executive function. The effects we found are over and above the learning experiences comparison children received prior to entering kindergarten, which for more than 80 percent of the comparison group was some form of center-based preschool.

This unique approach to early education in California—which serves children in a narrow age range on elementary school campuses, with credentialed teachers holding bachelor's degrees and a curriculum aligned with kindergarten—appears to better prepare students for kindergarten than what they might have received in the absence of the program. It is important to note that this study reports results for one cohort of students—those participating in the second year of the rollout of TK (2013–14). Results for a second cohort of students who participated in the third year of TK (2014–15), now being collected, may differ as schools and districts refine their approach to implementing TK. Future analyses will investigate the extent to which the TK advantage is sustained through the end of kindergarten, for which groups of students TK is most beneficial, and which TK program characteristics are most supportive of student learning.

Chapter 1: Introduction

In 2010, Governor Arnold Schwarzenegger signed the Kindergarten Readiness Act into law, aligning California's kindergarten enrollment policy with the policies of most other states in the country, and then taking it one step further. With a kindergarten entry cutoff date of December 2, California has historically had young kindergartners, with up to a quarter of the state's kindergarten population entering school at age 4. The new law changed the kindergarten entry cutoff such that children must turn 5 by September 1 (instead of December 2) to enter kindergarten in that school year. In addition, the new law established a new grade level—transitional kindergarten (TK)—which districts must provide for students born between September 2 and December 2 and which is voluntary for families, as is kindergarten in California. With this new law, California makes a strong statement about the importance of early learning experiences, providing an additional year of early education to children affected by the rule change, with the goal of promoting their school readiness. This new program is intended to address achievement disparities between older and younger five-year-olds because those disparities can persist though kindergarten and into later years (Aunio, Heiskari, Van Luit, & Vuorio, 2015; Cannon & Karoly, 2007).

To determine whether TK is effective at improving school readiness and learning outcomes for children, American Institutes for Research (AIR) is conducting an evaluation of the impact of TK in California. The goal of this study is to assess the impact of TK on California students' readiness for kindergarten across multiple domains of development critical for success in school. Using a regression discontinuity (RD) design, this study examines whether TK participation improves kindergarten readiness in the domains of early literacy and language, mathematics, executive function, and social-emotional skills.

Background

TK, as defined in the Kindergarten Readiness Act, is the first year of a two-year kindergarten program. This is an innovative approach to early education with little existing research to help us anticipate its efficacy. There are no existing evaluations of TK programs per se, for other states do not offer this program. However, early research on two-year kindergarten programs (Ferguson, 1991; Karweit & Wasik, 1992), including developmental kindergarten and transitional first-grade programs, found no effect on children's elementary school outcomes. This research is now dated, however; it summarizes results from kindergarten programs that predated the push for school accountability under the No Child Left Behind Act of 2001 and the increase in attention to academics in kindergarten that has been observed over time (Bassok, Latham, & Rorem, 2015; Walston & Flanagan, 2013). In addition, developmental kindergarten and transitional first-grade programs are qualitatively different from California's TK program in that these programs target students with social or academic difficulties. Because TK is a program available to all children in the specified age window, regardless of academic ability, its impact may differ from that of programs targeted on at-risk children. A large body of relevant research exists, however, documenting the impact of prekindergarten programs, as well as the effects of repeating kindergarten, both of which are experiences that may be somewhat comparable to TK.

Overall, research has shown that participation in high-quality preschool prior to kindergarten can improve young children's readiness skills for elementary school, positively affecting behavioral, social-emotional, and cognitive outcomes (Andrews, Jargowsky, & Kuhne, 2012; Barnett, 1995; Yoshikawa et al., 2013). In particular, for children who may be at risk for academic challenges in early elementary school, attending a high-quality preschool can improve test scores and attendance and reduce future grade-level retention and placement in special education (Andrews et al., 2012; Barnett, 2008; Karoly & Bigelow, 2005; Reynolds et al., 2007). Thus, as a "prekindergarten" program, there also is potential for TK to affect school readiness, especially if the students' TK program is of high quality.

And on some measures, TK is, by definition, a high-quality early education program. For example, TK teachers in general would be considered more highly qualified than typical preschool teacher qualifications, as TK teachers in California are required by law to be credentialed and hold at least a bachelor's degree. Despite some inconsistent findings, there is evidence that teachers' level of education and teacher pay are both positively related to student outcomes. In fact, the preschool programs that have shown long-term gains for their students in research studies all were staffed by teachers who held bachelor's degrees and whose compensation was similar to that of public school teachers (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Pianta, Barnett, Burchinal, & Thornburg, 2009; Whitebook, Gomby, Bellm, Sakai, & Kipnis, 2009), like TK. Thus, we hypothesize that TK teachers, being well educated and better compensated than most preschool teachers, may help their students achieve better school readiness outcomes than students who did not attend TK the year before kindergarten, even if those students attended a preschool program instead.

In addition, research indicates that the effects of preschool are better supported if curricula and instructional strategies from prekindergarten through Grade 3 are well aligned (Bogard & Takanishi, 2005; Brooks-Gunn, 2003). As the first year of a two-year kindergarten program, TK is co-located in elementary schools with kindergarten and other early elementary classrooms, the majority of TK teachers have taught kindergarten (Quick et al., 2014), and TK teachers are asked to use California's kindergarten standards as their guide. As a result, there is likely to be more alignment between TK and the school's K–3 experience than between other early education programs and the K–3 experience. This close alignment may help TK be more successful in increasing students' kindergarten readiness.

TK may, in fact, be more like kindergarten than a typical preschool program, in that children attending TK receive greater exposure to kindergarten-like experiences. However, an early look at TK in California (Quick et al., 2014) suggested that TK students spend less time on academic subjects (e.g., reading and language arts) than kindergarten students do and more time on developmentally appropriate activities such as social-emotional learning and child-led exploration—the kinds of activities that one might expect to see in a high-quality preschool program. The emphasis on academics may still, however, be greater than what children would otherwise receive in preschool or at home, because of the proximity to and alignment with statewide public kindergarten programs. Overall, it is unclear how a potentially more academic program for prekindergarten children might affect their school readiness and future outcomes; researchers and educators disagree about the right balance of academic and nonacademic content in kindergarten (Duncan, 2011; Elkind & Whitehurst, 2001; Zigler, 1987; Zigler & Bishop-Josef, 2006), debates that extend to TK as well. Although critics stress that a heavy academic focus in

kindergarten may not be developmentally appropriate (Datar & Sturm, 2004; Raver & Knitzer, 2002; Shonkoff & Phillips, 2000; Stipek, 2006), there is evidence that exposure to advanced academic content in kindergarten may lead to greater student learning (Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Engel, Claessens, Watts, & Farkas, 2015).

TK students also have, in effect, two years of kindergarten, which, in some ways, is like repeating (being retained in) kindergarten. Research on students who are retained in kindergarten shows that short-term outcomes may improve, but those gains are not maintained over the long term; specifically, Hong and Raudenbush (2005) found that kindergarten retention does not improve outcomes for retained students but, rather, that these retained students learn less. Similarly, Mantzicopoulos and Morrison (1992) found that although there were some positive effects on behavioral problems for retained kindergarten students, their academic outcomes did not improve. Students are retained for particular reasons, however, including behavioral issues or learning disabilities, and these reasons limit the relevance of these findings for TK students and their experiences.

On balance, the larger body of evidence presented here suggests that TK—as an additional year of high-quality early learning experience—should support positive educational outcomes for students.

Chapter 2: Methods

As detailed below, this study estimated the impact of the TK program by comparing a range of school-readiness outcomes for 2,864 kindergartners in the 2014–15 school year, approximately half of whom had access to TK (because they turned 5 before December 2 in the prior school year) and half of whom did not (because they turned 5 after the December 2 cutoff). Twenty California school districts and 164 elementary schools participated in the study. These districts and schools were sampled to be broadly representative of California and were drawn from all geographic regions of the state. (See Appendix A for details of the study's sampling approach.) Exhibit 1 shows that the background characteristics of the student sample participating in the study were similar to those of California kindergartners overall.

Exhibit 1. Characteristics of the TK Study Sample Compared to the California Kindergarten Population (Where Available)

	Percentage of Students	
	Sample ***n*** **= 2,864**	**California** ***n*** **= 511,985**
Female	50.0%	48.2%
Race/ethnicity		
White	26.0%	23.2%
Hispanic	55.9%	55.5%
Asian	10.9%	8.1%
Black	4.1%	5.3%
Other ethnicity	3.1%	7.9%
Eligible for free or reduced-price lunch	58.9%	59.4%[1]
English learner	41.9%	35.2%
Spanish home language	36.9%	NA
Special education	6.9%	7.1%
Parental education		
Less than high school diploma	13.0%	19%
High school diploma	20.2%	23%
Some college	17.1%	24%
Vocational certificate or AA	17.8%	NA
College degree	17.6%	20%
Graduate education	14.3%	13%

Sources: Authors' analysis of statewide student data for academic year 2014–15 obtained through DataQuest (http://data1.cde.ca.gov/dataquest/), student record data from participating districts, and parent survey data.
Notes: [1]The most recent year of free and reduced-price lunch data available in DataQuest is 2013–14, and it is not available by grade level. Comparison data for parental education is available only for parents of all students K–12 statewide; data are not available for vocational certificates or AA degrees: http://api.cde.ca.gov/Acnt2013/2013GrthStAPIDC.aspx?allcds = 0000000.

We also examined the characteristics of students who were eligible for TK and those who were not to ensure that, after controlling for the age difference between TK and comparison group students, there were no notable differences between these two groups that might drive differences in achievement. In terms of demographic characteristics (Exhibit 2), although slightly more female students were eligible for TK, simply by virtue of the way births were distributed in 2009, there were no other significant differences between students eligible for and students not eligible for TK. To account for any minor differences, we controlled for demographic characteristics, including age, in the RD models.

Exhibit 2. Demographic Characteristics of Students in the TK and Comparison Samples

	TK Group *n* = 1,562	Comparison Group *n* = 1,302
Mean age (as of 9/1/2014)	5.83***	5.66
Female	51.0%*	48.8%
Race		
White	26.0%	28.0%
Hispanic	55.4%	56.6%
Black	4.4%	3.9%
Asian	12.0%	9.6%
Other ethnicity	2.3%	1.9%
Free and reduced-price lunch eligibility	59.1%	58.6%
English learner	43.5%	39.9%
Special education	7.0%	6.7%
Parental education		
Less than high school diploma	12.5%	13.6%
High school diploma	19.9%	20.5%
Some college	16.7%	17.6%
Vocational certificate or AA	17.3%	18.4%
Graduated from college	18.9%	16.1%
Graduate education	14.7%	13.8%

*** $p < .001$, * $p < .05$
Source: Authors' analysis of student record data from participating districts and parent survey data.
Note: Table displays unadjusted means and percentages, but the significance testing for all variables except age adjusts for student age.

In addition, we considered prior early education experiences among TK and comparison students (Exhibit 3). First, as context for our findings, it is important to note that more than 80 percent of students in the comparison group attended some type of center-based preschool program the year before kindergarten (while TK students were enrolled in TK), according to parent reports. And, half of all students in the comparison group attended their preschool program for at least 15 hours per week (roughly equivalent in duration to part-day TK).

Exhibit 3. Prior Preschool Experience of Students in the TK and Comparison Samples

	TK Group *n* = 1,562	Comparison Group *n* = 1,302
Attended center-based preschool in the year before kindergarten	N/A (Attended TK)	81.2%
Attended center-based preschool in the year before kindergarten **for at least 15 hours per week**	N/A (Attended TK)	64.5%
Attended center-based preschool **2 years** before kindergarten	76.6%***	49.9%
Attended center-based preschool **2 years** before kindergarten **for at least 15 hours per week**	45.1%***	33.8%

*** $p < .001$
Source: Authors' analysis of parent survey data.
Note: Table displays unadjusted percentages, but the significance testing adjusts for student age.

Many of these students also attended a center-based preschool program *two years* before kindergarten as well. However, more TK-eligible students attended a center-based preschool program two years before kindergarten (in the year before they attended TK) than students in the comparison group (Exhibit 2). In general, these early education experiences were not intensive; only 45 percent of TK and 34 percent of comparison students attended a preschool program for at least 15 hours per week two years prior to kindergarten. However, to account for this difference, we controlled for prior preschool experience in the RD models.

Introduction to the Study Design

To measure the effect of TK, relative to "business as usual" (how similarly aged children would have progressed without the additional year of education), researchers would ideally randomly assign children to be either in TK or to continue with business as usual, which could include child care, preschool, Head Start, or remaining at home. However, such assignment would be difficult to defend and implement, and it would produce results that are not necessarily generalizable to the full population of TK-eligible children (because they would be limited to children whose parents would be comfortable with the uncertainty inherent in a randomized controlled trial setting). Fortunately, eligibility for TK is limited to children in a very specific age range, which means that a regression discontinuity (RD) design can be used to approximate the rigor and credibility of random assignment without actually randomly assigning children.

This study takes advantage of this birthdate cutoff and limited age range and employs the RD design. Students born between October 1 and February 2 (within 60 days on either side of the December 2 cutoff date to enter TK) in sample districts and schools were invited to participate in the study by consent of their parents; participation was voluntary. We then compared the academic and social kindergarten readiness of students who attended TK with the readiness of those who did not, as determined by the birthdate cutoff. In all of the impact analyses, we statistically controlled for student age, which is the only baseline variable on which TK and comparison group students varied by design. (That is, all TK students were somewhat older than all comparison group students).

Data are being collected from two cohorts of students: those who entered kindergarten in the fall of 2014 and those who entered in the fall of 2015. Both cohorts include students who were eligible for TK and those who were not. Findings in this report are based on the first cohort of students. When data from both cohorts are available, they can be combined so that the total sample size of students is large enough to allow the research team to do additional analyses, including examining the impact of TK on subgroups of students (such as English learners) and identifying the particular characteristics of TK classrooms that are most supportive of positive outcomes for students. Future reports will present findings from these analyses.

Data Sources

Information about students' skills in kindergarten was obtained from both direct student assessments and surveys of kindergarten teachers, who rated students' behaviors and social skills. Student background information was gathered from school districts and with a parent survey.

Student Assessments

Direct assessments of students, in English and Spanish, were the primary source of information about students' kindergarten readiness. Trained assessors administered the assessments in the participating schools between October 2014 and January 2015. For students speaking Spanish at home, the results of an English language screener determined whether English or Spanish would be used for the primary assessment. Regardless of primary language, all Spanish-speaking students were administered two assessments (Woodcock-Johnson Applied Problems and CELF Expressive Vocabulary, described below) in both English and Spanish. In addition, the Head Toes Knees Shoulders measure (Ponitz et al., 2009) of executive function was translated into the five most common Asian languages in the study's participating districts: Cantonese, Korean, Mandarin, Tagalog, and Vietnamese; and the translated version was used for students who spoke one of these languages at home and did not pass the English language screener. These assessments are described in more detail below.

English Language Screener. The receptive and expressive language subtests of the *preLAS 2000* (De Avila & Duncan, 2000)—Simon Says and Art Show—were used to assess students' English proficiency. If students demonstrated sufficient proficiency in English (scoring at least 12 out of 20 points on the two subtests), they were given the full assessment battery in English, and, in addition, a shorter supplemental assessment in Spanish. Spanish-speaking students who did not score at least 12 points were given the full assessment in Spanish with a supplement in English. In this way, all Spanish-speaking dual language learners were assessed on a core set of measures—mathematics and vocabulary—in both languages. Students who did not demonstrate English proficiency and spoke one of the five most common Asian languages in the study districts were given only the Head Toes Knees Shoulders task in their home language.

Language and Literacy. It is critical for students to develop early language and literacy skills in kindergarten to support their success later in school, making these skills we wanted to assess in students for this study. Before entering kindergarten, few students can read independently, but they possess many of the language skills required for mastery of reading (Reaney & Kruger, 2002). Knowledge of the alphabet, commonly used sight words, vocabulary, and awareness of

word sounds, also called phonological awareness, all are foundational skills for reading (Reaney & Kruger, 2002). Students' alphabet knowledge and phonological awareness prior to kindergarten entry are strongly related to later literacy skills, such as reading comprehension, spelling, and fluency (Kjeldsen, Kärnä, Niemi, Olofsson, & Witting, 2014; National Early Literacy Panel, 2008). Furthermore, recognizing sight words quickly and without difficulty facilitates children's fluency in reading (Ehri, 2015). Numerous studies also link vocabulary to reading achievement, as well as to achievement in other academic areas (Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Neuman & Dwyer, 2009; Reaney & Kruger, 2002; Kjeldsen et al., 2014; Wasik, 2010). For example, having a larger vocabulary enables young children to know and use more words and phrases representing abstract mathematical concepts, which can help to facilitate the understanding of those concepts (Morgan et al., 2015).

To measure these critical language and literacy skills, we selected a set of three widely used validated assessments that are available in English and Spanish. First, the Woodcock-Johnson III Letter-Word Identification subtest (and its equivalent in the Spanish-language Batería III (Woodcock-Muñoz) measured students' ability to name letters and read common words. Second, the Expressive Vocabulary subtest of the Clinical Evaluation of Language Fundamentals—Preschool 2 (CELF–2P) assessment measured students' word knowledge by asking them to name pictures and to describe the actions depicted in the pictures. Third, the Phonological Awareness subtest of the CELF–2P measured students' awareness of the sounds of language, including the rhythm of speech and rhyming sounds.

Mathematics. Early mathematics skills also are critical for students' later academic success. Kindergarten students' number competence sets the foundation for later mathematics comprehension and is predictive of mathematical achievement in third grade (Clements & Sarama, 2014; Duncan et al., 2007; Jordan, Kaplan, Ramineni, and Locuniak, 2009). For this study, two subtests of the Woodcock-Johnson III and their equivalent in the Woodcock-Muñoz were administered to gather data on students' mathematical skills and knowledge. The Woodcock-Johnson III Quantitative Concepts subtest assesses students' understanding of the number line, recognition of mathematical symbols, and understanding of various mathematical representations. The Applied Problems subtest assesses students' quantitative reasoning and mathematical knowledge, such as counting, basic operations (such as addition or subtraction), and problem solving.

Executive Function. Executive function is a set of mental skills that allows children to plan, manage their time, regulate their behavior, and think flexibly. Behavioral regulation is "the manifestation of executive function skills in overt, observable responses in the form of children's gross motor actions" (Ponitz, McClelland, Matthews, & Morrison, 2009). Researchers have noted that behavioral self-regulation facilitates children's adjustment to school, ability to benefit from learning experiences, and success in social interactions (Ponitz et al., 2009), making this another critical skill to measure for this study. For example, attentional focus permits children to attend to the teacher and focus on school tasks. Working memory is essential for remembering multiple-step teacher instructions. Last, inhibitory control helps children to control their behavior, such as remembering to raise their hands before answering (Ponitz et al., 2009). Thus, behavioral self-regulation provides a solid foundation for school achievement when taken together with early academic skills (Schmitt, Pratt, & McClelland, 2014).

In this study, the Head Toes Knees Shoulders task (McClelland & Cameron, 2012) was used to assess students' executive function, including working memory, inhibition, and cognitive flexibility. Assessors asked students to touch a part of their body different from the one named during a Simons Says–type game (e.g., when the assessor says to touch their head, they should touch their toes), and to remember to adjust to the changing rules of the game. The test items are scored for either full credit if the student goes immediately to the correct body part or partial credit if the student starts to go to the wrong body part and then self-corrects by inhibiting his or her initial impulse.

Teachers' Assessments of Student Social Skills

We also collected information about students in the study from their teachers. Specifically, we gathered teachers' assessments of students' social-emotional skills using the Social Skills Improvement System (SSIS) Rating Scales (Gresham & Elliott, 2008), a valid and reliable tool commonly used to assess students' behavior in elementary schools. We asked teachers to rate students on items that aggregated into five subscales of the instrument: cooperation, engagement, self-control, internalizing behavior,[3] and externalizing behavior.[4] Teachers were asked to rate students on the items using a four-point Likert scale. Teachers provided data for 82 percent of students participating in the study.

Student Demographic Characteristics

Student age is the primary selection variable in this study. It is important to note that the difference in average age between TK and comparison students in this study is only about two months (TK group: 5.83 years; comparison group: 5.66 years). We have controlled for age of the students when estimating the effect of TK on student outcomes. This means that when estimating the effect of TK, we are holding age constant between the TK and comparison groups and observing the unique effect of TK on student outcomes. In other words, when estimating the impact of TK, we are eliminating the differences in the outcome between the TK and comparison groups that are due to differences in age.

In addition to students' age, the impact analyses controlled for a comprehensive set of student background variables that may be associated with TK attendance and student outcomes, including race or ethnicity, gender, English learner status, family income, students' eligibility for free and reduced-price lunch, parents' education, and students' participation in early education programs two years before kindergarten (the year before TK), to account for students' different early learning experiences. Parents reported their education level, the family income, and the students' participation in other early education programs on a brief survey that accompanied the consent form. Other demographic information was requested for all consented students from study districts.

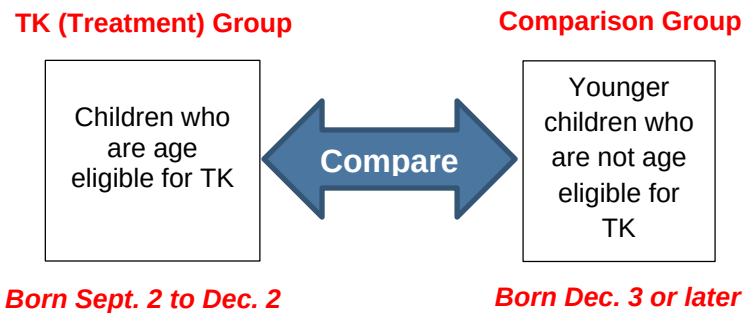
[3] Internalizing behaviors are problematic behaviors that are directed toward the self, such as depression or social withdrawal.

[4] Externalizing behaviors are problematic actions directed at others, such as aggression or defiance.

Analytic Approach

As mentioned earlier, this study uses an RD design to compare the outcomes of students with birthdates on either side of the December 2 cutoff date for TK eligibility, as shown in Exhibit 4, controlling for students' age. TK-eligible students born on December 2 or earlier (i.e., two months before December 2), who are eligible for TK, serve as the treatment group. Students who are too young to have qualified for TK (i.e., those born on December 3 or later, or two months after December 2) are the comparison group. These children, similar in age to TK students, will enter kindergarten at the same time as the TK students, but without the TK experience.

Exhibit 4. The Regression Discontinuity Approach



Because children's access to TK is determined by a specific birthdate cutoff (December 2), student and family characteristics that might otherwise influence participation in an education intervention, and thus bias the results (e.g., student learning needs, parent income or education, motivation to participate), do not drive eligibility. Birthdates cannot be manipulated by parents wanting to enroll their child. Thus, this analytical approach is a very strong research design, second only to a randomized controlled trial in which students are randomly assigned either to participate in the TK program or not. Assuming children's birthdates are randomly distributed, the comparison between the students with birthdates before and after the cutoff date can be likened to such a randomized experiment, once the students' ages are controlled for in the analysis. In addition, our analyses control for a broad set of student background characteristics. Therefore, differences seen at the beginning of kindergarten between students who did and did not attend TK can be attributed to the TK program.

Appendix A describes the RD approach in detail, including a number of important diagnostic and sensitivity analyses implemented to ensure that the findings presented here are valid and not overly sensitive to our model specification and other analytical decisions.

Chapter 3: Results

Students who attended TK had more advanced literacy, mathematics, and executive function skills at kindergarten entry than did their peers who did not attend TK. As discussed further in this chapter, the advantage conferred by TK participation was up to approximately five months of learning. Thus, at kindergarten entry, students who attended TK were up to half a school year ahead of their peers who did not attend TK, many of whom attended other early education programs. Although TK students are slightly older than the comparison students (by about two months, on average), we have controlled for age when estimating the effect of TK, which eliminates the differences in the outcomes between the TK and comparison groups that are due to differences in age.

Exhibits 5–11 display the estimated effects in terms of adjusted mean differences between students who attended TK and those who did not, taking age and other demographic characteristics into account.[5] The asterisks on the bars indicate where there are statistically significant differences between TK and comparison student scores. Comparisons with lighter shaded bars are not statistically different from each other. Effect sizes, a standardized measure of impact that helps to assess the magnitude of changes observed in a study sample beyond statistical significance, are also reported. Exhibits 12 and 13 present a summary of effect sizes across all measures.

Language and Literacy

Participation in TK improved students' language and literacy skills to a significant degree (Exhibits 5, 6, and 7). The impact of TK on these outcomes ranged from an effect size of .157 to .502.[6] Specifically, children who attended TK displayed greater skills in identifying letters and words, as measured by the Woodcock-Johnson Letter-Word Identification subtest (effect size = .502; $p < .001$); this advantage was equivalent to approximately 5.0 months of learning. TK attendees also had greater awareness of letter sounds and rhyming than did students who did not attend TK, as measured by the CELF phonological awareness measure (effect size = .307, $p < .01$).

The effect of TK on vocabulary, as measured by the CELF Expressive Vocabulary subtest, was smaller and only marginally significant (effect size = .157, $p < .10$),[7] which is not unexpected; very few early literacy interventions have been successful in increasing children's vocabulary (Wasik, 2010), perhaps because students who enter school with larger vocabularies are primed to continue

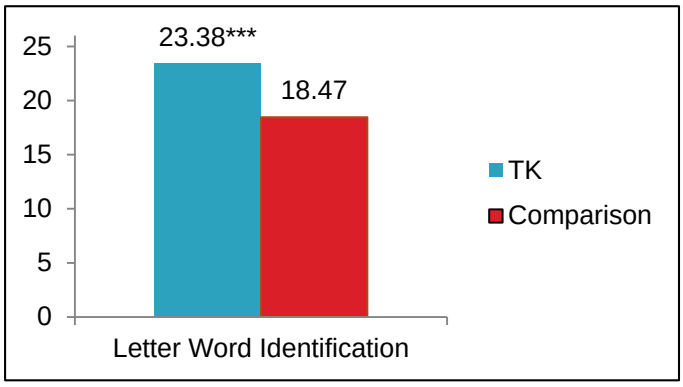
[5] Adjusted means are the model-predicted means computed from fuzzy RD models and controls for age, TK participation, and all the student demographic variables in the model. The adjusted means shown in the exhibits are computed using the student level predicted outcomes from the fuzzy RD models, aggregated over children's TK attendance status. Because the summation of scores over TK attendance status cannot take noncompliance in eligibility into account, the difference between the adjusted means might not always add up to the effect sizes estimated through the fuzzy RD models.

[6] Effect sizes are the standardized mean differences in the outcomes between the students who attended TK and those who did not as estimated by the RD model and computed by dividing the mean difference in the outcome by overall standard deviation. Effect sizes of 0.2 are considered small, 0.5 moderate, and 0.8 high.

[7] Note that this study was designed to have a minimum detectable effect size of .20. Thus, differences of this magnitude would not be expected to be statistically significant at the .05 significance level with the sample size that this study has.

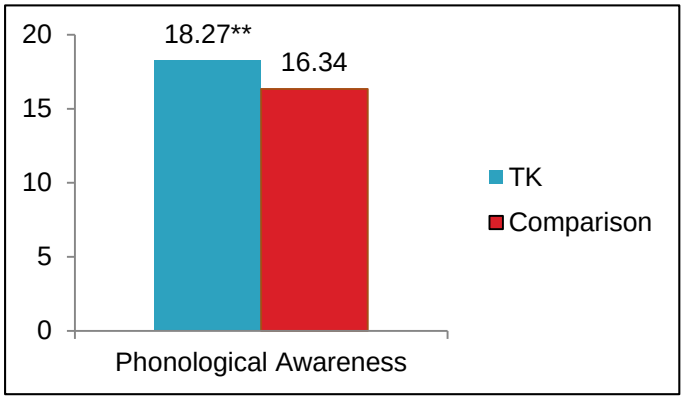
building those vocabularies, much more so than students who enter school knowing fewer words, making it difficult for educational interventions to make a large impact on vocabulary.

Exhibit 5. Adjusted Means for TK and Comparison Students on Letter Word Identification



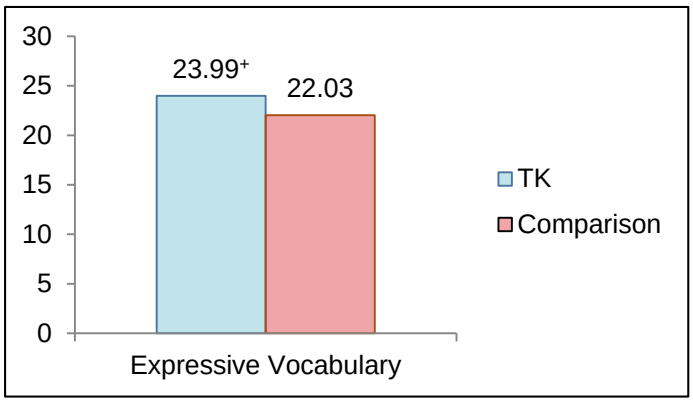
*** p < .001
Note: Effect size: .502
Source: Authors' analysis of student scores on the Woodcock-Johnson Letter-Word test.

Exhibit 6. Adjusted Means for TK and Comparison Students on Phonological Awareness



** p < .01
Note: Effect size: .307
Source: Authors' analysis of student scores on the Clinical Evaluation of Language Fundamentals Phonological Awareness test.

Exhibit 7. Adjusted Means for TK and Comparison Students on Expressive Vocabulary

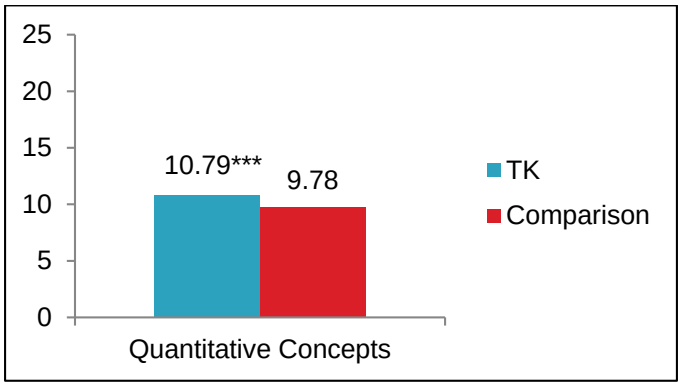


+ p < .1
Source: Authors' analysis of student scores on the Clinical Evaluation of Language Fundamentals Expressive Vocabulary test.

Mathematics

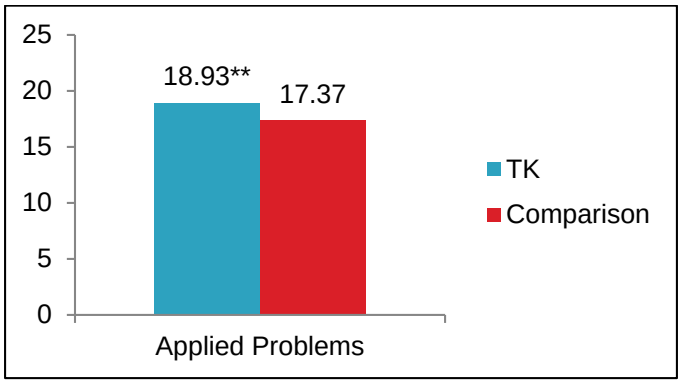
TK graduates also outperformed their peers who did not attend TK on measures of mathematical knowledge and skills. In particular, TK participation improved students' knowledge of basic mathematical concepts and symbols in kindergarten, as measured by the Woodcock-Johnson Quantitative Concepts subtest (effect size = .356, p< .001; Exhibit 8). Although the effect is smaller in magnitude, students who had attended TK also exhibited stronger mathematics problem-solving skills at the beginning of kindergarten, such as counting objects, understanding measurement, conducting basic mathematical operations, and solving mathematics word problems, as measured by the Woodcock-Johnson Applied Problems subtest (effect size = .260, p< .01; Exhibit 9); this gave TK graduates a three-month advantage in learning over students who did not attend TK.

Exhibit 8. Adjusted Means for TK and Comparison Students on Quantitative Concepts



*** *p* < .001
Note: Effect size: .356
Source: Authors' analysis of student scores on the Woodcock-Johnson Quantitative Concepts test.

Exhibit 9. Adjusted Means for TK and Comparison Students on Applied Problems



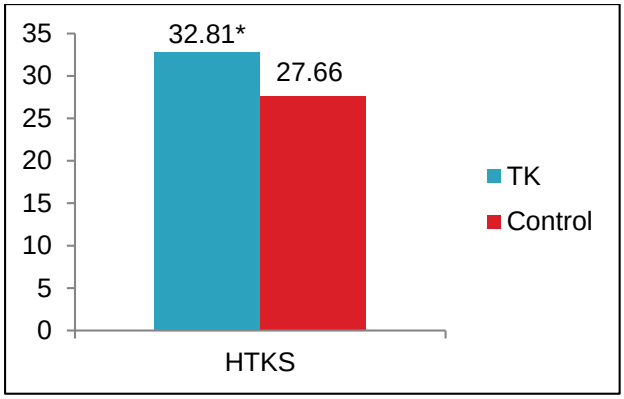
** *p* < .01
Note: Effect size: .260
Source: Authors' analysis of student scores on the Woodcock-Johnson Applied Problems test.

The stronger impact on the Quantitative Concepts subtest, which assesses students' understanding of mathematical concepts and symbols, suggests greater exposure to this basic mathematics content in TK than in other early learning and care environments experienced by the non-TK students.

Executive Function and Social-Emotional Skills

Analyses for social-emotional outcomes yielded fewer statistically significant results. We did find a modest, but statistically significant, impact on students' executive function skills—comprising self-regulation, working memory, and cognitive flexibility (effect size = .197; $p < .05$) (Exhibit 10). TK students' five-point advantage on the HTKS executive function measure is similar to or greater than the point gains observed during the kindergarten year in other studies (Ponitz et al., 2009), thus reflecting a notable advantage from TK. It may be the additional time that TK offers students to participate in a school-based classroom environment with norms and routines—which requires them to inhibit their impulses at times, follow instructions, and adapt to different tasks—that gives TK students the opportunity to develop executive function skills.

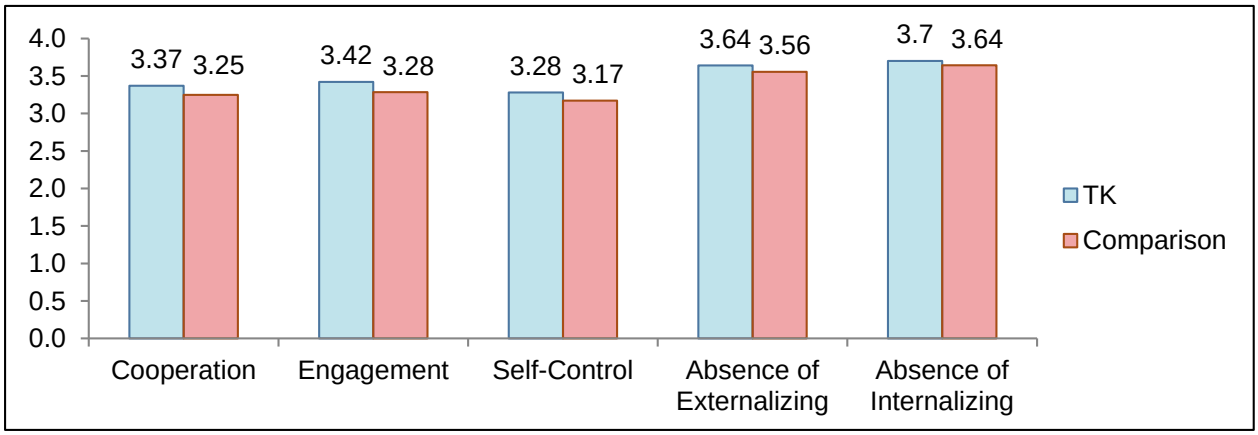
Exhibit 10. Adjusted Means for TK and Comparison Students on Executive Function



* $p < .05$
Note: Effect size: .197
Source: Authors' analysis of student scores on the Head Toes Knees Shoulders task.

Students who attended TK were not, however, rated by their kindergarten teachers as having significantly better behavior than comparison students on any of the five SSIS subscales examined (Exhibit 11). It may be that students' social skills and behaviors are similarly supported in TK and in other types of prekindergarten programs that the comparison group attended. It also may be that the four-point SSIS rating scale does not provide teachers enough rating options to effectively differentiate students' behavior and social skills.

Exhibit 11. Adjusted Means for TK and Comparison Students on Teacher Ratings of Social-Emotional Skills



No statistically significant differences
Source: Authors' analysis of teacher responses on the Social Skills Improvement System (SSIS) Rating Scales.

Summary of Impact

Exhibits 12 and 13 summarize the impact of TK on different student outcomes at the beginning of kindergarten using effect sizes—a standardized measure that allows us to compare the magnitude of effects across student outcome measures. Lighter blue bars indicate differences between TK and comparison students that were *not* statistically significant, as described above. As shown below, we observed positive effects of TK participation for students across the range of literacy and mathematics outcomes as well as in executive function, with the largest effect for skills in identifying letters and words.

Exhibit 12. Effect Sizes for Language, Literacy, and Mathematics Outcomes

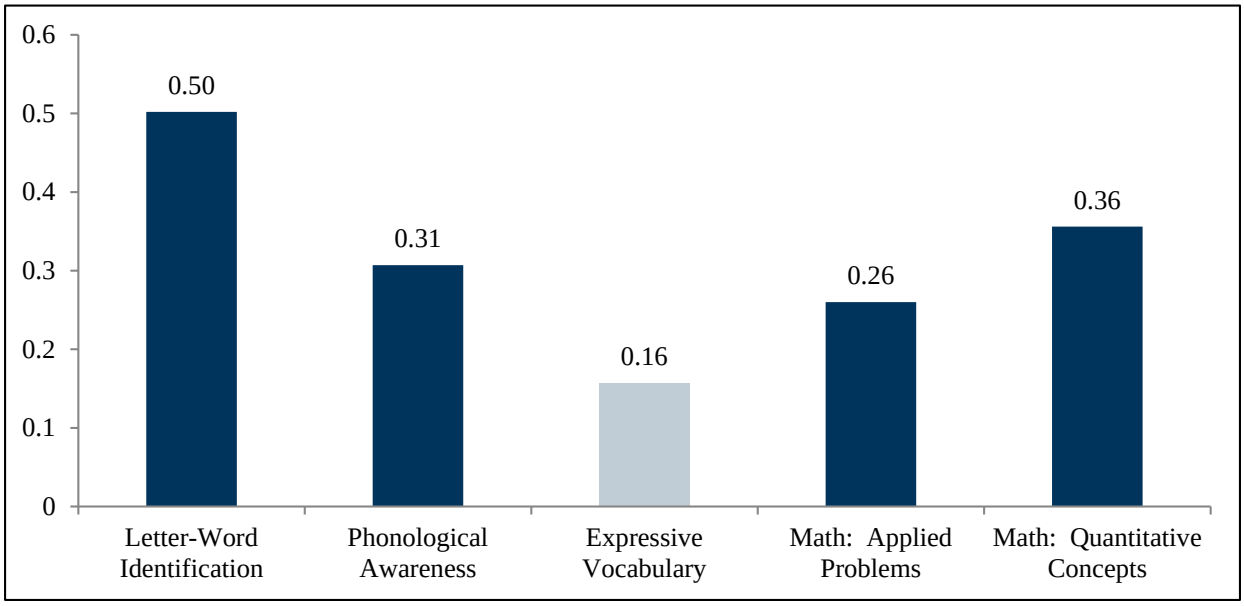
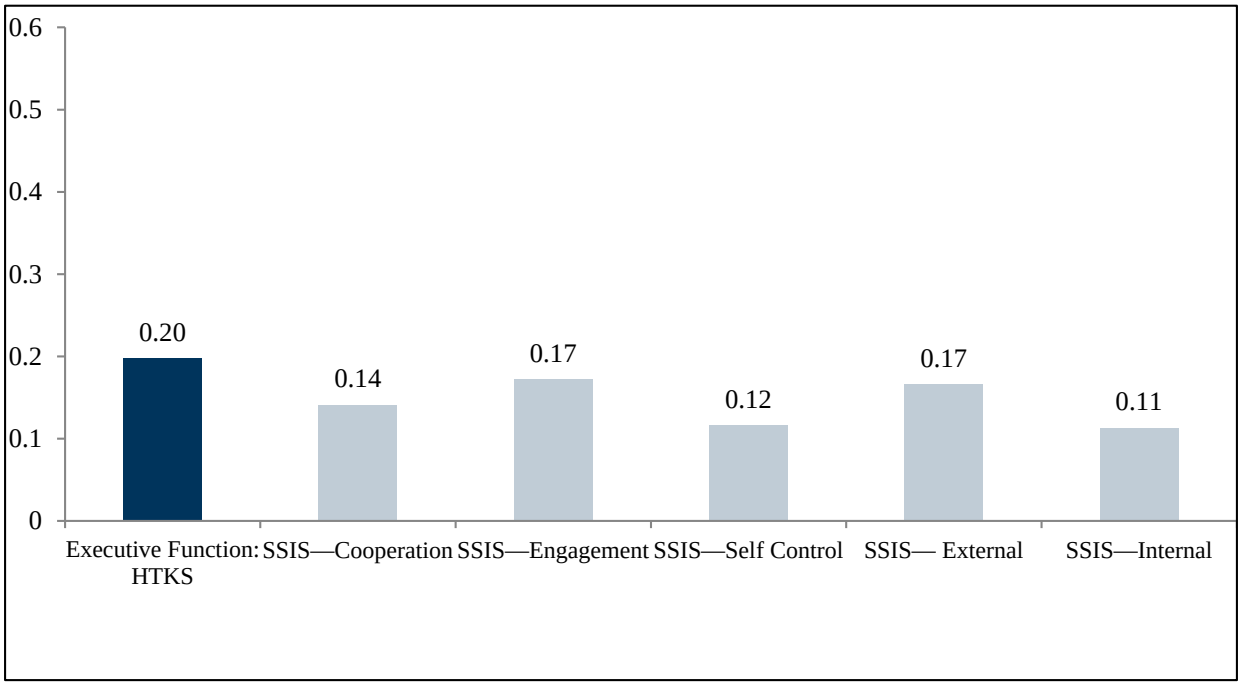


Exhibit 13. Effect Sizes for Executive Function and Social-Emotional Outcomes



Chapter 4: Conclusions and Policy Implications

This study found that students who attended TK were better prepared for kindergarten than students who did not attend TK. We found that TK broadly benefited enrolled students, improving their reading and math outcomes as well as their executive function. TK appears to have especially strong effects on preliteracy skills. By the time they entered kindergarten, students who had attended TK were five months ahead of non-TK attenders on their ability to identify letters and some sight words and three months ahead on their preliteracy phonological awareness skills. We also found effects on mathematics learning, with TK students performing three months ahead of their non-TK peers on mathematics problem solving, and an even larger impact on their knowledge of basic mathematical concepts and symbols. The relative effects of TK on various skill areas may be indicative of the amount of time TK teachers focused on these skills; future analyses using two years of data from kindergartners[8] will enable us to examine the relationship between specific content and practices in TK classrooms and student outcomes in different domains.

It is not surprising that students who attended TK—a full year of early education provided in a school setting by a qualified teacher with a kindergarten-aligned curriculum—are entering kindergarten with basic school readiness skills and are performing better than students who did not have similar early education experiences. It is important, nevertheless, to note that more than 80 percent of students in the comparison group attended some form of center-based preschool program while their TK counterparts were in TK. Thus, the benefits of TK we found were over and above the benefits of the other preschool programs experienced by the majority of children.

The observed impact was primarily on early academic measures. We did not find many effects of TK on social-emotional and behavioral outcomes. TK students demonstrated stronger executive function skills, which are important for later school achievement, but they did not have significantly better social-emotional or behavioral outcomes, such as engagement, cooperation, self-control, or less internalizing or externalizing behavior. This may be because the comparison students' typical preschool experiences equally impacted those outcomes. In the next phase of analysis, we will examine TK teachers' practices more closely to identify strategies that may be most supportive of social-emotional development.

The results of this study are timely for state policymakers considering the best ways to expand access to high-quality early educational opportunities for students in California. This study suggests that TK is an effective way to prepare students for kindergarten.

The findings also are relevant for school district leaders currently considering whether to include younger students in their TK program (i.e., those who turn 5 after December 2) in response to newly passed legislation (summer 2015) that allows districts to expand the eligibility window for TK and receive money for those younger students only after they turn 5. At least one large district in the state is moving toward enrolling more four-year-olds in TK and fewer in their public preschool program. Because the literature suggests that early education for four-year-olds can

[8] The study is currently collecting data from a second cohort of kindergartners who entered kindergarten in the fall of 2015.

affect kindergarten readiness if it is of high quality, we might expect that a TK program including younger children would have an impact similar to that found here, because important structural quality indicators—e.g., teacher qualifications and alignment with kindergarten—are built into the TK program design. We know, however, that other important process quality characteristics, such as teacher–child interactions, vary across TK classrooms. Assuming these process quality elements are in place, and that instruction is adequately adjusted to be developmentally appropriate for younger children, we might anticipate that TK would have a similar impact for a group of students that includes younger children. However, to be sure, it would be best to repeat this RD study with the new age cutoffs, since it is possible that the results presented here do not generalize to younger children.

It is also important to note that this study draws on data for students participating in the second year of the rollout of the TK program. Data for a second cohort of students, those who attended in the third year of TK (2014–15), are being collected now, and results using those data could differ; schools and districts are refining their approach to TK, and the program's impact could vary as implementation evolves.

Recent research (Lipsey, Farran, & Hoffer, 2015) raises questions about the long-lasting impact of prekindergarten programs, so the next steps in this study, examining the persistence of impact through the kindergarten year and characteristics of TK programs that best support children's kindergarten outcomes, will be of critical interest. When the data from both cohorts are combined, we will have a large sample that will allow us to more closely examine the relations between specific characteristics of TK classrooms, collected through classroom observations and TK teacher surveys, and student outcomes. In future reports using data from this ongoing study, we will examine the relation between TK classroom quality and structure and outcomes for participating students when they reach kindergarten. In addition, we will explore the impact of TK for particular subgroups of students, such as English learners.

References

Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). *The effects of Texas's targeted pre-kindergarten program on academic performance* (NBER Working Paper No. 18598). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18598

Aunio, P., Heiskari, P., Van Luit, J. E., & Vuorio, J.-M. (2015). The development of early numeracy skills in kindergarten in low-, average-, and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3–16.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children, 5*(3), 25–50.

Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*. Boulder: University of Colorado, Education in the Public Interest Center; Tempe: Arizona State University, Educational Policy Research Institute. Retrieved from http://greatlakescenter.org/docs/Policy_Briefs/Barnett_EarlyEd.pdf

Bassok, D., Latham, S., & Rorem A. (2015). *Is kindergarten the new first grade?* (Working Paper Series, no. 20). Charlotte: University of Virginia, Education Policy Works. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/20_Bassok_Is_Kindergarten_The_New_First_Grade.pdf

Bogard, K., & Takanishi, R. (2005). PK–3: An aligned and coordinated approach to education for children 3 to 8 years old. *Social Policy Report, 19*(3), 3–23. Retrieved from http://eric.ed.gov/?id=ED521747

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, 17*(1), 3–14.

California Department of Education. (2008). California Preschool Learning Foundations. Sacramento: Author. Retrieved from http://www.cde.ca.gov/sp/cd/re/documents/preschoollf.pdf

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, *6*(1), 42–57.

Cannon, J. S., & Karoly, L. A. (2007). *Who is ahead and who is behind? Gaps in school readiness and student achievement in the early grades for California's children.* Santa Monica, CA: RAND. Retrieved from http://eric.ed.gov/?id=ED498974

Cascio, E., & Schanzenbach, D. W. (2007). *First in the class? Age and the education production function* (NBER Working Paper No. 13663). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13663

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach*, 2nd ed. New York, NY: Routledge.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*(2), 127–166. Retrieved from http://www.jstor.org/stable/10.5951/jresematheduc.42.2.0127

Crosby, J. W. (2011). Test review: "Social Skills Improvement System Rating Scales," by F. M. Gresham & S. N. Elliott (Minneapolis, MN: NCS Pearson, 2008). *Journal of Psychoeducational Assessment*, *29*(3), 292–296.

Datar, A., & Sturm, R. (2004). Physical education in elementary school and body mass index: Evidence from the early childhood longitudinal study. *American Journal of Public Health*, *94*(9), 1501–1506.

De Avila, E. A., & Duncan, S. E. (2000). *PreLAS ©2000 English and Spanish: Technical notes*. Monterey, CA: CTB/McGraw-Hill.

Duncan, G. J. (2011). The importance of kindergarten-entry academic skills. In E. F. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The Pre-K debates: Current controversies and issues* (pp. 89–93). Baltimore, MD: Paul H. Brookes.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

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

Elkind, D., & Whitehurst, G. J. (2001). Young Einsteins. *Education Matters*, *1*(2), 8–21. Retrieved from http://eric.ed.gov/?id=EJ628699

Engel, M., Claessens, A., Watts, T. W., & Farkas, G. (2015, April). *The misalignment of kindergarten mathematics content.* Working paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

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

Gormley, W. T. Jr., Gayer, T., Phillips, D., & Dawson, B. (2005).The effects of universal Pre-K on cognitive development. *Developmental Psychology*, *41*(6), 872–884.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system (SSIS) rating scales.* Bloomington, MN: Pearson Assessments.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*(3), 205–224.

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, *79*, 933–959.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology*, *45*(3), 850–867. http://doi.org/10.1037/a0014939

Karoly, L. A., & Bigelow, J. H. (2005). *The economics of investing in universal preschool education in California.* Santa Monica, CA: RAND.

Karweit, N. L., & Wasik, B. A. (1992). *A review of the effects 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.-C., Kärnä, A., Niemi, P., Olofsson, Å., & Witting, K. (2014). Gains from training in phonological awareness in kindergarten predict reading comprehension in Grade 9. *Scientific Studies of Reading*, *18*(6), 452–467.

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A Randomized Control Trial of the Effects of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third grade* (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute.

Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2015). The prekindergarten age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis*, *37*(3), 296–313.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, *122*(1), 159–208.

Mantzicopoulos, P., & Morrison, D. (1992). Kindergarten retention: Academic and behavioral outcomes through the end of second grade. *American Educational Research Journal*, *29*(1), 182–198. Retrieved from http://aer.sagepub.com/content/29/1/182.short

McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*(2), 136–142.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers of Psychology*, *5*, 599 Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4060410/

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

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development, 86*(5), 1351–1370.

National Early Literacy Panel (2008). Developing early literacy. Washington, DC: National Institute for Literacy.

Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in Pre-K. *Reading Teacher, 62*(5), 384–392.

Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest, 10*(2), 49–88. Retrieved from http://psi.sagepub.com/content/10/2/49.short

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 605–619.

Quick, H., Manship, K., González, R., Holod, A., Cadigan, M., Anthony, J., et al. (2014). *Study of California's transitional kindergarten program: Report on the first year of implementation.* San Mateo, CA: American Institutes of Research. Retrieved from http://www.air.org/sites/default/files/downloads/report/Transitional%20Kindergarten%20Implementation%20Study%20Report.pdf

Raver, C. C., & Knitzer, J. (2002). *Ready to enter: What research tells policymakers about strategies to promote social and emotional school readiness among three- and four-year-old children.* (Policy Paper No. 3). New York, NY: National Center for Children in Poverty.

Reaney, L. M., & Kruger, T. (2002). *The ECLS-B 48-month assessment: Framework and measures.* U.S. Department of Education. National Center for Education Statistics. Washington, DC.

Reynolds, A. J., Temple, J. A., Ou, S. R., Robertson, D. L., Mersky, J. P., Topitzes, J. W., et al. (2007). Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. *Archives of Pediatrics and Adolescent Medicine, 161*(8), 730–739. Retrieved from http://archpedi.jamanetwork.com/article.aspx?articleid=570882&resultclick=1

Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development, 25*(5), 641–660. doi: 10.1080/10409289.2014.850397

Schochet, Peter Z. (2008). *Technical methods report: Statistical power for regression discontinuity designs in education evaluations* (NCEE 2008-4026). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early child development.* Washington, DC: National Academies Press.

Stipek, D. (2006). No child left behind comes to preschool. *Elementary School Journal, 106*(5), 455–466.

Walston, J., & Flanagan, K. (2013). *InformED blog—The new first grade: Kindergarten*. Retrieved from http://educationpolicy.air.org/blog/new-first-grade-kindergarten

Wasik, B. A. (2010). What teachers can do to promote preschoolers' vocabulary development: Strategies from an effective language and literacy professional development coaching model. *Reading Teacher, 63*(8), 621–633.

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

Whitebook, M., Gomby, D., Bellm, D., Saiki, L., & Kipnis, F. (2009). *Preparing teachers of young children: The current state of knowledge, and a blueprint for the future.* Berkeley: University of California, Berkeley, Center for the Study of Child Care Employment.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals preschool, second edition: Examiner's manual.* San Antonio, TX: Pearson/PsychCorp.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., et al. (2013). *Investing in our future: The evidence base on preschool education*. Ann Arbor, MI: Society for Research in Child Development. Retrieved from http://fcdus.org/sites/default/files/Evidence%20Base%20on%20Preschool%20Education%20FINAL.pdf

Zigler, E. F. (1987). Formal schooling for four-year-olds? No. *American Psychologist, 42*(3), 254–260. Retrieved from http://doi.org/10.1037/0003-066X.42.3.254

Zigler, E. F., & Bishop-Josef, S. J. (2006). The cognitive child versus the whole child: Lessons from 40 years of Head Start. In D. G. Singer, R. M. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (15–35). Oxford, UK: Oxford University Press.

Appendix A. Detailed Methodology

This appendix provides additional detail on the study's sample selection, measures, analytic approach, and results.

Power Analysis

Regression discontinuity designs are less statistically efficient than a randomized-assignment design (Schochet, 2008) because of the correlation between the treatment variable (eligibility for TK) and the forcing variable (age in days). Therefore, large sample sizes are needed to detect program effects. For the purpose of estimating the statistical power of the RD study design, we assume a symmetrical four-month window of birth dates around the cutoff date for TK eligibility and a minimum of a 70 percent response rate at the child level in the first wave of data collection. For a fuzzy RD design with the given assumptions,[9] in order to achieve a minimum detectable effect size of .19 at the spring of kindergarten, 2,352 children (14 children from each of 168 schools) were needed.

Sample Selection and Recruitment Procedures

We began by defining the population of school districts eligible for this study. We used several inclusion criteria in order to achieve the required sample size and maximize the statistical power of the research study. To be included, districts had to

1. Be a regular school district (i.e., one not run by a county office of education)
2. Be in operation in the 2013–14 school year
3. Have at least ten TK-age children during the 2012–13 school year
4. Follow state guidelines for enrollment criteria and include no more than 5 percent of their TK students who were born after the December 2 cutoff
5. Enroll at least 60 percent of their eligible students

We then assigned the resulting sampling frame of districts to sampling strata, defined by district urbanicity and the proportion of English learner students enrolled in the district and drew a sample of 94 districts, with the two largest districts included with certainty. Districts were randomly ordered within strata for recruitment. We continued to recruit study districts from this pool of 94, continuing from the districts that had the smallest random numbers to the largest random numbers, until we reached the number of districts needed to reach our target study sample size.

Exhibit A-1 shows how the sample of districts compares with that of California districts overall in terms of size and urbanicity. Fewer small districts and rural districts were included in our sample, because these types of districts enroll fewer students (thus contributing few of the students needed

[9] Assumptions: alpha = .05; a two-tailed test; power = .80; ICC = .15; treatment effect heterogeneity = 0; proportion of students who receive TK = 50 percent (symmetrical RD design); R^2 level 1 = .2; R^2 level 2 = .1; number of level 2 covariates = 0; design effect = 4; correlation between treatment and birthdate = .8. The design effect was selected in accordance with guidance in Schochet (2008).

to reach our large sample size) and data collection in these districts would have been more costly. However, the sample of districts included in the study does include broad representation from geographic regions across the state, and districts with various student demographics.

Exhibit A-1. Characteristics of TK Study Districts Compared With All Districts in California

	Sample (*n* = 20)	California (*n* = 942)
District Size		
Small	0.0%	33.6%
Medium	30.0%	33.1%
Large	70.0%	33.3%
Urbanicity		
Urban or suburban	60.0%	44.2%
Not urban or suburban	40.0%	55.8%

All schools that offered TK within participating districts were invited to be part of the study, except in one very large school district that had more schools than were needed for the study design. In that district, schools were stratified on the basis of their TK classroom configurations. Schools that offered TK in single-grade classrooms and schools that offered TK in combination with kindergarten or other grade levels were both invited to participate in the study, although schools with combination classrooms were oversampled to allow for later subgroup analyses by classroom type. To ensure a balanced student sample that would be representative of the state, schools also were selected on the basis of student demographic characteristics—namely, the proportion of English learners and students eligible for free or reduced-price lunch.

All students with birthdates between October 2 and February 2 in participating schools were invited to take part in the study. This birthdate range represents a window of 60 days before and after the December 2 TK eligibility cutoff. To recruit students, AIR sent consent forms to all students with birthdates in this range to schools. Teachers received a $10 gift card for every consent form from their classroom returned, and parents received a $10 gift card for returning the form, regardless of whether they elected to participate. We invited 5,897 students into our study; 3,924 returned consent forms, and 2,910 said yes (49% of those invited). The resulting sample for this first study cohort consists of 20 school districts, 164 schools, and 2,910 students, of whom 2,864 have outcome data.[10]

Exhibit A-2 presents descriptive statistics for the student sample by both eligibility and enrollment status. Students in our sample who were eligible for TK (the treatment group) did not differ on most background characteristics from students who were not eligible for TK (the comparison group), after controlling for age. However, TK students were significantly more likely to attend a center-based early care and education program two years before kindergarten. Exhibit A-2 also shows that student enrollment in the TK program at differing rates, in relation

[10] Some students were unable to be assessed because of chronic absence rates or because they had an individualized education plan (IEP) that prohibited testing. Some students did not have teacher rating data because their teachers did not respond to the survey.

with their demographic characteristics. For example, Hispanic children were more likely to enroll in TK than their peers.

Exhibit A-2. Demographic Characteristics of Student Sample by TK Eligibility and Enrollment Status

Student Demographics	Eligible for TK *n* = 1,562	Not Eligible *n* = 1,302	Attending TK *n* = 1,265	Not Attending *n* = 1,527
Mean age (as of 9/1/2014)	5.83***	5.66	5.83***	5.69
Female	51.0%*	48.8%	51.7%*	48.5%
Race				
White	26.0%	28.0%	25.5%	28.2%
Hispanic	55.4%	56.6%	57.4%	54.6%
Black	4.4%	3.9%	4.2%	4.1%
Asian	12.0%	9.6%	10.8%	11.0%
Other ethnicity	2.3%	1.9%	2.1%	2.2%
Free and reduced-price lunch eligibility	59.1%	58.6%	59.3%	58.0%
English learner	43.5%	39.9%	45.4% **	38.7%
Special education	7.0%	6.7%	6.5%	7.1%
Parental education				
Less than high school diploma	12.5%	13.6%	13.5%	12.8%
High school diploma	19.9%	20.5%	20.1%	20.1%
Some college	16.7%	17.6%	16.1%	18.0%*
Vocational certificate or AA	17.3%	18.4%	17.6%	18.2%
Graduated from college	18.9%	16.1%	18.4%	16.6%
Graduate education	14.7%	13.8%	14.3%	14.2%
Attended center-based preschool 2 years before kindergarten	76.6%***	49.9%	80.8%***	50.9%

**$p < .01$, ***$p < .001$

Measures

A direct, untimed, one-on-one cognitive assessment was administered to the kindergarten students[11] whose parents consented to the study. The direct assessment took approximately 45 minutes to an hour to complete and assessed each student's language, literacy, and mathematics knowledge and skills. The direct assessment also assessed each student's executive function. Social-emotional skills were assessed indirectly. Kindergarten teachers were asked to rate each student's engagement, cooperation, self-control, and internalizing and externalizing behavior using a four-point Likert scale. Detailed descriptions of these measures follow.

[11] First-grade children who had been in TK the year before also were invited to participate in the child assessments.

Direct Cognitive Assessments

Language and Literacy. We chose to assess students' language skills because California's large dual language learner population made it important to assess students' expressive and receptive English language skills in order to be sure that it was appropriate to assess these dual language learner students in English. All students who participated in the direct cognitive assessment were first administered two subtests of the *preLAS 2000* (De Avila & Duncan, 2000). The *preLAS 2000* is an English proficiency test designed for dual language learners who are between the ages of four and six. The *preLAS 2000* actually consists of five subtests, but for this study, only the receptive and expressive language subtests (Simon Says and Art Show) were administered.

All study students were first administered the subtest Simon Says, which assesses students' receptive English language skills or their ability to comprehend basic English commands (e.g., "Simon says, 'Point to the door'"). This subtest is similar to the game Simon Says and so was presented first to help the students become more comfortable with the testing situation. Next, the children were all given the *preLAS 2000* subtest Art Show in which they were shown pictures of objects and asked to name them and say what they are used for. Art Show thus assesses students' ability to express themselves in English.

These two subtests were given to all study students as a warm-up, intended to acclimate them to the testing situation. However, for dual language learners, they also served to determine whether the assessor could continue the assessment in English, or whether she needed to switch to Spanish or terminate the session because of the student's limited English understanding. All assessments were available in both English and Spanish. If the child did not achieve a combined score on Simon Says and Art Show of at least 12 correct out of 20[12] and the child spoke Spanish (according to a parent's response on the consent form), the assessment was continued in Spanish. If the student's combined score was 12 or higher, however, the assessment continued in English. For those students who spoke neither English nor Spanish but spoke Mandarin, Cantonese, Korean, Tagalog, or Vietnamese and failed the *preLAS*, the assessor was able to administer the executive function assessment in the student's language before terminating the session. Students whose language was something other than those just listed who failed the *preLAS* could not be assessed.

Expressive vocabulary was next assessed using the *Clinical Evaluation of Language Fundamentals—Preschool 2* (CELF-2P) assessment. Vocabulary knowledge is essential for reading comprehension; if a child does not know the meaning of words he is reading, then he cannot comprehend the text. The CELF-2P is a criterion-referenced diagnostic measure consisting of nine subtests designed to identify children ages 3–6 in need of speech or communication therapy. It has been validated in Spanish, and each subtest takes approximately three to five minutes to administer. The CELF-2P Expressive Vocabulary assessment asks children to name pictures and describe actions depicted in pictures. The assessor scores the student's responses, awarding either full or partial credit on the basis of whether the student said

[12] The cut score of 12 out 20 was used in the Early Childhood Longitudinal Study Kindergarten Class of 2011 (ECLS-K:2011) to identify children who needed to be assessed in Spanish rather than English.

exact target words or something similar. The English Expressive Vocabulary assessment is discontinued once the student misses seven consecutive items; the Spanish assessment (Vocabulario Expresivo) is terminated once the student misses five consecutive items.

After the CELF-2P Expressive Vocabulary subtest was administered, the student was administered the CELF-2P Phonological Awareness subtest. In this activity, students were asked to complete various phonological awareness tasks such as putting together two words to make a new word (e.g., "bed" and "room" make "bedroom"), or to clap the words in a sentence. For both the Spanish and the English subtests, the subtest is discontinued when the student gets all the items in three consecutive sections incorrect; most students receive the entire subtest.

Students' ability to name letters and read common words was assessed using the Woodcock-Johnson III Letter and Word Identification Subtest. Those students that were assessed in Spanish took the equivalent subtest in the Batería III by Woodcock-Munoz. The Letter Word subtest takes approximately five minutes to administer. Students are asked to point to letters named, name letters, and read sight words. The subtest is discontinued once the student misses six consecutive items.

Executive Function. Executive function is a set of cognitive skills that work in tandem to help an individual formulate and execute a plan. These skills are developing in young children, and research with this age group tends to focus on the following three skills: working memory, inhibition, and cognitive flexibility. The Head Toes Knees Shoulders (HTKS) activity (Ponitz et al., 2009) was included in the direct assessment because it assesses all three of these critical skills. The task consists of three parts. In the first part, the student is instructed to touch his toes when told to touch his head and to touch his head when told to touch his toes. Thus, the student must remember the rule and, at the same time, inhibit the impulse to touch the body part named. In the second part, the task is made more challenging by adding knees and shoulders. Now the student must touch his shoulders when told to touch his knees and touch his knees when told to touch his shoulders. There are four rules to remember, and the student must continue to inhibit the impulse to follow the commands literally. The third and most challenging part of the task is when the original rules change, taking cognitive flexibility to a higher level. In this part, the student needs not only to be flexible in the rules, but also must forget the rules of the previous two parts of the task and learn new rules (for example, "touch your head" now means to touch your knees). The student advances to the next part of the task when he has earned at least four points on the 10 test items. Each item is scored full credit (two points) if the student goes immediately to the correct body part or partial credit (one point) if the student starts to go to the wrong body part and then self-corrects, inhibiting the initial impulse but ultimately giving the correct response.

Mathematics. With the rising emphasis in education on STEM subjects—science, technology, engineering, and mathematics—schools are beginning to introduce and build mathematics skills earlier in schooling. Mathematics once may have been defined narrowly as number sense and number operations, but now it also includes understanding of shapes, patterns, relative comparisons, and other skills. Consequently, the direct assessments chosen for this study included measurement of multiple mathematics skills. Two subtests of the Woodcock-Johnson III and their equivalent in the Batería III (Woodcock-Munoz) were administered after the HTKS task. The Applied Problems subtest assessed students' quantitative reasoning and mathematical

knowledge, asking them to count, do basic operations, and figure out what information in a word problem is needed to solve the problem. The subtest in both English and Spanish is discontinued when the student misses six consecutive items. The final mathematics assessment administered in the assessment battery was the Woodcock-Johnson III Quantitative Concepts, which assesses students' understanding of the number line, recognition of mathematical symbols, and understanding of different types of representations. The Quantitative Concepts subtest is discontinued when the student misses four consecutive items.

Supplement for Spanish Speakers. For those students who speak both Spanish and English, in order to capture their knowledge in both languages, we administered two subtests—the CELP-2P Expressive Vocabulary and the Woodcock-Muñoz Applied Problems—in the alternate language at the end of the assessment. In other words, if the student was assessed in English for the main assessment battery, the student was administered the Expressive Vocabulary and Applied Problems subtests in Spanish at the end of the assessment session. If students were assessed in Spanish for the primary battery, then they were administered these two assessments in English at the end. Thus, all Spanish-speaking students were assessed on their vocabulary and a set of mathematical skills in both languages. Ideally, all Spanish speakers would be administered the full assessment battery in both languages, but time constraints allowed only a subset of assessments—one language/literacy and one mathematics—to be repeated.

In the fall, only 45 students were given the primary assessment in Spanish; all other Spanish speakers were able to score at least 12 out of 20 on the two *preLAS* subtests. Thus, the majority of the Spanish-speaking students in the study were assessed in English and then given the supplement in Spanish.

Asian language students. California has a high percentage of students who speak an Asian language. Concerns about losing these students in the sample led to translating the HTKS executive function task into Mandarin, Cantonese, Korean, Tagalog, and Vietnamese—the five most commonly spoken Asian languages in study districts. It was not possible to simply translate the other assessments because they involved pictorial representations and language concepts not easily translated into these languages. In contrast, the HTKS task was easily translatable because it comprises only verbal commands. We hired assessors fluent in these languages to administer the HTKS in these languages in the event that the student failed to score at least 12 out of 20 on the *preLAS* and could not be assessed in English. Only two students who spoke an Asian language at home, however, failed the *preLAS* in the fall and were assessed in one of these Asian languages.

Indirect Social-Emotional Assessment

One goal of TK is to prepare students socially and emotionally for kindergarten, and so we decided to measure social and emotional skills among study students. However, because social-emotional skills are displayed in interaction with peers, they are hard to directly assess. Teachers are commonly asked to report on students' social-emotional skills, so we asked students'

kindergarten[13] teachers to complete selected subscales from the Social Skills Improvement System Rating Scales (Gresham & Elliott, 2008). In order to minimize the burden on the teacher, only five subscales were included in the teacher survey. Three of these subscales tapped positive or prosocial behaviors: cooperation, engagement, and self-control. Teachers also were asked to report on two problem areas: internalizing behavior and externalizing behavior. Teachers were asked to rate students using a four-point Likert scale. These five subscales were chosen on the advice of our technical advisory group members.

Exhibit A-3 presents each student outcome measure used in the study, skills it assesses, its scale, how it was administered, and its reliability coefficient. All outcome measures were standardized, i.e., converted to *z*-scores, prior to analysis.

Exhibit A-3. Measures of Student Outcomes

Measure	Skills Assessed	Scale	Source	Reliability
Language and Literacy Skills				
Clinical Evaluation of Language Fundamentals Preschool-2 Expressive Vocabulary subtest	Expressive vocabulary	Sum of items correct Range: 0–40	Direct student assessment	.94 (Wiig, Secord, & Semel, 2004).
Clinical Evaluation of Language Fundamentals Preschool-2 Phonological Awareness subtest	Phonological awareness	Sum of items correct Range: 0–24	Direct student assessment Direct student assessment	.86 (Wiig, Secord, & Semel, 2004).
Woodcock-Johnson Letter-Word Identification subtest	Ability to name letters and read words	Sum of items correct Range: 0–76	Direct student assessment	.94 (Schrank, McGrew, & Woodcock, 2001).
Mathematics Skills				
Woodcock-Johnson Quantitative Concepts subtest	Mathematical concepts, symbols, and vocabulary	Sum of items correct Range: 0–34	Direct student assessment	.91 (Schrank, McGrew, & Woodcock, 2001).
Woodcock-Johnson Applied Problems subtest	Mathematics numeracy and basic operations	Sum of items correct Range: 0–63	Direct student assessment	.97 (Schrank, McGrew, & Woodcock, 2001).
Executive Function				
HTKS assessment	Executive function (inhibitory control, attention, and working memory)	Sum of items correct Range: 0–60	Direct student assessment	.93 (McClelland, & Cameron, 2012).

[13] In some cases, when TK students were promoted directly to first grade, first-grade teachers completed these surveys.

Measure	Skills Assessed	Scale	Source	Reliability
Social-Emotional Skills				
SSIS rating scales, Cooperation subscale	Helping others, sharing materials, and complying with rules and directions	Mean rating across items Range: 1–4	Teacher report	0.81 (Crosby, 2011)
SSIS rating scales, Engagement subscale	Joining activities in progress and inviting others to join, initiating conversations, making friends, and interacting well with others	Mean rating across items Range: 1–4	Teacher report	
SSIS rating scales, Self-Control subscale	Responding appropriately to conflict (e.g., disagreeing and teasing) and nonconflict situations (taking turns and compromising)	Mean rating across items Range: 1–4	Teacher report	
SSIS rating scales, Externalizing subscale	Being verbally and physically aggressive, failing to control temper, and arguing	Mean rating across items (reverse coded) Range: 1–4	Teacher report	
SSIS rating scales, Internalizing subscale	Feeling anxious, sad, and lonely; exhibiting poor self-esteem	Mean rating across items (reverse coded) Range: 1–4	Teacher report	

Covariates

Covariate data came from district administrative data and a parent survey that was distributed with student consent forms. (See Exhibit A-4 for the full list of covariates.) Districts provided the administrative data for 97 percent of study students, and 90 percent of parents responded to the parent survey. In instances where covariate data were missing, the research team imputed a value using other sources of information. For example, household income and household size as reported on the parent survey were used to impute free or reduced-price lunch eligibility, if that information was missing in the administrative data. Gender was imputed from parent report or students' names, if possible. Students' scores on the *preLAS* measure were used to determine English learner status, if missing. In cases where values could not be reasonably inferred from available data, missing values on covariates were recoded to zero, and an indicator variable was generated to note that the variable had been recoded. These missing indicators for each covariate were included in the RD models.

Exhibit A-4. Student-Level Covariates

Measure	Coding	Source
Date of birth (student age)	Month, day, and year	District administrative data
Race/ethnicity	Binary indicators for White, African-American, Hispanic, Asian-American, or Other Ethnicity	District administrative data
English learner status	Yes or no indicator	District administrative data
Gender	Yes or no indicator for female	District administrative data
Free or reduced-price lunch eligibility	Yes or no indicator	District administrative data
Household income	Binary indicators for $0–25,000, $25–50,000, $50–75,000, $75–100,000, $100–125,000, $125,000 and above	Parent survey
Highest level of schooling completed by adult in household	Binary indicators for less than high school diploma, high school diploma, some college, vocational certificate or AA, graduated from college, graduate education	Parent survey
Special needs status	Yes or no indicator	District administrative data
Early education program participation in 2012–13	Yes or no indicator. A student is marked as "yes" if he or she participated in a center-based early education program. We considered the following programs to be center-based: • Child-care center • Head Start program • Prekindergarten program • Transitional kindergarten program • Preschool or nursery school program	Parent survey

Regression Discontinuity Design

The regression discontinuity (RD) design compares two groups of students on either side of the December 2 eligibility cutoff. For example, a child born December 2 and a child born December 3 are very close in age, but one will attend TK the year before kindergarten, while the other will just miss being eligible for TK and enter kindergarten at the same time as his peer but without having had TK. This rigorous approach reduces the risk that selection bias will affect the impact estimates. Results from this model produce an impact estimate generalizable to children born at the cutoff date. However, if we included only children with December 2–3 birthdays, we would have a sample size too small to draw any conclusions. The RD approach can be applied to a group of children with a wider band of birthdates around the cutoff if models control for the effects of age, which we have done in this study.

Unlike other RD studies of early education programs (e.g. Gormley, Gayer, Phillips, & Dawson, 2005; Weiland & Yoshikawa, 2013), all students included in the study entered kindergarten in the same year, because of the design of the TK program. Other studies rely on two groups of

students entering prekindergarten in two different years—one group who made the cutoff, and one group who missed it. Because these two cohorts of students entered school at different times, their experiences might have been different as a result of different events (in the school or community, thought of as historical effects) happening that year, which is a threat to the validity of these studies' results (Lipsey, Weiland, Yoshikawa, Wilson, & Hofer, 2015). By contrast, in the current study, historical effects are not a concern.

Analytic Approach

The primary results presented in this report are from fuzzy RD models with a 60-day bandwidth on either side of the eligibility cutoff. These models use a linear functional form for age, as opposed to a quadratic or cubic functional form, and include demographic covariates. Because we have a hierarchical data structure in which students are clustered within schools, we take this dependency in the data into account by using cluster-adjusted standard errors in all of our analyses. The results of these models are presented in table form in Exhibit A-7.

The study team also conducted a series of sensitivity analyses that tested alternative model specifications, including

- Fuzzy and sharp RD models
- Varying bandwidths around the eligibility cutoff
- Different functional forms for student age
- Models with and without covariates

The results of the alternative models are presented in Exhibits A-7 through A-9.

Sharp Versus Fuzzy RD Estimates

Sharp RD models ignore any noncompliance with treatment assignment. The purpose of these models is to compare students who are eligible with those who are not eligible for TK to estimate the effect of *offering* the program, that is, the so-called intent-to-treat effect. Ignoring noncompliance attenuates the estimated impact of TK because some of the control students might have attended TK and some treatment students might have chosen not to attend TK. Therefore, the results from these analyses, which are also called intent-to-treat estimates, provide a conservative estimate of the effect of TK participation on student outcomes.

Let x_i and x_0 denote the student *i*'s birth date and the December 2 enrollment cutoff date for TK eligibility, respectively. Defining treatment, D_i, as TK participation,

$$D_i = \begin{cases} 1 \text{ if } x_i < x_0 \\ 0 \text{ if } x_i \geq x_0 \end{cases} \quad (1)$$

a common regression model representation of this evaluation problem would become

$$Y_i^k = \alpha + \beta D_i + \varepsilon_i \quad (2)$$

where, in the main specification, Y_i^k is the test score of student *i* in assessment *k* in the fall of kindergarten, where *k* is CELF Expressive Vocabulary, CELF Phonological Awareness, Woodcock-Johnson Letter-Word Identification subtest, Woodcock-Johnson Quantitative Concepts subtest, Woodcock-Johnson Applied Problems subtest, HTKS assessment, or SSIS rating scales.

Provided that the conditional mean function $E[\varepsilon_j|x_j]$ is continuous at the TK eligibility cutoff, the causal impact of TK participation on a student outcome is given by

$$\beta = \lim_{x_j \uparrow x_0} E[Y_i^k|x_j] - \lim_{x_j \downarrow x_0} E[Y_i^k|x_j] \quad (3)$$

Parametrically, we estimate Equation 3 with the following equation using ordinary least-squares:

$$Y_i^k = \alpha + \beta_1 D_i + f(x_i) + \varepsilon_i \quad (4)$$

where $f(x_j)$ is a polynomial function of the selection variable. Because we have a hierarchical data structure in which students are clustered within schools, we take this dependency in the data into account by using cluster-adjusted standard errors in all of our analyses. In other words, in our models we separate the residual into student-level and classroom-level residuals. The final model run is in the following form:

$$Y_{is}^k = \alpha + \beta_1 \text{Eligible}_{is} + \beta_2 \text{Age}_{is} + \beta_3 \text{Age}_{is}^2 + \beta_4 \text{Age}_{is}^3 + \beta_5 \text{Age}_{is}^4 + \beta_5 \text{Covariates}_{is} + \vartheta_s + \varepsilon_{is} \quad (5)$$

where Eligible_{is} is the TK eligibility status for student *i* in school *s*, Age_{is} refers to the student's birthdate centered at eligibility cutoff, Covariates_{is} denotes student-level covariates, ϑ_s is the school residual, and ε_{is} is the student residual.

Noncompliance with enrollment guidelines leads to fuzziness at the December 2 enrollment cutoff, where the effect of the TK is to be estimated. Fuzzy RD models account for the fact that some children do not comply with their treatment assignment; this enables a better estimate of the effect of TK for children who actually attend. Some districts enroll students in TK who are younger than the state eligibility guidelines. Though we excluded districts from our sampling frame that did so frequently, some sample districts still allowed this for some students (in our sample, 1.6% of ineligible students). In addition, some parents chose to keep their TK-eligible child at home or in a preschool program for an additional year prior to school enrollment rather than attend TK (in our sample 17.8% of eligible students).

The model representation for fuzzy RD is similar to sharp RD as shown in Equation 2. However, in fuzzy RD, instead of a deterministic jump at the cutoff score (as in sharp RD), we estimate the probability of jump by

$$P[D_i|x_i] = \begin{cases} g_0(x_i) \; if \; x_i \geq x_0 \\ g_1(x_i) \; if \; x_i < x_0 \end{cases} \quad \text{where } g_0(x_i) \neq g_1(x_i) \quad (6)$$

which can be rewritten as

$$E[D_i|x_i] = P[D_i = 1|x_i] = g_0(x_i) + [g_1(x_i) - g_0(x_i)]T_i \quad (7)$$

where $T_i = 1(x_i > x_0)$ (i.e., eligibility status) and is an instrument for the TK participation cutoff. We use two-stage least squares to estimate the impact of TK, where the first stage is

$$D_i = \theta_0 + f(x) + \pi T_i + \vartheta_{1i} \quad (8)$$

Inserting this into the equation for the RD model (Equation 3), we find the reduced form of fuzzy RD as

$$Y_i = \mu + \beta_1 \pi T_i + (\beta_1 + 1) f(x) + \epsilon_i \quad (9)$$

where $\mu = \alpha + \beta_1 \theta_0$ and $\epsilon_i = \vartheta_{1i} + \varepsilon_i$. Because we have a hierarchical data structure in which students are clustered within schools, we take this dependency in the data into account by using cluster-adjusted standard errors in all of our analyses. The fuzzy RD design employs a two-stage least-squares correction (2SLS) to account for both of these forms of noncompliance with the cutoff date. The model estimates the effect of the treatment on those who received it by using predicted participation, rather than eligibility, as the primary explanatory variable in the impact model. In the first-stage model for the fuzzy RD, the probability of participation is estimated using student age:

$$\text{Participation}_\text{i} = \theta_0 + \theta_1 \text{Age}_\text{i} + \varepsilon_\text{i}$$

In the second stage, estimated participation from the first stage model is used as a predictor variable:

$$\text{Outcomes}_\text{is}^\text{k} = \alpha + \beta_1 \widehat{\text{Participation}}_\text{is} + \beta_2 \text{Age}_\text{is} + \vartheta_s + \varepsilon_{is}$$

Optimal Bandwidth

Bandwidth refers to the age range of students on either side of the eligibility cutoff who are included in the analytic sample. There are several tests that may be used to determine optimal bandwidth (Imbens & Kalyanaraman, 2012; Ludwig & Miller, 2007). However, these tests rely on comparing averages within arbitrarily small neighborhoods around the cutoff, which is not feasible with a discrete forcing variable. For this study, age measured in days is the forcing variable that defines TK program eligibility. We chose 60 days on either side of the eligibility cutoff as our optimal bandwidth, which represents students born up to two months before the cutoff and students born up to two month after the cutoff. A formal statistical test for optimal bandwidth, called cross-validation, supports this choice of bandwidth.[14] This bandwidth also is ideal because it uses all available data and maximizes our statistical power. However, we also tested models using 15-, 30-, and 45-day bandwidths to test whether our results were sensitive to the bandwidth selection, which they were not.

[14] Statistical tests for optimal bandwidth require a continuous variable for program eligibility, whereas age is a discrete variable. However, we still computed the optimal bandwidth using both the IK (Imbens & Kalyanaraman, 2012) method and the cross-validation method (CV) proposed by Ludwig and Miller (2007). The optimal bandwidths for IK range from 22.3 to 57.2, which varies with the outcome, whereas the bandwidths from CV is 59 for all outcomes.

Functional Form for Age

For all outcomes, we present the linear model that includes only the linear age term, but we tested the sensitivity of results to the use of functional forms using quadratic and cubic terms. We determined that the linear model was the best fit for the data because the higher order polynomial terms were not consistently significant.

Inclusion of Covariates

We present models with covariates in order to fully account for student background characteristics. This approach is more conservative and follows the norms in the early childhood research literature. We also ran models without covariates to determine how results differ. Note that the addition of covariates did not increase the predictive power of our models, defined as the total variance explained, in the first stage of the two-stage least-squares (2SLS) models used for impact estimation. (See Exhibit A-6.)

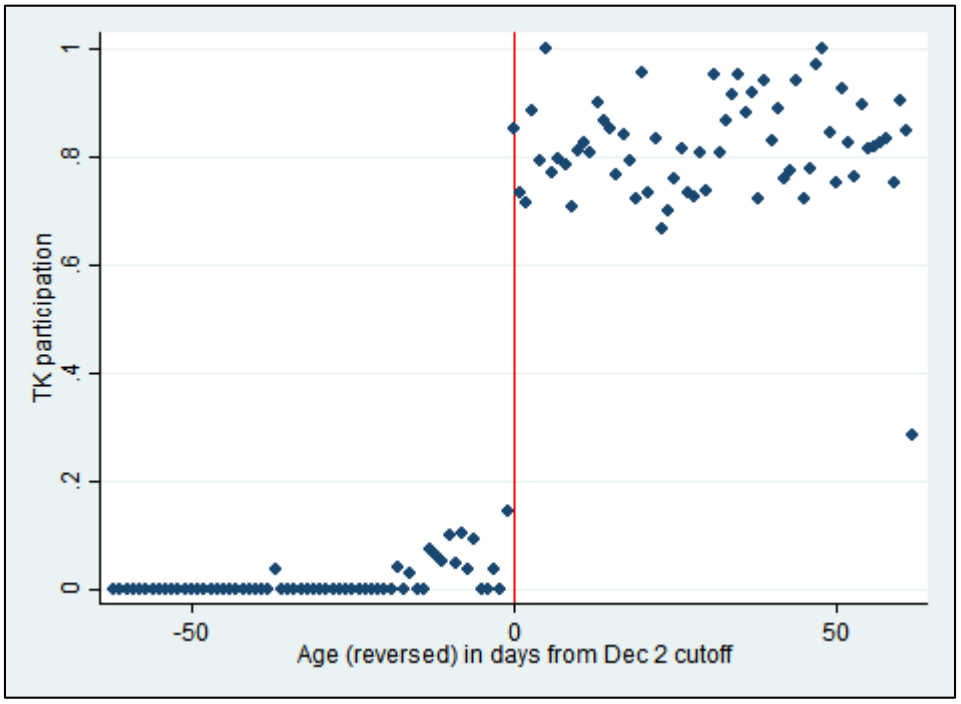
Diagnostic Checks for RD Analyses

The RD analyses require that there is a discontinuity (i.e., jump) in the program participation around the cutoff. Exhibit A-5 shows that compliance with treatment assignment was high in the study sample; 82.2 percent of students who were eligible for TK participated in the program, and 98.4 percent of students who were *not* eligible did *not* participate. In other words, there is a big jump in program participation at the cutoff, but there is also some "fuzziness" in program participation, mainly due to eligible students not attending TK (Exhibit A-6).

Exhibit A-5. Compliance by Treatment Assignment

Group	Attended TK	Did not attend TK
Eligible (treatment)	82.2%	17.8%
Ineligible (comparison)	1.6%	98.4%

Exhibit A-6. TK Participation Rates by Age



The impact estimate in RD designs depends on the assumption that, in the absence of any intervention, there would be a smooth relationship (i.e., no discontinuity) between the outcome and the forcing variable. For this reason, any discontinuity observed in the outcome at the cutoff is attributed to the intervention. Therefore, to check the smoothness assumption, we checked for the discontinuity at the cutoff in the forcing variable and student and family background characteristics, such as poverty status, English learner status, race/ethnicity, family income, and parental education, among others. The visual inspection of the figures did not reveal any jump around the cutoff. In addition to inspecting the continuity in the forcing variable visually, we also tested whether the density of the forcing variable (i.e., age) is continuous at the cutoff visually as well as by using the McCrary (20078) test. The results of the McCrary test confirmed that there was no discontinuity in the forcing variable at the cutoff. Finally, we examined the functional form of the relationship between the forcing variable and the outcomes because, in the parametric approach, the validity of estimates from RD depends on whether the polynomial function is an accurate representation of $E[Y_i|\ x_i]$. Otherwise, an apparent jump at the cutoff that might be due to misspecification of the mean function could be mixed with the treatment effect. The results from these functional form analyses are discussed in the sensitivity analyses section that follows.

Results

Results from the primary RD model, including effect sizes, standard errors, sample sizes, first-stage r-squared values, and first-stage F values, are presented in Exhibit A-7. As described above in the Measures section, all outcome variable are standardized. Thus, the regression coefficients are effect sizes that report the standardized mean difference between the treatment and comparison groups. The effect size can be represented by the following formula:

$$d = \frac{\bar{M}_t - \bar{M}_c}{SD}$$

where $\bar{M}_t$ represents the treatment group mean $\bar{M}_c$ represents the comparison group mean, and *SD* represents the pooled standard deviation. The use of effect sizes allows the reader to compare across outcomes, even if they were originally on different scales, to see which outcomes demonstrate a larger impact of TK.

Exhibit A-7. The Impact of Transitional Kindergarten on Student Outcomes

Outcome	Effect Size	*SE*	*N*	First-Stage *R*-Squared	First-Stage *F*
Language and Literacy Outcomes					
W-J Letter-Word ID	0.502***	0.100	2636	0.677	962.750
CELF Phon Aware (raw)	0.307**	0.094	2647	0.677	937.816
CELF Exp Vocab (raw)	0.157†	0.085	2695	0.673	941.221
Mathematics Outcomes					
W-J Applied Problems	0.260**	0.090	2675	0.675	987.858
W-J Quant Concepts	0.356***	0.084	2629	0.680	965.089
Executive Function					
HTKS	0.197*	0.090	2683	0.674	972.801
Social-Emotional Skills					
SSIS—Cooperation	0.141	0.103	2223	0.659	670.974
SSIS—Engagement	0.172	0.117	2203	0.658	655.840
SSIS—Self-Control	0.116	0.107	2189	0.660	641.504
SSIS— External	0.166	0.117	2217	0.658	668.328
SSIS—Internal	0.113	0.114	2214	0.658	667.473

†$p < .1$, *$p <. 05$, **$p < .01$, ***$p < .001$

Note: The estimates are from fuzzy RD models with a bandwidth of 60 days around the cutoff and a linear functional form for age. The covariates included in the model are dummy variables for race/ethnicity, special education, free or reduced-price lunch, English learner, parental education, income, early childhood education participation two years before kindergarten, and missing indicators for any missing covariates.

Results of Sensitivity Analyses

Sensitivity analyses included fuzzy and sharp RD models, models with and without covariates, varying bandwidths around the eligibility cutoff, and different functional forms for student age. The relative magnitude of the effects for the different outcomes is very similar in the sharp and fuzzy RD models and are robust to the inclusion of covariates (see Exhibit A-8 for details). As seen in Exhibit A-9, the sample size decreases as the bandwidth decreases. The estimates are less precise and less likely to be statistically significant in the models with smaller samples. The estimates are similar across models with differing functional forms for age, as shown in Exhibit A-10.

Exhibit A-8. The Impact (Effect Size) of Transitional Kindergarten on Student Outcomes by Model Type

Outcome	Sharp	Sharp with covariates	Fuzzy	Fuzzy with covariates
Language and Literacy Outcomes				
W-J Letter-Word ID	0.442***	0.383***	0.567***	0.502***
CELF Phon Aware (raw)	0.245**	0.234***	0.309**	0.307**
CELF Exp Vocab (raw)	0.113	0.123+	0.134	0.157+
Mathematics Outcomes				
W-J Applied Problems	0.235**	0.197**	0.299**	0.260**
W-J Quant Concepts	0.313***	0.270***	0.403***	0.356***
Social-Emotional Skills				
HTKS	0.166*	0.153*	0.203*	0.197*
SSIS—Cooperation	0.164*	0.111	0.205*	0.141
SSIS—Engagement	0.139	0.116	0.195+	0.172
SSIS—Self-Control	0.140+	0.095	0.170+	0.116
SSIS—External	0.159+	0.115	0.214+	0.166
SSIS—Internal	0.103	0.088	0.128	0.113

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: The estimates are from models with a bandwidth of 60 days around the cutoff and linear functional form for age.

The covariates included in the models with covariates are dummy variables for race/ethnicity, special education, free or reduced-price lunch, English learner, parental education, income, early childhood education participation two years before kindergarten, and missing indicators for any missing covariates.

Exhibit A-9. The Impact (Effect Size) of Transitional Kindergarten on Student Outcomes by Bandwidth

Outcome	Bandwidth = 15 days (*N* = 576-708)	Bandwidth = 30 days (*N* = 1115–1406)	Bandwidth = 45 days (*N* = 1650–2046)	Bandwidth = 60 days (*N* = 2189–2695)
Language and Literacy Outcomes				
W-J Letter-Word ID	0.350+	0.384**	0.462***	0.502***
CELF Phon Aware (raw)	0.287	0.240*	0.262*	0.307**
CELF Exp Vocab (raw)	0.064	0.059	0.134	0.157+
Mathematics Outcomes				
W-J Applied Problems	0.289	0.294*	0.225*	0.260**
W-J Quant Concepts	0.273	0.355**	0.315***	0.356***
Social-Emotional Skills				
HTKS	0.189	0.215+	0.117	0.197*
SSIS—Cooperation	0.283	0.030	0.120	0.141
SSIS—Engagement	0.477+	0.135	0.174	0.172
SSIS—Self-Control	0.165	0.019	0.097	0.116
SSIS—External	0.008	0.001	0.109	0.166
SSIS—Internal	0.062	0.132	0.060	0.113

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$
Note: The estimates are from fuzzy RD models with a linear functional form for age.
The covariates included in the model are dummy variables for race/ethnicity, special education, free or reduced-price lunch, English learner, parental education, income, early childhood education participation two years before kindergarten, and missing indicators for any missing covariates.

Exhibit A-10. The Impact (Effect Size) of Transitional Kindergarten on Student Outcomes by Functional Form for Age

Outcome	Linear	Quadratic	Cubic	Sample Size
Language and Literacy Outcomes				
W-J Letter Word ID	0.502***	0.495***	0.421***	2636
CELF Phon Aware (raw)	0.307**	0.304**	0.203+	2647
CELF Exp Vocab (raw)	0.157+	0.148+	0.096	2695
Mathematics Outcomes				
W-J Applied Problems	0.260**	0.254**	0.228+	2675
W-J Quant Concepts	0.356***	0.354***	0.294**	2629
Social-Emotional Skills				
HTKS	0.197*	0.184*	0.080	2683
SSIS—Cooperation	0.141	0.139	0.099	2223
SSIS—Engagement	0.172	0.164	0.166	2203
SSIS—Self-Control	0.116	0.108	0.055	2189
SSIS—External	0.166	0.164	0.034	2217
SSIS—Internal	0.113	0.109	0.101	2214

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$
Note: The estimates are from fuzzy RD models with a bandwidth of 60-days around the cutoff. The covariates included in the model are dummy variables for race/ethnicity, special education, free or reduced-price lunch, English learner, parental education, income, early childhood education participation two years before kindergarten, and missing indicators for any missing covariates.

Additional Exploration

Spanish-speaking students were administered both the Expressive Vocabulary and Applied Problems subtests in English and Spanish. Additional exploration of Expressive Vocabulary scores in English and Spanish for bilingual students will be presented in future reports including both cohorts of students. For this report, in order to test whether students' home language had any effect on the measurement of their mathematical skills, we created a new variable that uses a student's highest score on either the Woodcock-Johnson Applied Problems (English) or the Woodcock-Muñoz Problemas Aplicados (Spanish) assessment. The impact estimate from this model was an effect size of .242 (p<.01), which is very similar to the impact estimates from Woodcock-Johnson Applied Problems English test presented in this report (an effect size of .254 (p<.01).

We also explored the possibility that differences between groups may lie in very high and very low ratings on the SSIS. To explore this, we examined differences in proportion of treatment and comparison students rated very highly or very low by their teachers on each subscale. There were no consistent patterns in this exploration that suggested a differential impact of TK.

ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

2800 Campus Drive, Suite 200
San Mateo, CA 94403-2555
650.843.8100 | TTY 650.493.2209

www.air.org

Making Research Relevant

LOCATIONS

Domestic

Washington, D.C.
Atlanta, GA
Austin, TX
Baltimore, MD
Cayce, SC
Chapel Hill, NC
Chicago, IL
Columbus, OH
Frederick, MD
Honolulu, HI
Indianapolis, IN
Metairie, LA
Naperville, IL
New York, NY
Rockville, MD
Sacramento, CA
San Mateo, CA
Waltham, MA

International

Egypt
Honduras
Ivory Coast
Kyrgyzstan
Liberia
Tajikistan
Zambia

4479_12/15