Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1218

Impacts of Michigan Transitional Kindergarten Through Third Grade

Jordan S. Berne
University of Michigan

Brian A. Jacob
University of Michigan

Tareena Musaddiq
Mathematica Inc.

Anna Shapiro
RAND Corporation

Christina Weiland
University of Michigan

Transitional Kindergarten (TK) is a relatively new model of early childhood education, with little evidence on whether and how it affects children's development. This study provides new evidence using data from Michigan, which has the nation's second-largest TK program. Using survey data (N=171) from administrators in 2021-2022, the paper documents several program features that distinguish TK from typical public Pre-K programs (including Michigan's), such as greater use of domain-specific curriculum. Focusing on students who enrolled at four years old (50% female, 78% White), a regression discontinuity analysis shows that Michigan TK improved children's readiness for kindergarten (0.91 SD, N=1,943), improved third-grade math scores (0.2-0.3 SD, N=15,680), and may have caused earlier entry into special education followed by earlier exit (N=15,704).

VERSION: June 2025

Suggested citation: Berne, Jordan S., Brian A. Jacob, Tareena Musaddiq, Anna Shapiro, and Christina Weiland. (2025). Impacts of Michigan Transitional Kindergarten Through Third Grade. (EdWorkingPaper: 25-1218). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/yzdp-ce29

Impacts of Michigan Transitional Kindergarten Through Third Grade*

Jordan S. Berne[1], Brian A. Jacob[1,2], Tareena Musaddiq[3], Anna Shapiro[4], and Christina Weiland[1]

[1]University of Michigan
[2]National Bureau of Economic Research
[3]Mathematica Inc.
[4]RAND Corporation

June 2, 2025

Abstract

Transitional Kindergarten (TK) is a relatively new model of early childhood education, with little evidence on whether and how it affects children's development. This study provides new evidence using data from Michigan, which has the nation's second-largest TK program. Using survey data (N=171) from administrators in 2021-2022, the paper documents several program features that distinguish TK from typical public Pre-K programs (including Michigan's), such as greater use of domain-specific curriculum. Focusing on students who enrolled at four years old (50% female, 78% White), a regression discontinuity analysis shows that Michigan TK improved children's readiness for kindergarten (0.91 SD, N=1,943), improved third-grade math scores (0.2-0.3 SD, N=15,680), and may have caused earlier entry into special education followed by earlier exit (N=15,704).

*The data used for this research was structured by the MERI-Michigan Education Data Center (MEDC), and is not identical to data collected and maintained by the Michigan Department of Education (MDE) or Michigan's Center for Educational Performance and Information (CEPI). Results and opinions presented here do not reflect the views or positions of MDE or CEPI. This research was funded by the Smith Richardson Foundation, as well as training grants R305B200011 and R305B170015 from the U.S. Department of Education's Institute of Education Sciences. We thank MDE, CEPI, MiLEAP, for their partnership, especially Thomas Howell, Richard Lower, and Emily Laidlaw. We thank the Washtenaw Intermediate School District for sharing Michigan's Kindergarten Readiness Assessment data, especially Naomi Norman. We also thank Katia Córdoba García and Samuel Owusu for excellent research assistance, Nicole Wagner Lam and Jasmina Camo-Biogradlija for assistance with data agreements and project management, Kyle Kwaiser for data management, Karen Manship for sharing California Transitional Kindergarten insights and survey questions, administrators who completed a survey for our study, and participants in the University of Michigan Causal Inference in Education Research seminar for excellent feedback.

1 Introduction

A large body of evidence shows that early childhood education (ECE) programs improve children's language, literacy, math, executive function, and social emotional skills at school entry (Phillips et al., 2017). ECE programs also can (but do not always) have long-run effects in K-12 and beyond, including on participants' educational, career, and health outcomes (Phillips et al., 2017; Yoshikawa et al., 2013). Such evidence, alongside the reality that all adults are in the workforce in most U.S. households with young children, has helped fuel rapid expansion of public ECE programs over the last 25 years (Cascio, 2021; Chaudry and Dutta, 2017).

As public ECE programs have expanded, states and localities have developed different models of service delivery. One new model is Transitional Kindergarten (TK), now prevalent in at least five states (Berne et al., 2025). TK programs were launched when states pushed back birthday cutoffs for kindergarten entry, providing a schooling option for children no longer eligible to attend regular kindergarten. TK programs differ from most public prekindergarten (Pre-K) programs on multiple dimensions, including eligibility criteria, teacher compensation and credentials, curriculum, and program location (Goodvin et al., 2023; Mercado-Garcia et al., 2014). For this reason, research on TK can shed light on the benefits of a *different type* of early learning program than typically found in the literature.

The existing literature on TK is quite limited. California TK is the only program to be rigorously evaluated. Studies of this program show positive effects on literacy, math, and social-emotional skills at kindergarten entry (Doss, 2019; Manship et al., 2017) and some evidence of persistence at the end of kindergarten/beginning of first grade (Doss, 2019; Manship et al., 2017). One study of five districts finds participation in California TK has no effect on standardized test scores or special education placement in third or fourth grade (Lafortune and Hill, 2023). Another study that includes all districts except Los Angeles finds positive effects on standardized test scores in third grade among children with English-speaking parents (Johnson, 2024), with work on the full sample currently underway.

We extend this literature by providing evidence from Michigan, which has the second-largest TK program in the nation. Following evidence that preschool program impacts can vary by program elements like dosage and curriculum (Atteberry et al., 2019; Phillips et al., 2017), we leverage surveys of school administrators in 2021-2022 to describe variation in TK programs across the state as well as administrator perceptions of the TK programs in their districts. We then employ a preregistered regression discontinuity research design based on birthday cutoffs to study the impact of TK participation on several critical child outcomes including kindergarten readiness, placement in special education from grades K-3, and standardized test scores in third grade.[1] We also use administrative data to describe likely counterfactual care settings, following evidence that the treatment contrast plays a large role in the detected effects of early education programs (Feller et al., 2016; Kline and Walters, 2016). As we discuss below, our findings make novel contributions to the evidence base on early learning programs, including to debates around developmental mechanisms and to the critical question of what *kinds* of programs produce the most robust lasting effects.

2 Theoretical Frame

Prior studies of the longitudinal effects of ECE have relied on multiple theoretical frames. Human capital theory in economics posits that early investments can facilitate acquisition of more advanced skills which produce lasting benefits (Heckman, 2000). Developmental cascades theory from developmental psychology predicts that earlier conditions and experiences can lead to higher or lower functioning across domains that, in turn, influence later competencies (Masten, 2010). Also from developmental psychology, Sameroff (2009)'s transactional developmental theory implies that ECE effects may (or may not) last due to reciprocal effects of higher early childhood skills on subsequent teacher behavior which, in turn, impacts child outcomes. Finally, Bailey et al. (2017) advanced three different hypotheses for lasting intervention effects: 1) "sustaining environments," meaning effects are more likely to last if subsequent experiences successfully build on children's ECE boost; 2) a "foot-in-the-door," in which benefits may last if a program gets children over an important hurdle that grants them access to a beneficial program/experience or avoids a harm; and 3) the skill target of the ECE program—i.e., for benefits to last, the skills the program focuses on must be malleable, fundamental for success, and unlikely to develop in the counterfactual (Bailey et al., 2017). Overall, it remains unclear which of these theories best explains the current empirical evidence, a topic we return to in the Discussion section.

3 Prior Empirical Evidence on Pre-K Impacts

Over the last 25 years, preschool enrollment in the U.S. has grown dramatically, fueled in large part by the expansion of publicly funded options (Cascio, 2021). By 2015, an estimated 70 percent of U.S. four-year-olds attended preschool, with 26 percent in a private program and 44 percent in a public option (i.e., 28 percent in public Pre-K, 10 percent in Head Start, and 6 percent accessing a program via a child care subsidy) (Chaudry and Dutta, 2017).

[1] Our preregistration plan is in the Registry of Efficacy and Effectiveness Studies (`https://sreereg.icpsr.umich.edu/sreereg/`) under Registry ID #13420.1v3.

As preschool enrollment in the U.S. has grown, so too has research on its effectiveness. Multiple reviews have summarized the evidence on preschool programs (Cascio, 2021; Phillips et al., 2017; Yoshikawa et al., 2013, 2016), including a meta-analysis of over 80 reasonably rigorous studies conducted since the 1960s (Duncan and Magnuson, 2012). There is broad agreement that attending preschool improves children's readiness for kindergarten compared to staying home with a family member. Across a diverse array of programs that vary in terms of region, time period, structure, and quality of inputs, research finds that preschool participation improves children's language, literacy, math, social emotional, and executive function skills.

Convincing evidence of impacts throughout grade school is more sparse (Phillips et al., 2017). Studies that follow children into elementary school generally find that children who did not attend preschool catch up (sometimes partially but often fully) with peers who attended preschool in terms of cognitive functioning and standardized test scores. Notably, some of the most rigorous studies of recent public Pre-K programs lack measures of children's skills before third grade, limiting their ability to explore how and why program effects fade-out (Braga et al., 2024; Gray-Lobe et al., 2023; Weiland et al., 2020). Regardless of whether benefits were detected in K-12 performance, most studies find that preschool improves long-term outcomes such as educational attainment, earnings, health, and intergenerational mobility (Bailey et al., 2021; Barr and Gibbs, 2022; Deming, 2009; Gray-Lobe et al., 2023; Pages et al., 2020).

Special education has long been another outcome of interest, both because it is an indicator of children's development and because of the cost of providing services. A 2017 meta-analysis of experimental and quasi-experimental evaluations concluded that early education programs reduced special education placements by 8.1 percentage points on average (McCoy et al., 2017). However, in more recent studies not included in the review, special education findings are more mixed. Studies of Tulsa Pre-K, the Tulsa Community Action Program (CAP) Head Start, DC Universal Pre-K (UPK), and Boston Pre-K find null effects (Braga et al., 2024; Gormley et al., 2018; Phillips et al., 2016; Weiland et al., 2020). Dodge et al. (2017) find that North Carolina Pre-K programs reduced participation in special education programs, while Durkin et al. (2022) find that Tennessee programs increased special education rates.

4 Background on Transitional Kindergarten

4.1 Key Features of TK Programs

TK programs are now prevalent in at least five states (CA, MI, WA, MT, and IA), where they exist alongside publicly funded, income-targeted early learning programs.[2] TK programs differ from state to state but share several research-aligned features that set them apart from most public Pre-K programs.

First, TK programs tend to be open to all age-eligible children regardless of family income. One study suggests universal programs may enhance children's acquisition of critical early skills more than income-targeted programs do (Cascio, 2023). Second, TK programs typically match the K-12 calendar—i.e., full-day, five days per week, for the full K-12 school year—while Pre-K programs are often available only part-day and/or part-week (Friedman et al., 2024). A recent randomized trial highlights the importance of Pre-K dosage, finding that children learn more in full-day programs than half-day programs (Atteberry et al., 2019).

Third, unlike most Pre-K programs, TK is funded at the same per-child rate as K-12 grades. More commonly, public Pre-K expenditures fall far short of K-12 benchmarks. As a result, Pre-K teachers are typically not compensated at parity with TK and K-12 teachers (Friedman et al., 2024). Relatedly, Pre-K teachers in many programs do not have to meet the same educational and certification requirements as K-12 teachers. To take our focal state of Michigan as an example, state Pre-K funding in 2020-2021 was $8,700 per child, compared to $14,347 in TK and K-12 (Weiland et al., 2023). Michigan public Pre-K teachers must obtain a BA but not the same certification as TK/K-12 teachers. In 2021-2022, salaries for public Pre-K teachers were 31 percent less than for TK/K-12 teachers (Wu et al., 2023). Wage and funding gaps have long fueled preschool teacher turnover and undermined investments in quality (Markowitz and Bassok, 2025), which in turn could negatively affect child development (Weiland and Rosada, 2025).

Fourth, TK programs are offered exclusively in public school settings, in contrast to public Pre-K which is often offered in a mix of public schools and community-based settings (Friedman et al., 2024). Although high-quality preschool is achievable in both settings, policies often place community-based programs at a disadvantage in terms of teacher wages and other supports (Weiland et al., 2024). As a result, when there are differences in quality and child learning in mixed-delivery Pre-K systems, they tend to favor programs in public schools.

Finally, there is some evidence that TK programs are more likely to utilize domain-specific

[2]Of the five states, all but Montana also offer a state-funded public Pre-K program. All five offer Head Start and child care subsidies that facilitate preschool access (Friedman et al., 2024)

curriculum than public Pre-K programs. State-funded Pre-K and Head Start programs tend to use comprehensive curriculum—most commonly, Creative Curriculum or HighScope—that purport to cover all child domains but that do not have a strong track record of improving children's learning compared to more focused, domain-specific curricula (Jenkins et al., 2018; National Academies of Science and Medicine, 2024; Yoshikawa et al., 2016). In contrast, most TK programs in California use domain-specific curriculum (Manship et al., 2015). In Washington State, 80 percent of TK programs use a comprehensive curricula (with about 40 percent using the Creative Curriculum or HighScope), although most programs layer in domain-specific curricula too. In Washington's state-funded Pre-K program, nearly all classrooms use Creative Curriculum (Goodvin et al., 2023).

4.2 Prior Evidence on TK

Research on TK is sparse and so far largely limited to California. Manship et al. (2015) find positive impacts of California TK on children's literacy, language, math, and engagement at kindergarten entry as well as higher literacy scores at the end of kindergarten. Doss (2019) finds positive impacts from TK in San Francisco on literacy and language outcomes in the fall of kindergarten and on language in the fall of first grade. A recent study that follows children in five CA districts finds no impact of TK participation on standardized test scores in third or fourth grade (Lafortune and Hill, 2023). However, a study that examines children with English-speaking parents in all CA districts except Los Angeles finds that children who attended TK programs outperformed their peers on third-grade standardized math and English Language Arts (ELA) exams (Johnson, 2024). The effects are particularly large for children in better-funded districts. For special education, California TK appears to increase the likelihood of having an individualized education program in kindergarten through second grade with no meaningful differences in third through fifth grade (Lafortune and Hill, 2023).

Our team previously described how TK fits in Michigan's early learning landscape (Berne et al., 2025). We found that districts with more White and fewer economically disadvantaged students were more likely to offer TK than other districts. Within TK-offering districts, we found some demographic differences in child enrollment patterns. For example, White children, Hispanic children, and children of another race/ethnicity were roughly twice as likely to enroll in TK compared with Black and Asian children. Children who were economically disadvantaged were also less likely to enroll relative to their peers. We also tracked enrollment changes in several early learning programs following the introduction of TK programs in Michigan. We found a shift into TK and away from entering kindergarten early with a waiver, resulting in a net increase in the total share of Michigan children served by public early learning programs. This research, to our knowledge, provides the first evidence on how TK changes the early learning landscape and provides important context for the present study.

For child impacts, we previously used the same augmented regression discontinuity design that we leverage in this study to estimate the impact of participation in Michigan TK on third-grade math and ELA scores. These findings appeared in non-peer reviewed conference proceedings (Berne et al., 2024). We found that participation in TK increased third-grade math scores by roughly 0.29 standard deviations. The point estimate for ELA was 0.19 standard deviations, but less precise and not statistically significant (p=0.29). We build on these findings in the present study.

5 The Present Study

In the current study, we extend our previous work in several important ways to provide a holistic understanding of Michigan TK. First, we draw on a survey of district administrators we conducted to characterize the structural and institutional features of TK programs across the state. The survey also allows us to explore how administrators view the benefits of TK programs. Given the relative lack of evidence on how TK implementation varies across and within states, these data provide new insight into the critical question of what TK is and how it differs from other early learning programs.

Second, we examine the impact of TK on children's kindergarten readiness and participation in special education programs through third grade. For special education, as we described, the recent literature has been mixed for Pre-K and evidence on this outcome for TK so far has been limited to five districts in California (Lafortune and Hill, 2023). Special education systems vary substantially across states, making it especially important to examine TK's effects on this outcome outside of California. We also include students' math and ELA achievement in third grade to provide a more holistic picture of TK's effects in early elementary school and because we conduct additional sensitivity checks for these outcomes not included in Berne et al. (2024). Notably, ours is the first study of a TK program to include both kindergarten readiness and third-grade test scores.

Our specific research questions are: (1) How do structural and instructional features vary across Michigan TK programs and what are administrators' perceptions of the benefits of TK? (2) What is the impact of Michigan Transitional Kindergarten on children's kindergarten readiness, special education placement through third grade, and third-grade state standardized reading and math test scores?

6 Setting

Michigan TK is a district-led, publicly funded early learning option for four- and/or five-year old children in the year before regular kindergarten. Districts are not required to offer TK programs but rather choose whether and in which schools to offer them.

The birthdate cutoff for kindergarten eligibility in Michigan is September 1, meaning that children who turn five years old by this date are recommended to start kindergarten in that year. Children who turn five between September 2 and December 1 can attend kindergarten if they receive a waiver. Children born on December 2 or later must wait until the following school year to enroll. Figure 1 illustrates these program eligibility rules.

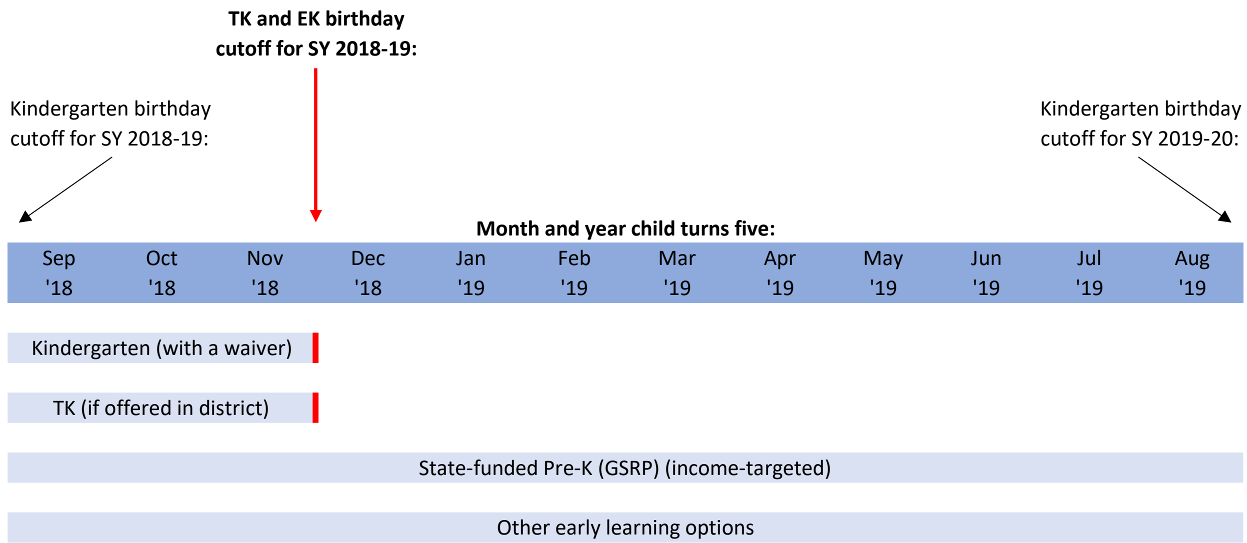
All districts with TK allow children born between September 2 and December 1 to enroll in TK. Most districts also allow slightly older children (i.e., those with earlier birthdates) to enroll in TK, so it is common to observe children with summer birthdays in TK. As explained below, however, our impact analysis focuses on children with fall birthdates who have two options within the public K-12 system: they can either obtain a waiver to attend kindergarten early (which we will denote as EK) or enroll in a Transitional Kindergarten (TK) program if their district offers it. Children who do not enroll in EK or TK can enroll in Head Start or the state's income-targeted Pre-K program, GSRP, if they meet eligiblity criteria. They could also enroll in a private preschool program or a licensed family child care home. Otherwise, these children are cared for by family or friends or participate in some other type of informal option.

As described in Section 4, TK programs have several common structural features, including co-location in public schools alongside kindergarten programs and certification requirements and salary levels that match those of kindergarten teachers. For instructional features such as curriculum and assessment, districts have wide latitude. These program features are not tracked by the state, which is why we conducted a survey of TK administrators to measure implementation across the state.

7 Data, Samples, and Measures

We use a combination of survey and administrative data to answer our research questions. In this section, we describe the data sources, explain how we created the key outcome measures, and present summary statistics for our impact analysis sample.

Figure 1. Program Eligibility in Michigan in SY 2018-2019, by Month Child Turns Five



Note: This figure shows the early childhood education programs Michigan children in the 2018-2019 Pre-K cohort were eligible for based on their month of birth. All children born in the months shown were scheduled to start kindergarten in school year 2019-2020 based on state guidelines. However, the oldest children in this cohort, born between September 2 and December 1, were eligible to enroll in kindergarten in 2018-2019 with a waiver. These children were also eligible to enroll in Transitional Kindergarten. All children in this cohort were age-eligible for Michigan's income-targeted Pre-K program (GSRP) in 2018-2019, although this program also had eligibility criteria based on family economic characteristics.

7.1 Survey Data

To address our first research question, which focuses on features and perceptions of TK, we contacted administrators in Michigan districts and charter schools with TK programs in the 2021-2022 school year to request their participation in an online survey. In TK-offering districts, we used district websites and phone calls to identify respondents, with a priority order within each district of superintendent, curriculum coordinators, and then building administrators/principals. We contacted respondents up to three times each via email and seven times each by phone asking them to consent to participate in the study and to complete our survey. In some cases, a respondent identified for us the best person to complete the survey in their district (i.e., a superintendent would refer us to the curriculum coordinator). As needed, we identified building administrators/principals by determining which elementary schools offered TK, and if there was more than one that did so, we randomly selected one to contact.

In all, we contacted 335 of the 364 districts and charter schools with Michigan TK programs in 2021-2022. We did not survey 29 districts and charters because we confirmed they had a TK program after our survey window had closed. In all, an administrator from 171 districts/charters consented to participate and responded to our survey, representing 47 percent of active TK programs across Michigan in 2021-2022 and 45 percent of students in grades 1-5. 40 percent (69) of survey respondents were principals/building administrators, 24 percent (42) were curriculum coordinators, and 21 percent (37) were district superintendents. The survey took about 25 minutes to complete and we offered a $100 incentive to participate.

In Appendix Table A1, we show that districts with TK that responded to the survey are demographically similar to districts with TK that were not surveyed or did not respond to the survey. In contrast, charter schools with TK that responded to the survey serve more advantaged populations than charter schools with TK that did not. We return to this limitation in the Discussion section.

7.2 Administrative Data

To address our second research question, we use longitudinal administrative data on the universe of Michigan public school students. These data contain student demographics, school and district enrollment, placement in special education, and third-grade test scores.

7.2.1 Third-Grade Sample

To estimate impacts on special education and third-grade test scores, we limit our sample in three main ways. First, we restrict the sample to students with a math or ELA test score in third grade. This restriction is not strictly necessary for estimating special education impacts, but it ensures that our sample is consistent across outcomes.

Second, we restrict our sample to districts that (i) do not offer TK or (ii) offer TK and also reliably report TK enrollment at the student level. Our early analysis of administrative data along with survey responses from district leaders (Shapiro et al., 2023) revealed that some school districts in Michigan do not report TK enrollment accurately. (Appendix B provides details on how we identify districts with reliable data.) In prior work, we found that 60 percent of districts and charter schools with TK programs accurately report TK enrollment (Berne et al., 2025). These districts/charters tend to be larger than those without accurate reporting, so this sample covers 75 percent of all grade 1-5 students in TK districts.[3]

[3] In school year 2021-2022, we undertook an extensive data triangulation process to identify which districts offered TK programs in that year. Appendix Table B1 of Berne et al. (2025) shows that TK districts that accurately reported enrollment were similar to TK districts that did not accurately report enrollment in terms of racial composition but served a somewhat higher proportion of economically disadvantaged students and were more likely to be in towns and rural areas. We suspect the results would be similar for TK districts in 2014-2015 and 2018-2019.

Table 1. Summary Statistics for Third-Grade Sample

	Non-TK Districts		TK Districts		
	All Students	Early K Students	All Students	Early K Students	TK Students
2015 Pre-K cohort (%)	50	51	38	46	31
2019 Pre-K cohort (%)	50	49	62	54	69
Female (%)	50	56	50	58	50
White (%)	48	33	75	65	78
Black (%)	36	52	12	19	9
Hispanic (%)	10	9	7	6	7
Asian American (%)	4	5	5	9	4
Other race (%)	2	1	1	1	1
Economically disadvantaged (%)	70	78	45	58	38
Prior state-funded pre-K enrollment (%)	5	14	4	13	8
English learner (%)	11	15	9	20	6
Neighborhood White share (%)	64	52	85	82	86
Neighborhood poverty share (%)	17	21	9	13	8
Neighborhood unemployment rate (%)	13	12	20	19	20
Neighborhood BA attainment rate (%)	13	12	20	19	20
Neighborhood median household income ($)	49,720	45,745	66,989	61,479	69,129
School is in a city (%)	38	47	19	32	18
School is in a suburb (%)	31	34	50	46	50
School is in a town (%)	7	5	12	9	11
School is in a rural area (%)	23	14	19	13	21
Magnet school (%)	19	20	8	7	7
School enrollment	444	488	445	442	431
School pupil:teacher ratio	17.9	18.6	17.2	16.9	16.4
School FRL share (%)	67	73	43	50	41
Charter school (%)	30	47	3	4	5
District is in a city (%)	40	49	20	34	18
District is in a suburb (%)	30	33	55	49	57
District is in a town (%)	8	5	13	9	11
District is in a rural area (%)	23	13	13	7	13
District free- and reduced-price lunch share (%)	64	70	39	44	38
District English learner share (%)	11	11	7	12	6
District avg. 3rd grade math M-STEP score (SD)	−0.25	−0.34	0.25	0.18	0.24
Observations	7,056	1,467	8,648	832	1,851

Note: This table uses the third-grade sample described in Section 7.2.1, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. This table restricts the sample to those with non-missing outcome data; summary statistics for the full sample are essentially identical, as shown in Appendix Table B1. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Third, we also limit our sample by student age and include only students who turned five near December 1 in 2014 *or* 2018. Using the spring of the school year, we refer to these as the 2015 and 2019 cohorts. As explained further below, our main analysis focuses on students born within one month before or after December 1 in each year. We do not include students from the 2015-2016, 2016-2017, or 2017-2018 cohorts because their third-grade tests (one of our main outcomes) were disrupted by the COVID-19 pandemic.[4] Further data cleaning decisions and sample restrictions are discussed in Appendix B.

Table 1 compares the characteristics of students in our sample based on whether their district offered TK. Students in TK districts were more likely to be White, less likely to be Black, less likely to be economically disadvantaged, and more likely to be located in suburbs. Within the same district, students enrolled in TK tended to be slightly more economically advantaged and more likely to live in a suburb compared to the average child in the district. We provide further detail on districts with TK programs and the students who participate in TK in Shapiro et al. (2023).

7.2.2 Kindergarten Readiness Sample

To analyze TK's impact on kindergarten readiness, we obtained Kindergarten Readiness Assessment (KRA) data, which we merged to state administrative data at the student level. Some specifics of Michigan's KRA rollout affect data availability in our study. First, KRA data are available for our 2018-2019 cohort but not our 2014-2015 cohort as Michigan did not have a state kindergarten readiness assessment in place in fall 2014.

Further, not all districts administered the KRA, and within districts that did, not all students were assessed. Districts had the choice of assessing all entering kindergartners or a randomly selected subset (with about eight per class as the target). For our primary KRA analysis, we restrict the sample to districts in which at least 20 percent of students in the 2018-2019 cohort were assessed. 61 TK districts and 58 non-TK districts meet this threshold (out of 505). See Appendix F for details on this sample, including Table F6 for student summary statistics. Districts in this sample had lower rates of early kindergarten enrollment than the broader set of districts in Table 1. Otherwise, they were very similar in terms of observable characteristics.

Despite state policy, our analysis of the KRA data suggests assessments were not administered to students randomly; White and non-economically disadvantaged students were systematically more likely to be assessed. This selection could bias our results if it is cor-

[4] As explained more below, our analysis of each cohort requires test score data from two academic years because children who waive into regular kindergarten early typically reach third grade and take third-grade exams a year earlier than other children in their age cohort.

related with TK and EK enrollment. To ensure our results are robust, we also conduct our analysis on samples restricted to districts in which at least 80% or 90% of students were assessed. In these districts, the scope for bias from non-random assessment is minimal. In Appendix F, we provide more details about these samples and show that our results are qualitatively similar across samples.

7.3 Measures

To address our first research question, we use administrator responses to a set of questions about the key features of their TK programs and Likert-scaled items on their perceptions of TK's benefits for students. We developed these survey questions through a small number of initial interviews with administrators and teachers about Michigan TK's goals and features and by reviewing TK surveys used in California (Manship et al., 2017) and surveys used in the early grades in the Boston Early Learning Network study (McCormick et al., 2022). We piloted the survey with a small number of respondents who we excluded in the main survey data collection.

In the survey, structural feature items included a mix of continuous and yes/no response items (e.g., class size cap was continuous whereas having a full-time teaching assistant/paraprofessional was dichotomous). We created a list of commonly used curriculum and assessments in early childhood education programs and asked respondents to indicate which were used in their TK program. We also provided an open-response "other" category for unlisted curricula and assessments. Using descriptions from each publisher, we coded whether districts/charters were using domain-specific curricula (for language/literacy, math, and social emotional skills); comprehensive curriculum that purport to cover all domains; and/or district-created curriculum only (for which we were unable to determine the domains covered). We followed a similar procedure for assessment domains. We also created a categorical variable that captures whether each district primarily used curricula targeted to Pre-K, kindergarten, or an equal mix.

Finally, for perceptions of TK, we asked respondents 19 items that had a four-point Likert scale (1=strongly disagree; 4=strongly agree), along with an N/A option. The specific question we asked was, "To what extent do you agree with each of the following statements about offering Transitional Kindergarten in your district?" The statements included eleven positively valenced items (e.g., "TK gives time for social-emotional development); five negatively valenced items (e.g., "Children with strong academic skills will be bored"); and three neutral items (e.g., "TK and K instruction look about the same"). We report average scores and standard deviations, as well as the distribution across response categories.

To assess our second research question (child impacts), our key outcome measures fall into three domains: children's kindergarten readiness, special education in grades K-3, and third-grade state standardized test scores. These are measures of children's learning and school progress that meet guidelines in the field for meaningful educational indicators that are not overly aligned with the intervention (What Works Clearinghouse, 2022). We preregistered one confirmatory outcome: whether a child ever participated in special education by the end of second grade. We did so because under Michigan's special education law, children classified as having a developmental delay must be re-evaluated once they turn 7, with most children reaching this age in first grade (Office of Innovative Projects, 2018). Since the special education identification and placement process takes time to complete, we use second grade for our measure. Other outcomes were preregistered as exploratory. Below we describe these measures and provide psychometric details.

Kindergarten readiness. Following state policy, teachers assessed students at the beginning of kindergarten using the Kindergarten Readiness Assessment. The KRA has been extensively researched for a relatively new measure, with evidence of solid predictive validity with third-grade outcomes (r=0.46-0.49 across domains), solid discriminant validity (subscale correlations between 0.42-0.76), and high internal consistency and convergent validity across items (Justice et al., 2019; Maryland State Department of Education, 2019; WestEd, 2015). There are 17 language and literacy items (15 directly assessed, 2 observational); 14 mathematics items (all directly assessed); 7 physical well-being and motor development items (all observational); and 12 social skills items (all observational). In our analysis, we use both the composite score and individual subscale scores as outcomes. For ease of interpretation, we standardize all KRA scores using means and standard deviations from the universe of KRA test-takers in Ohio in the relevant cohorts (Table F1).

Special education placement. From students' state records, we create dichotomous indicators of whether children had an individualized education program (IEP) in each grade from TK to third grade. We also create a summary indicator for having an IEP anytime between kindergarten and second grade.

Third-grade standardized test scores. In Michigan, students in grades 3-8 take a state-mandated assessment (M-STEP) that uses items drawn primarily from the Smarter Balanced Assessment Consortium item pool that are aligned to Michigan's content standards. M-STEP scale scores are calculated from the total number of raw points a student scores on each item and converted along a standardized scale to allow comparisons over time and across test forms. An analysis of student growth measures derived from the M-STEP found that scale scores are highly correlated within students over time in both reading (r=0.82-0.84) and math (r=0.84-0.87) (Data Recognition Corporation, 2018). In our analysis, we use students' third-grade scores on the English Language Arts (ELA) and Mathematics (math) M-STEP assessments.

8 Empirical Strategy

To explore the features of Michigan TK programs and administrators' perceptions of TK, we calculate and discuss simple descriptive statistics (means and standard deviations) from our administrator survey. To examine the causal effect of TK participation on kindergarten readiness, receipt of services for students with disabilities, and third-grade achievement scores, we use a regression discontinuity (RD) design. RD is a common quasi-experimental method for measuring the impact of an intervention that is assigned on the basis of a continuous variable with a discrete cutoff (Bloom, 2022). Under certain assumptions, by comparing individuals on either side of this cutoff, one can estimate the causal impact of being *eligible* for the intervention—which is referred to in the literature as the intent-to-treat (ITT) effect. With additional assumptions, one can estimate the causal impact of actually receiving the treatment, which is referred to as the local average treatment effect (LATE).

In our case, the continuous variable that determines treatment eligibility is a child's birthdate. As discussed earlier, in districts that offer TK, children who turn five before December 2 of a given year are eligible to enroll in the program. Children born before December 2 (in any district) are also eligible to apply for a waiver to enroll early in regular kindergarten. We denote these children as EK students to distinguish them from children who enroll in Transitional Kindergarten, who we refer to as TK students. Children born on December 2 or after are eligible for neither TK nor EK.

Note that not all children who are eligible for TK and EK enroll in these programs. Such situations with incomplete take-up are referred to in the literature as "fuzzy" RDs. The instrumental variables method we describe below is designed to analyze such data.

The fact that the same birthdate cutoff (December 1) is associated with two different potential treatments (TK and EK) makes identification of causal effects in a standard RD model challenging. In the subsections below, we explain how we address this additional complication, carefully describing our estimation strategy and the assumptions underlying our approach.

8.0.1 Identification with Multiple Treatments

At its core, an RD analysis involves comparing individuals on either side of a cutoff. In Michigan, children born on December 2 or later cannot enroll in TK or EK. In practice, these children either enroll in a formal preschool program, stay at home with family, or participate in some other informal care option. Children born on December 1 or earlier can (and do) enroll in either TK or EK. Hence, if we compare the outcomes of children on either side of the December 1 cutoff, our estimates will capture the *combined* effect of being eligible for *both* TK and EK.

In districts that offer TK, one can show that the ITT effect is a weighted combination of the TK and EK local average treatment effects (LATEs):

$$ITT = \Omega_{TK} LATE_{TK} + \Omega_{EK} LATE_{EK} \tag{1}$$

where each weight, Ω_x, is the share of students at the cutoff who are compliers for option x. See Appendix C for the derivation. The quantities ITT, Ω_{TK}, and Ω_{EK} can be estimated using data from districts that offer TK. However, the equation above has two unknown parameters: $LATE_{TK}$ and $LATE_{EK}$.

Our strategy is to use data from Michigan school districts that do *not* offer TK to estimate the effect of enrolling in kindergarten early (i.e., $LATE_{EK}$), which then allows us to back out the TK effect (i.e., $LATE_{TK}$) using equation 1. In districts that do not offer TK, the December 1 birthday cutoff only matters for early kindergarten enrollment. For this reason, a standard RD approach in non-TK districts allows us to cleanly identify the causal effect of enrolling in kindergarten early (i.e., $LATE_{EK}$).

Using the effect of early kindergarten in *non-TK districts* to infer the same effect in districts that *do offer* TK requires an additional assumption—specifically, that the effect of participating in EK does not differ systematically across districts in a way related to whether districts offer TK. There are reasons to be concerned that this might not be true. As shown in Table 1, EK students in TK and non-TK districts differ in several ways. Relative to their peers in TK districts, children who attend EK in non-TK districts are more likely to be economically disadvantaged (78% vs 58%), less likely to be English learners (15% vs. 20%) and more likely to be Black (52% vs. 19%). Districts that do and do not offer TK also differ in terms of urbanicity, neighborhood income, neighborhood education levels, and the presence of charter schools nearby. Given these differences, one might be concerned that the effect of attending EK in non-TK districts will not generalize to TK districts. We therefore estimate two models—one that assumes treatment effect homogeneity and another that relaxes this assumption.

8.0.2 Baseline Estimation Approach

Our baseline model assumes that the treatment effect of EK is the same, on average, in TK and non-TK districts. Following the standard fuzzy RD literature, we use a two-stage least squares (2SLS) approach to jointly estimate the treatment effects of attending EK and TK, using both districts that do and do not offer TK. For student i in district d from cohort c, we estimate the following system of equations via 2SLS:

$$Y_i = \beta_0 + \beta_1 TK_i + \beta_2 EK_i + f(dob_i) + \Pi \mathbf{X}_i + \lambda_{dc} + \varepsilon_{idc} \tag{2}$$

$$EK_i = \delta_0^{EK} + \delta_1^{EK} Left_i + \delta_2^{EK} Left_i \times DistHasTK_{dc} + f^{EK}(dob_i) + \Psi^{EK}\mathbf{X}_i + \theta_{dc}^{EK} + \epsilon_{idc}^{EK} \tag{3}$$

$$TK_i = \delta_0^{TK} + \delta_1^{TK} Left_i + \delta_2^{TK} Left_i \times DistHasTK_{dc} + f^{TK}(dob_i) + \Psi^{TK}\mathbf{X}_i + \theta_{dc}^{TK} + \epsilon_{idc}^{TK} \tag{4}$$

where $\mathbf{X}_i$ is a vector that includes student sex, race, and economic disadvantage status; λ_{dc}, θ_{dc}^{TK}, and θ_{dc}^{EK} are district×cohort fixed effects; and dob_i is date of birth. The f functions allow different linear relationships between date of birth and outcomes on either side of the cutoff and across districts with and without TK.

The "first-stage" equations 3 and 4 model the likelihood that a child participates in EK or TK. The key predictors in these models are $Left_i$ and $Left_i \times DistHasTK_{dc}$, which indicate (i) whether student i was born on or before the RD cutoff of December 1, and (ii) being born on or before the RD cutoff *and* being in a district×cohort in which TK is offered. Given the eligibility rules discussed above, we expect δ_1^{EK} to be positive but δ_2^{EK} to be negative, assuming that the TK option draws some students away from EK. We would expect δ_1^{TK} to be zero (because students cannot enroll in TK if the district does not offer it) and δ_2^{TK} to be positive. These four variables serve as excluded instruments in the model. Equation 2 is the second-stage or structural equation. $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimates for $LATE_{TK}$ and $LATE_{EK}$, respectively.

We use a bandwidth of ±30 days around the cutoff when estimating this set of equations. Following the standard in the RD literature, we conduct inference with clustering by values of the assignment variable, which in our case is student birthdate. In Appendix G, to test the robustness of our findings, we explore a variety of alternative specifications for this model and find comparable results.

8.0.3 Relaxed Assumptions Approach

As discussed above, it is possible that the treatment effect of EK differs across TK and non-TK districts. In particular, one might be concerned that effects differ across districts because the students themselves differ demographically. In this alternative approach, we relax the assumption of treatment effect homogeneity by allowing the EK treatment effect to differ by student demographic characteristics. We first estimate EK effects in the *non-TK sample* separately for 8 demographic groups defined by sex × race (White or Asian vs. other races) $\times$ economic disadvantage status (disadvantaged vs. not). We estimate these effects using a 2SLS model analogous to Equations 2 and 3. This allows participation in EK, for example, to have a different effect on economically disadvantaged, Black boys compared to non-economically disadvantaged White girls. Appendix E shows estimates for each of these subgroups.

Next, we turn to the *TK sample* and calculate the share of early kindergarten students in each of the 8 demographic groups in the TK sample. Using these shares as weights, we calculate a weighted average of the demographic group-specific effects from above. The weighted average gives us a single EK treatment effect estimate, $\widehat{LATE}_{EK}$, that reflects the group-specific EK effects from the non-TK districts and the demographic composition of EK students in TK districts.

Finally, we use this "demographically-adjusted" estimate of $LATE_{EK}$ to back out $LATE_{TK}$ using Equation 1 and estimates of ITT, Ω_{TK}, and Ω_{EK} from the TK sample. Arriving at $LATE_{TK}$ using this relaxed assumption approach requires multiple steps. We therefore conduct inference using a bootstrap. In tables below, we report p-values rather than standard errors because the bootstrap distributions in this alternative approach are non-normal and contain extreme outliers, rendering the standard errors uninformative. For consistency, we also report bootstrap-based p-values for our baseline approach. Appendix H provides more details.

8.0.4 Internal Validity Checks

As with all regression discontinuity studies, the internal validity of our findings depends on meeting several critical assumptions. We conduct a variety of analyses to assess these assumptions. We briefly describe the results of these tests here, deferring a more detailed discussion to Appendix D. While there are several small concerns, overall we believe we have a valid natural experiment.

First, we examine whether the assignment variable (i.e., birthdate) may have been manipulated or misreported. From a practical perspective, it seems unlikely that families manipulated their child's birthdate to gain access to TK or EK. It is possible, however, that some families relocated to districts that offered TK to gain access. To examine this threat, we use two formal tests: (1) the McCrary (2008) test with bandwidths of 5 and 10 days from the cutoff, and (2) the Cattaneo et al. (2020) test which uses a mean squared error minimizing procedure to determine an optimal bandwidth. We show our results in Appendices D and F (Tables D2 and F4 and Figures D1 and F1), where we also provide a full explanation of our findings. To summarize, we find mixed evidence about potential discontinuities in density. We find no evidence in TK districts in the third-grade sample or in non-TK districts in the KRA sample. In some specifications, though, we estimate significant discontinuities in non-TK districts in the third-grade sample and in TK districts in the KRA sample. However, as we discuss in Appendix D, there is no clear story that explains these estimates, and they are only problematic if there are systematic differences in students on either side of the cutoff related to our outcomes. As we discuss next, we find no evidence of such systematic differences. Accordingly, we view this finding as likely driven by noise and unlikely to bias our estimates.

Second, we find no systematic differences in student background on either side of the birthdate cutoff. Appendix Tables D3 (third-grade sample) and F5 (KRA sample) show estimates for 21 separate characteristics as well as a composite measure of all characteristics. Most estimates are small and indistinguishable from zero. Importantly, there are no significant differences for our composite measure.

Finally, for the third-grade sample, we find no evidence that our results will be biased due to differential attrition. In our study, attrition could have occurred via students leaving the Michigan public school system in the years after TK/EK eligibility or by not taking third-grade standardized exams. As we show in Appendix Table D1, nearly all students enroll in Michigan public schools and have observed special education outcomes in grades 1 through 3 (around 98% in first grade, 95% in second grade, and 91% in third grade), with no evidence of differential attrition at the cutoff. We observe third-grade test scores for roughly 86% of of students, but again there is no difference in data availability at the cutoff.

As previously discussed, the nuances of KRA administration in Michigan open the door for differential assessment rates at the cutoff, which we do find (Appendix Table F3). Nevertheless, children's characteristics are generally smooth through the RD cutoff. Still, to ensure our results are not driven by differential assessment, we test sensitivity using districts with particularly high assessment rates, where the scope for bias from differential assessment is minimal (Tables F10-F14). Reassuringly, our estimates are very similar across samples. Lastly, we also show in a bounding exercise (Table F16) that imputing missing KRA scores with near-worst-case values does not overturn our finding of large positive effects.

Together, these results provide strong evidence that children born just before and just after the RD birthday cutoff are equivalent at baseline, on average, other than their eligibility for TK and EK enrollment.

9 Results

9.1 Features and Perceptions of TK

As we show in Table 2, TK programs in Michigan look remarkably similar to each other in their structural features, despite the wide latitude districts have over their programs. For example, nearly all programs are full-day and full-year, meaning they mirror the K-12 schedule and run roughly six hours per day for about 185 days per year. The average TK classroom is capped at 19 students. 89% of districts have standalone TK classrooms (i.e., they do not have classrooms that include TK *and* regular kindergarten students) and 69% employ a full-time or part-time teaching assistant alongside the TK teacher.

Unlike other preschool programs, TK programs in Michigan are more likely to use targeted math and/or literacy curricula rather than comprehensive curricula. State-funded Pre-K and Head Start in Michigan require the use of a comprehensive curriculum. 84% of Michigan state-funded Pre-K classrooms use Creative Curriculum or HighScope (Weiland et al., 2023). In contrast, only 8% of Michigan TK programs reported using Creative Curriculum or HighScope, and only 16% reported using any comprehensive curriculum. Instead, 94% reported using a specific literacy curricula and 69% reported using a targeted math curricula; 66% of responding districts reported using curricula covering both literacy and math. In terms of curriculum level, an equal number of programs use curricula that are designed for kindergarten-age children (38%) and preschool-age children (40%), with 22% using a mix of Pre-K and kindergarten level curricula.

For assessments, administrators reported using 2.4 assessments on average in TK (SD=1.1, range 1-6), with 85% reporting using the same assessments in TK and kindergarten. Most administrators (88%) reported using at least one assessment in which the child is directly assessed, while just under half reported using an assessment based on teacher observations of the child in the classroom. Literacy was the most commonly assessed area (94%), followed by language (61%), math (60%), and social-emotional skills (42%). 40% of districts created their own assessments, for which we could not determine content.

For perceptions of TK, Table 3 indicates that district administrators have an extremely positive view of their TK programs. For example, nearly all agreed or strongly agreed that TK programs benefit children's social emotional development (99%), academic skills (99%), and school enjoyment (94%). The majority of respondents disagreed or strongly disagreed with negative statements regarding TK. For example, 75% disagreed or strongly disagreed that children with strong academic skills would be bored in TK.

Table 2. Features of Michigan TK Programs

	% or mean (SD)
Structural features	
Class size cap	19 (3)
Full-day	97%
Full week	99%
TK-only classrooms	89%
Enroll children with IEPs/inclusion model	86%
Full-time teaching assistant	35%
Part-time teaching assistant	34%
Teacher PD hours per year	34 (11)
Offered in all elementary schools	63%
Curricula	
Language/literacy	94%
Math	69%
Social emotional	37%
Global	16%
District-created only	3%
Curriculum grade level	
Primarily Pre-K	40%
Primarily K	38%
Equal split between Pre-K and K	22%
Assessments	
Language	61%
Literacy	94%
Math	60%
Social emotional	42%
District-created	40%

Note: N ranges from 153 to 171 across questions. In 2021-2022, we contacted 335 of the 364 districts and charter schools that offered TK programs in that year. Administrators from 171 districts/charter responded to the survey, including 69 principals/building administrators, 42 curriculum coordinators, and 37 district superintendents. For curriculum and assessment, we coded domains and grade from district-reported checklists of the curriculum/assessments used in TK.

Table 3. Administrators' Perceptions of Michigan TK Programs

	Strongly Disagree	Disagree	Agree	Strongly Agree	Mean	SD
	Panel A. Positively Valenced Items					
TK is worthwhile/necessary	1%	2%	13%	82%	3.8	0.5
TK gives students the gift of time	1%	3%	17%	78%	3.7	0.5
TK gives time for social-emotional development	0%	1%	13%	85%	3.8	0.4
TK boosts developmentally-appropriate academic skills	1%	2%	17%	80%	3.8	0.5
TK students are more likely to read at grade-level in elementary school	1%	9%	39%	47%	3.4	0.7
TK students are better prepared for K than their peers	0%	15%	40%	41%	3.3	0.7
TK students are better at following directions than peers in K	0%	16%	47%	31%	3.2	0.7
TK students are more independent than peers in K	0%	14%	49%	32%	3.2	0.7
TK students are leaders in K	0%	20%	43%	31%	3.1	0.7
TK promotes long-term enjoyment of school	0%	3%	48%	46%	3.4	0.6
Most children who attend TK benefit from the program	1%	8%	23%	64%	3.6	0.7
	Panel B. Negatively Valenced Items					
Children with fall birthdays should go to pre-K, not TK	32%	53%	10%	1%	1.8	0.7
TK students should not learn reading and math	47%	37%	10%	4%	1.7	0.8
TK is not worthwhile/necessary for my district	25%	43%	23%	6%	2.1	0.9
Children with strong academic skills will be bored	18%	57%	20%	4%	2.1	0.7
Children with strong social skills will not find TK challenging	20%	62%	16%	1%	2.0	0.6
	Panel C. Other Items					
Learning math and literacy are equally important in TK	0%	7%	35%	57%	3.5	0.6
It is clear to me what is expected of students after completing TK	2%	4%	39%	52%	3.5	0.7
TK and K instruction look about the same	7%	49%	32%	10%	2.5	0.8

Note: N ranges from 160 to 164 across questions. In 2021-2022, we contacted 335 of the 364 districts and charter schools that offered TK programs in that year. Administrators from 171 districts/charter responded to the survey, including 69 principals/building administrators, 42 curriculum coordinators, and 37 district superintendents. To compute response means and standard deviations, responses are valued at "strongly disagree"=1, "disagree"=2, "agree"=3, and "strongly agree"=4. The shares in columns 1 through 4 do not sum to 100% due to rounding and because for each question a small share of respondents (ranging from 1% to 6%) selected "not applicable."

9.2 Child-Level Impacts

9.2.1 Visual Evidence of Treatment Effects

Figure 2 shows the relationship between a child's birthdate and their likelihood of enrolling in TK or waiving into kindergarten. The top panel presents results from the KRA sample; the bottom panel shows the third-grade sample. Each circle reflects the average enrollment for all students with a specific birthdate relative to the December 1 cutoff. For example, the circles associated with −1 on the x-axis reflect enrollment patterns for children born on November 30; the circles at +1 reflect enrollment patterns for children born on December 2.

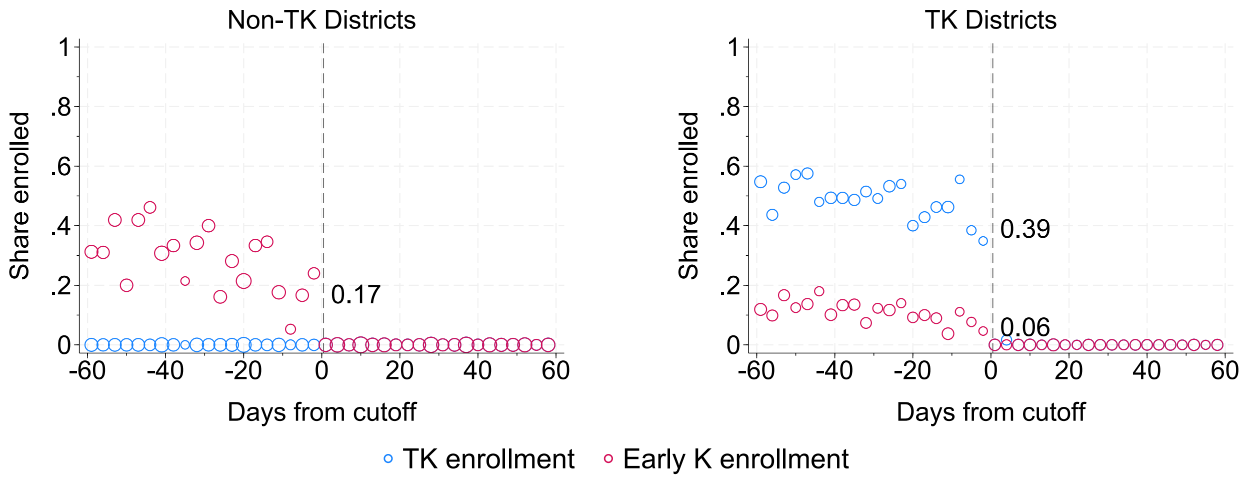
Looking at the third-grade sample in non-TK districts, we see that 35% of children born on December 1 (day zero on the x-axis) enrolled in kindergarten early. The share increases as one goes further from the cutoff in the negative direction, reflecting the fact that parents are more inclined to enroll their children in kindergarten if they are older. For example, among children born on October 1, which corresponds to 60 days before the cutoff or −60 on the x-axis, roughly 60% enrolled in EK. Immediately to the right of the cutoff, no children enrolled in EK, which is consistent with the state of Michigan not providing funding for students born on or after December 2 to enroll in public kindergarten programs.

Turning to districts that offer TK, we see that a sizable portion of children born to the left of the cutoff enrolled in TK and EK. Among children born on December 1, 17% enrolled in EK and 38% enrolled in TK. The remaining 45% were in a setting outside the public school system, which might have been a public preschool program, private preschool program, or informal care provided by family or other caregivers. As per Michigan funding regulations, no students born on or after December 2 enrolled in TK or EK.[5]

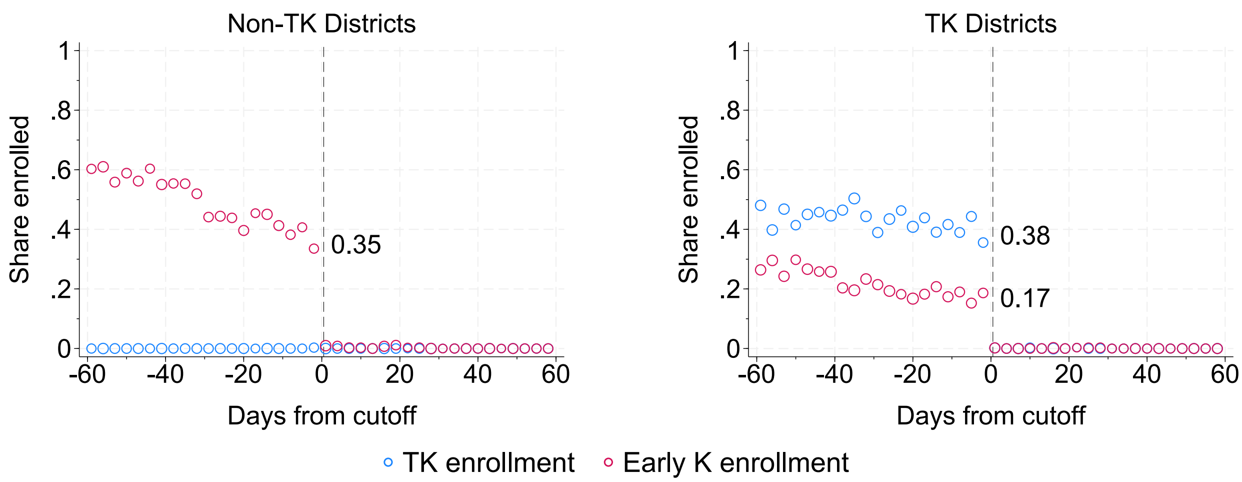
[5]Due to minor differences in outcome data missingness, the first stage estimates differ very slightly for each outcome in the third-grade sample. See Appendix Table A2 for the first-stage estimate and F-statistic associated with each outcome.

Figure 2. First-Stage Effects on TK and EK Enrollment in One's Pre-K Year

(a) KRA Sample



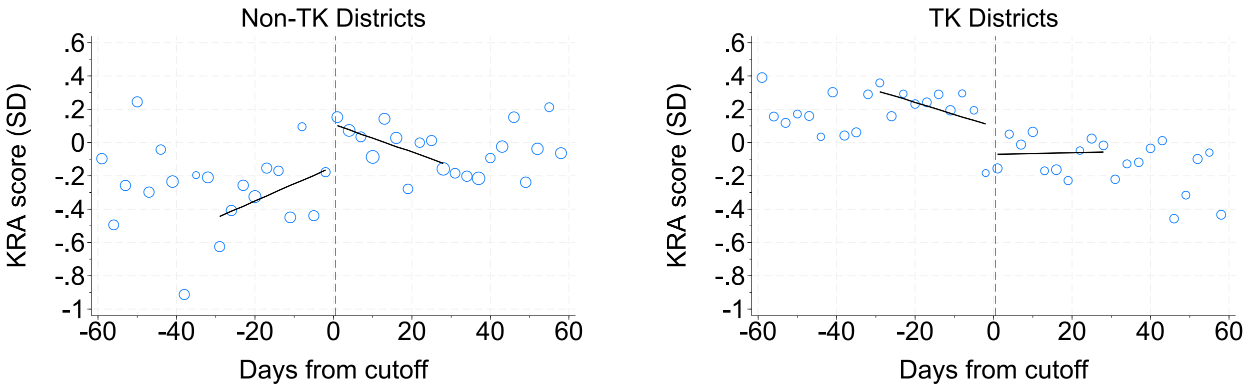
(b) 3rd Grade Sample



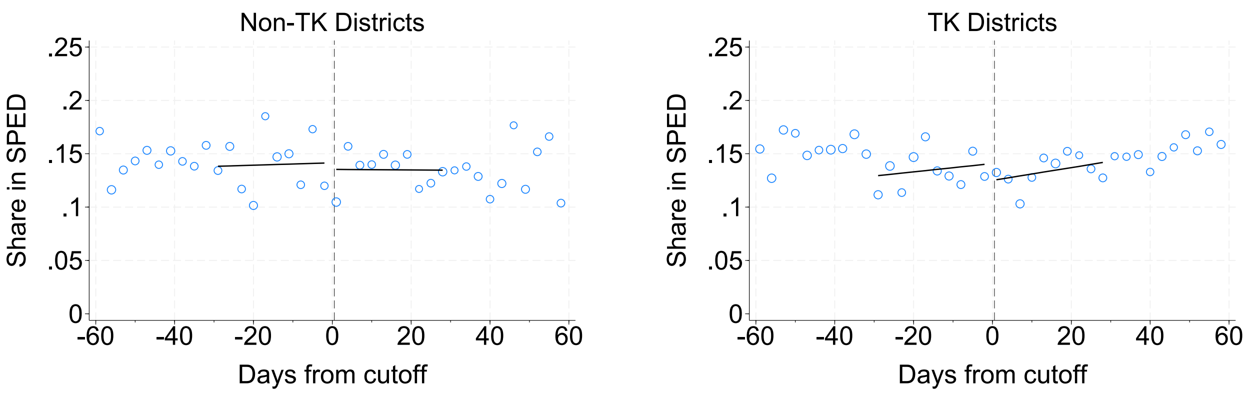
Note: N=1,943 for the KRA sample and N=15,680 for the 3rd grade sample. Each dot gives the average enrollment of children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). Panel (a) plots the first stage for TK and EK enrollment in districts in our KRA impact analysis. Panel (b) plots the first stages for districts in our 3rd grade impact analysis.

Figure 3. Intent-to-Treat Effects on KRA and SPED Outcomes

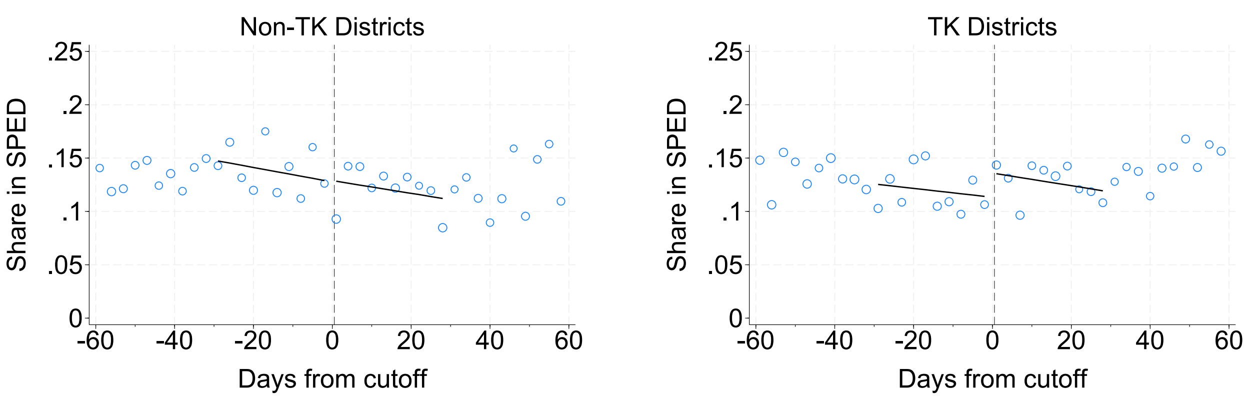
(a) Overall Kindergarten Readiness Scores



(b) Special Education in K, G1, or G2



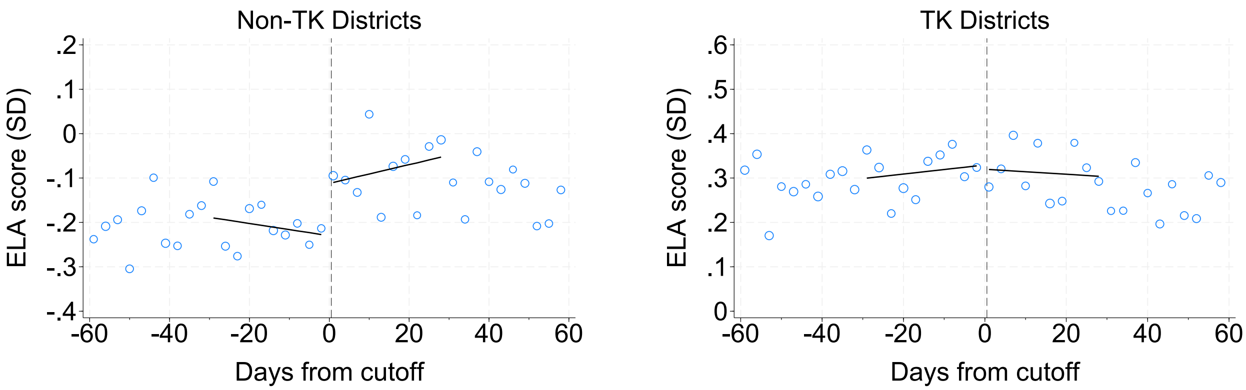
(c) Special Education in 3rd Grade



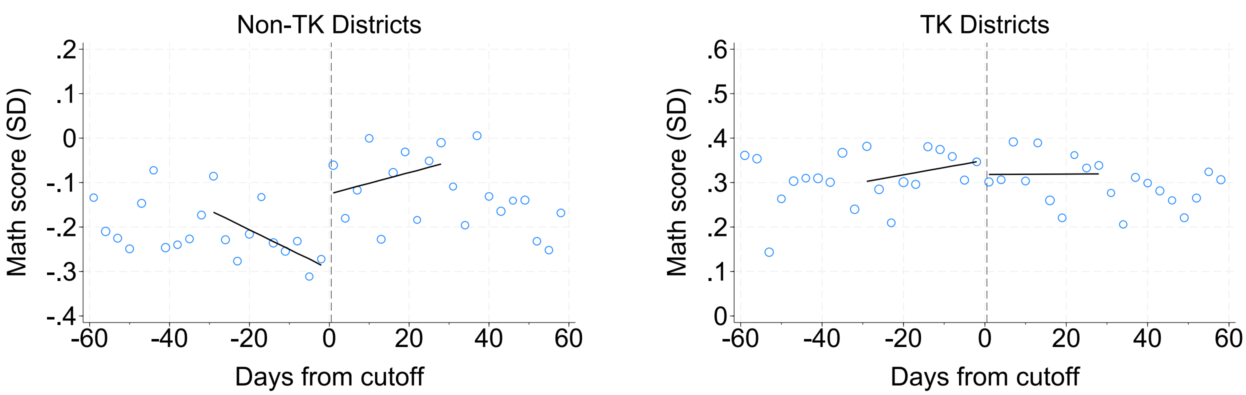
Note: N=1,943 in panel (a) and N=15,704 in panels (b) and (c). Each dot gives the average outcome among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). In panel (b), students are coded as being in special education in kindergarten, first grade, or second grade if they are observed as having an IEP in *any* of those grades.

Figure 4. Intent-to-Treat Effects on 3rd Grade Test Scores

(a) 3rd Grade ELA Scores



(b) 3rd Grade Math Scores



Note: N=15,669 in panel (a) and N=15,680 in panel (b). Each dot gives the average score on Michigan's statewide standardized test (the M-STEP) among children born in a three-day birthday range. Dot size is proportional to the number of students in a given birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). Third-grade test scores are measured the first time a student enrolls in third grade, regardless of whether they are "on schedule" based on their birthday and state guidelines regarding kindergarten entry.

Figures 3 and 4 present visual representations of our ITT effects on composite KRA scores, special education placement, and third-grade test scores. Here, we zero in on the plot for third-grade math scores in Figure 4 to provide intuition for our estimation of TK and EK treatment effects. In non-TK districts, we see a large negative discontinuity at the cutoff. Specifically, students born on December 2 score roughly 0.2 SD higher on the third-grade math exam than their peers born on December 1. In these districts, EK is the only option associated with the cutoff. This suggests students who waive into kindergarten early score substantially lower on third-grade math than their peers who do not waive in. This finding is not surprising, as EK students take state standardized tests one year earlier than they otherwise would and thus have experienced one fewer year of cognitive development (Deming and Dynarski, 2008; Ricks, 2024).[6] In contrast, in TK districts—where the cutoff is associated with both EK and TK—we find hardly any discontinuity in math scores at the cutoff. If EK has a negative effect in TK districts like it does in non-TK districts, this implies TK has an offsetting *positive* effect in TK districts. The 2SLS model we described above formally estimates these effects.

9.2.2 Main Treatment Effect Estimates

In Table 4, we provide results for the KRA that are consistent with the visual intuition in Figure 3. Attending TK led to a large 0.91 standard deviation (SD) boost in the KRA Overall score. Our findings are similar for the Social Foundations (0.93 SD) and Language and Literacy (0.92 SD) subscales. Impacts are somewhat smaller but still large for Math (0.64 SD) and Physical and Motor Development (0.54 SD).

Note that because of smaller sample sizes for the KRA, we are only able to fit models for the baseline approach. Further, as discussed earlier, districts range in the fraction of children assessed with the KRA. In Appendix F (Tables F10-F14), we provide impact estimates for the KRA across four different approaches to defining the sample. Across each sample and across various alternative bandwidths, the findings broadly align with our preferred KRA sample (at least 20 percent of kindergartners assessed).

In Table 5, we present the impacts of TK on special education. There is no statistically significant relationship between participation in TK and our preregistered confirmatory outcome, receiving special education services anytime in grades K-2. Our baseline approach shows a small positive effect of 2.8 percentage points ($p = 0.455$) and our relaxed assumptions approach shows an impact of -0.3 percentage points ($p = 0.970$).

[6] More precisely, EK causes children to reach third grade around 0.7 years earlier, on average, than they otherwise would. On the other hand, TK enrollment has no effect on grade trajectory. On average, TK students reach third grade at the same time as children who do not enroll in TK or EK (see Figure A5).

Table 4. Impacts of TK and EK on KRA Scores

	Overall	Social Foundations	Language and Literacy	Math	Physical/Motor Development
$LATE_{TK}$	0.914**	0.934**	0.915**	0.638*	0.544
[P-value]	[0.027]	[0.037]	[0.018]	[0.076]	[0.219]
$LATE_{EK}$	−1.841*	−2.127*	−2.079*	−0.848	−1.645
[P-value]	[0.079]	[0.072]	[0.067]	[0.299]	[0.173]
TK-complier control mean	−0.360	−0.674	−0.231	−0.108	−0.394
EK-complier control mean	1.628	1.585	1.979	0.924	1.130
Observations	1,943	1,943	1,943	1,943	1,943

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

To explore whether TK influenced student participation in special education over time, we estimate models for participation in each grade from TK through grade three. Note that nearly 10% of TK students in our analysis sample had an IEP while in TK. Looking at the kindergarten year, we see suggestive evidence that participation in TK increased special education participation. The baseline estimate is roughly 10 percentage points ($p = 0.000$); the estimate from the alternative specification is 6 percentage points, but not statistically significant ($p = 0.344$).[7] The effect shrinks in first grade and, by the end of second grade, the point estimates are close to zero and statistically insignificant in both estimation approaches. This pattern is broadly consistent with recent public Pre-K evaluations in Boston (Weiland et al., 2020) and DC (Braga et al., 2024). Interestingly, the third grade results suggest that participation in TK might *reduce* a child's use of special education services in the longer run. The point estimates from the baseline and alternative specifications suggest reductions of roughly 7 percentage points, which is extremely large given the control complier mean of 15.1%.[8] These estimates are not statistically significant at conventional levels, although the p-value for the baseline estimate is 0.106.

[7] The control mean estimate for TK compliers is slightly *negative* for the "special education in kindergarten" outcome. We note that this issue would be more pronounced across outcomes if we used the relaxed assumptions impact estimates to calculate complier control means. This may suggest an issue with our estimation procedure, although it could also be that our estimates are simply imprecise.

[8] We estimate complier control means for TK and EK students by subtracting (baseline approach) impact estimates from observed outcomes at the RD cutoff. We obtain observed outcomes at the RD cutoff by fitting a linear relationship between birthdays and outcomes for TK or EK students in districts with TK programs. To be consistent with Section 8, we limit the sample to children born 0-29 days before December 1. The predicted value at 0 days from the cutoff is our estimate of the observed outcome at the cutoff.

Table 5. Impacts of TK and EK on Special Education Placement

	SPED in K-G2	SPED in K	SPED in G1	SPED in G2	SPED in G3
	Panel A. Baseline Approach				
$LATE_{TK}$	0.028	0.097***	0.032	−0.007	−0.070
[P-value]	[0.455]	[0.000]	[0.270]	[0.856]	[0.106]
$LATE_{EK}$	0.016	0.009	0.029	0.001	−0.003
[P-value]	[0.714]	[0.786]	[0.415]	[0.978]	[0.941]
	Panel B. Relaxed Assumptions Approach				
$LATE_{TK}$	−0.003	0.059	0.009	−0.008	−0.066
[P-value]	[0.970]	[0.344]	[0.863]	[0.918]	[0.342]
$LATE_{EK}$	0.087	0.088	0.082	0.004	−0.010
[P-value]	[0.522]	[0.470]	[0.451]	[0.977]	[0.934]
TK-complier control mean	0.101	−0.001	0.064	0.109	0.151
EK-complier control mean	0.043	0.030	0.014	0.031	0.058
Observations	15,704	15,530	15,614	15,589	15,704

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. Complier control means are estimated by subtracting (baseline approach) impact estimates from observed outcomes at the RD cutoff. As a reference point, 9.6% of TK students at the birthday cutoff are in special education while in TK.

As we noted earlier, our third-grade standardized test score results appeared in Berne et al. (2024). We include them here alongside our other outcomes both for a more holistic understanding of TK's impacts and because we conduct more thorough sensitivity analyses that confirm our results (Appendix G). Table 6 presents the estimates of TK participation's impact on third-grade math and ELA scores. The patterns are consistent with the story conveyed by Figure 4—TK enrollment led to substantial improvements in 3rd grade math scores, but did not have a statistically significant effect on ELA scores.

Table 6. Impacts of TK and EK on 3rd Grade Test Scores

	Math		ELA	
	Panel A. Baseline Approach			
$LATE_{TK}$	0.252**	0.212*	0.123	0.097
[P-value]	[0.046]	[0.051]	[0.321]	[0.401]
$LATE_{EK}$	−0.378***	−0.366***	−0.240*	−0.219*
[P-value]	[0.000]	[0.000]	[0.061]	[0.078]
	Panel B. Relaxed Assumptions Approach			
$LATE_{TK}$	0.331*	0.294	0.209	0.191
[P-value]	[0.088]	[0.111]	[0.253]	[0.293]
$LATE_{EK}$	−0.557*	−0.557*	−0.435	−0.435
[P-value]	[0.092]	[0.092]	[0.181]	[0.181]
Controls		X		X
TK-complier control mean	0.219		0.310	
EK-complier control mean	0.609		0.447	
Observations	15,680		15,669	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. In the "relaxed assumptions" approach, we always exclude controls when estimating EK LATEs because the demographic subgroups are defined by the covariates. Hence, the EK estimates in the with- and without-controls columns are identical by construction. Complier control means are estimated by subtracting (baseline approach) impact estimates from observed outcomes at the RD cutoff.

Our baseline math estimate is 0.21 SD ($p = 0.051$) and our "relaxed assumptions" estimate is 0.29 SD ($p = 0.111$). These estimates differ because the demographic subgroups that dominate the EK compliers in TK districts (e.g., White and Asian girls regardless of economic disadvantage and White and Asian boys who are not economically disadvantaged) experience the largest negative effects from attending kindergarten early (see Appendix E). Unfortunately, this approach is data intensive, resulting in less precision. However, the fact that the point estimates *increase* compared to our baseline approach gives us confidence in interpreting the math impact as positive. For English Language Arts, as in Berne et al. (2024), our point estimates are not significantly different than zero but would suggest sizable effects (0.19 SD, $p = 0.293$).

Several points are worth noting when interpreting these estimates. First, the TK impacts are relative to a counterfactual of starting kindergarten "on schedule" or later, having spent one's Pre-K year receiving care at home, attending an income-targeted public Pre-K program, attending private Pre-K, or spending the year in some other type of arrangement. Second, and related, because they reflect the impact for students close to the December 1 cutoff, our estimates capture the effects of TK for children who are among the oldest in their grade cohort.

In Appendices F and G, we conduct a battery of robustness checks. We find our results are qualitatively similar across a wide range of specifications, including the use of alternative bandwidths and the inclusion/exclusion of covariates. We also find reassuring evidence from placebo checks at false birthday cutoffs. We have also tested our baseline approach 2SLS specification by estimating the components of Equation 1 (ITT, Ω_{TK}, and Ω_{EK}) separately and backing out $LATE_{TK}$; the estimates are nearly identical.

10 Extension Analysis: What is the Counterfactual?

Early learning programs tend to have larger impacts when the counterfactual is home-based care rather than center-based care (Feller et al., 2016; Kline and Walters, 2016). In our setting, children may have counterfactually received care from a parent or relative, attended a public or private preschool program, or received some other form of care. Our estimates capture the effects of TK relative to this mixed counterfactual.

Understanding the counterfactual is important for understanding mechanisms and for comparing impacts across programs. The Michigan administrative data lack the information needed to estimate childrens' counterfactual child care arrangements directly, but we can gain insight by pooling data from other sources. Table 7 presents enrollment estimates for children in the third-grade sample born just *after* the TK cutoff. To obtain these estimates, we combine information from the administrative student data, Head Start Program Information Reports, the Michigan Department of Licensing and Regulatory Affairs, and the American Community Survey. For details on our approach, see Appendix I.

The estimates in Table 7 should be seen as suggestive since they require strong assumptions and substantial imputation. Nevertheless, they provide useful context for understanding what TK and EK students would have experienced in the absence of these programs. We estimate that roughly half of all children born after the RD cutoff participated in some type of formal child care program, implying the other half were either at home with family or participated in some type of informal care. Among those in formal care, a plurality (23.1%) were in GSRP, Michigan's state-funded Pre-K program; 10.4% were in Head Start; and 15.6% were in other licensed child care. Note that the estimates in Table 7 represent the counterfactual for *all* children born by December 1. Given that families of EK and TK students had, by definition, chosen to enroll their children in a formal care arrangement, we

Table 7. Estimated Program Enrollment for Children Born Just After the TK Birthday Cutoff

GSRP	Head Start	Other Licensed Child Care	Residual Care Arrangements
23.1%	10.4%	15.6%	50.9%

Note: This table presents program enrollment estimates for children born just after the TK cutoff in districts that offered TK in 2014-2015 and 2018-2019 and that reliably reported TK enrollment. The estimates combine population and enrollment data from a variety of sources, requiring substantial imputation. For more information on our estimation approach, see Appendix I. GSRP stands for the Great Start Readiness Program, which is Michigan's income-targeted state-funded Pre-K program.

suspect that these children would have been more likely than average to enroll in some other type of formal option in the absence of the EK and TK programs. Taken as a whole, we view this evidence as suggesting that our estimates reflect the difference between participating in TK and, for many students, participation in some other type of formal care.

11 Discussion

Research commonly finds that attending preschool versus staying home with a family member has sizable impacts on kindergarten readiness, followed by partial or complete convergence in the early elementary years, and then re-emergence of benefits in early adulthood (Phillips et al., 2017). However, TK models are new to the early education landscape and the degree to which their findings will replicate Pre-K findings is an open question. Our Michigan TK findings underscore both the distinctive features of TK programs and the need for rigorous longitudinal studies of their effects on children's learning and development.

Despite wide latitude in program design, we find remarkable consistency in TK structural program features and high levels of administrator satisfaction with TK programs. These consistent features include class size (for which the state of Michigan has no cap in TK or K-12), program dosage, and standalone TK classroom structures. There is more variation in curriculum and assessment, with programs reporting more use of domain-specific choices relative to Michigan's state-funded Pre-K (Weiland et al., 2023) and publicly funded preschool programs nationally (Friedman et al., 2024; Jenkins et al., 2018; Shapiro et al., 2025).

Overall, as in the limited prior work on California and Washington TK (Goodvin et al., 2023; Mercado-Garcia et al., 2014), Michigan TK appears to blend features of Pre-K and kindergarten programs. Some of these features align with research on the optimal ingredients of early learning programs while others do not. For example, randomized trials have repeatedly shown that domain-specific curricula generate larger achievement effects than the global curricula more typically used in public Pre-K (Jenkins et al., 2018; National Academies of

Science and Medicine, 2024). But, only 69% of TK programs had a full-time or part-time teaching assistant or paraprofessional, meaning that in nearly a third of classrooms, ratios exceed the typical teacher-child recommendation of 1:10 (Friedman et al., 2024). State-funded Pre-K programs typically meet this guideline and in Michigan Pre-K at the time of our study, the ratio was 1:8 (Wu et al., 2023). Although research on ratios in early education classrooms is largely correlational and has mixed findings (Bowne et al., 2017; Perlman et al., 2017), the larger ratios in TK could impede small group instruction and overall classroom management.

For impacts on children's kindergarten readiness, we find a large, statistically significant overall impact (0.91 SD). Impacts are large for social skills and language and literacy and moderately large for math and physical/motor development. These effects are on the higher end for Pre-K studies. For example, the Boston Pre-K impact was 0.62 SD for literacy and 0.50-0.59 SD across assessments for math; the Tulsa Pre-K impact was 0.80 SD for literacy and 0.38 SD for math (Weiland and Yoshikawa, 2013; Gormley Jr. et al., 2005). The only TK comparison point is California's program, where kindergarten readiness impacts were 0.48 SD in literacy, 0.15 SD in vocabulary, 0.20-0.29 SD across math assessments, and 0.18 SD for engagement, with no effects on cooperation, self-control, or externalizing/internalizing behaviors (Manship et al., 2017).

For special education placement, we find an interesting pattern, although the estimates are quite imprecise and not statistically significant. Children who participated in TK were 6-10 percentage points more likely to be identified for special education by the end of kindergarten. By the end of first grade, however, the difference had shrunk to close to zero. Perhaps most intriguing, by the end of third grade, we find TK students were substantially *less* likely to participate in special education relative to comparison students (though again this difference is not statistically significant). In our view, these results suggest TK influenced special education participation in two distinct ways—namely, it shifted forward the timing of identification for some students and also reduced the likelihood that students needed special education services later on. These findings diverge somewhat from several recent Pre-K studies (Braga et al., 2024; Durkin et al., 2022; Weiland et al., 2020) and from a California TK study (Lafortune and Hill, 2023). These studies tend to show the same uptick in the rate of special education in the early elementary grades we detect, but they do not show the control group then surpassing the treatment group later in elementary school.

For third grade achievement, we find large, positive effects for math. ELA impacts are positive and sizable in magnitude but not statistically significant. These findings are robust to a wide variety of sensitivity analyses. The magnitude of our estimates is large relative to the prior literature. Across all relatively rigorous evaluations of programs since the 1960s, the average impact of preschool on children's *end-of-preschool* cognitive skills is about 0.25 standard deviations (Duncan and Magnuson, 2012). Our impact estimates are the same size for students in third grade. To give another reference point, our math estimate amounts to 61% of expected cognitive development between 3rd and 4th grade.[9]

Ultimately, the mechanisms behind our findings are unclear and are a topic for future research, especially given different theoretical perspectives on the reasons why early education benefits may or may not last (Bailey et al., 2017; Heckman, 2000; Masten, 2010; Sameroff, 2009). We want to underscore that the counterfactual in our setting may have been weaker than in some of the recent Pre-K studies and in California TK. As we outlined, about 50% of Michigan children born after December 1 did not attend an early learning program in the year before kindergarten. Nearly all of the Boston Pre-K comparison group and 80% of the California TK comparison group enrolled in other programs.

Notably too, previous rigorous research on Pre-K programs has shown, for example, that curriculum alignment and overall school quality appears to sustain the Pre-K boost (Clements et al., 2024; Mattera et al., 2021; Unterman and Weiland, 2024). Similar studies are needed on TK programs, as are exploratory studies of the relationship between TK program ingredients and gains in children's outcomes, following an extensive literature in Pre-K programs (Weiland and Rosada, 2025).

Our study has important limitations. Our measures after kindergarten entry are more limited than we would have liked, with no measures of other important child developmental domains such as social skills or executive function. We lacked power to examine student subgroup effects, a topic we hope to take up in future work given evidence that effects of early childhood programs are often larger for children from families with low incomes and for English learners (Phillips et al., 2017). Our survey was collected after our focal cohorts had completed TK and thus may not fully reflect their specific experiences in TK. Further, survey responses represent only a subset of districts that had TK programs. The TK districts in our survey appear largely representative of TK districts not in our survey, but the charter schools with TK in our survey tend to serve more advantaged populations than the charter schools with TK not in our survey (Table A1). In addition, the Kindergarten Readiness Assessment (KRA) has been well validated in diverse samples of young children, but to our knowledge, this is the first time it has been used in an impact study of an early learning program. Finally, the generalizability of our impact findings is limited to TK-offering districts in Michigan, which serve more advantaged children than non-TK-offering districts (Berne et al., 2025).

[9] This is based on the average growth of Michigan students from grade 3 in 2017-18 to grade 4 in 2018-19, which is nearly identical to estimates found in other literature (Hill et al., 2008).

As states and localities continue to expand early learning programs, our findings on Michigan TK's features and TK's effects on children's learning make novel contributions that inform the development and scaling of ECE models.

References

Atteberry, A., Bassok, D., and Wong, V. (2019). The effects of full-day prekindergarten: Experimental evidence of impacts on children's school readiness. *Educational Evaluation and Policy Analysis*, 41(4):537–562.

Bailey, D., Duncan, G. J., Odgers, C. L., and Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10:7–39.

Bailey, M., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Barr, A. and Gibbs, C. (2022). Breaking the cycle? Intergenerational effecs of an antipoverty program in early childhood. *Journal of Political Economy*, 130(12):3253–3285.

Berne, J., Garcia, K. C., Jacob, B. A., Musaddiq, T., Owusu, S., Shapiro, A., and Weiland, C. (2025). Transitional Kindergarten: The new kid on the early learning block. *Educational Researcher*, 54(4):177–187.

Berne, J., Jacob, B., Musaddiq, T., Shapiro, A., and Weiland, C. (2024). The effect of early childhood programs on third-grade test scores: Evidence from Transitional Kindergarten in Michigan. *AEA Papers and Proceedings*, 114:480–485.

Bloom, H. S. (2022). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5(1):43–82.

Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., and Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 29:407—-428.

Braga, B., Doromal, J. B., Greenberg, E., Monarrez, T., Restrepo, L., and Lamb, R. (2024). The effects of public pre-k for 3-year-olds on early elementary school outcomes: Evidence from the DC centralized lottery. Technical report, Annenberg Working Papers, Providence, RI.

Caetano, C., Caetano, G., and Escanciano, J. C. (2023). Regression discontinuity design with multivalued treatments. *Journal of Applied Econometrics*, 38(6):840–856.

Cascio, E. U. (2021). Early childhood education in the United States. In McCall, B. P., editor, *The Routledge Handbook of the Economics of Education Routledge*, chapter 2, pages 30–72. Routledge, London.

Cascio, E. U. (2023). Does universal preschool hit the target?: Program access and preschool impact. *Journal of Human Resources*, 58(1):1–42.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Chaudry, A. and Dutta, A. R. (2017). The current landscape for public pre-kindergarten programs. Technical report, Brookings Institute, Washington, DC.

Clements, D. H., Sarama, J., Wolfe, C. B., and Spitler, M. E. (2024). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal*, 50(4):812–850.

Data Recognition Corporation (2018). Psychometric analysis report for the Michigan English Language Arts (ELA), mathematics, science, and SAT student growth percentile and adequate growth percentile reporting.

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

Deming, D. and Dynarski, S. (2008). The lengthening of childhood. *Journal of Economic Perspectives*, 22(3):71–92.

Dodge, K. A., Bai, Y., Ladd, H. F., and Muschkin, C. G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development*, 88(3):995–1014.

Doss, C. (2019). How much regulation? A fuzzy regression discontinuity analysis of student literacy skills in prekindergarten vs. Transitional Kindergarten. *Education Finance and Policy*, 14(2):178–209.

Duncan, G. and Magnuson, K. (2012). Investing in preschool programs. *Journal of Economic Perspectives*, 27(2):109–132.

Durkin, K., Lipsey, M. W., Farran, D. C., and Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, 58(3).

Feller, A., Grindal, T., Miratrix, L., and Page, L. C. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care type. *The Annals of Applied Science*, 10(3):1245–1285.

Friedman, A. H., Barnett, W. S., Hodges, K. S., Garver, K. A., Jost, T. M., Weisenfeld, G., and Duer, J. (2024). The state of preschool 2023: State preschool yearbook. Technical report, National Institute for Early Education Research, New Brunswick, NJ.

Goodvin, R., Gibson, C., Rashid, A., Miller, M., and Hoagland, C. (2023). Transitional Kindergarten programs in Washington state: Describing 2022-23 programs, educators, and students (document number 23-12-2201). Technical report, Washington State Institute for Public Policy, Olympia, WA.

Gormley, W., Phillips, D., and Anderson, S. (2018). The effects of Tulsa's pre-k program on middle school student performance. *Journal of Policy Analysis and Management*, 37(1):63–87.

Gormley Jr., W. T., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, 41.

Gray-Lobe, G., Pathak, P. A., and Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, 138(1):363–411.

Heckman, J. (2000). Policies to foster human capital. *Research in Economics*, 54:3–56.

Hill, C., Bloom, H., Black, A., and Lipsey, M. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3):172–177.

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., and Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review*, 65:107–125.

Johnson, R. C. (2024). Synergistic impacts of expansions in pre-k access and school funding on student achievement: Evidence from California's Transitional Kindergarten rollout. *AEA Papers and Proceedings*, 114:467–473.

Justice, L., Koury, A., and Logan, J. (2019). Ohio's kindergarten readiness assessment: Does it forecast third-grade reading success? Technical report, Crane Center for Early Childhood Research and Policy.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Lafortune, J. and Hill, L. (2023). Assessing Transitional Kindergarten's impact on elementary school trajectories. Technical report, Public Policy Institute of California, San Francisco, CA.

Manship, K., Holod, A., Quick, H., Ogut, B., de los Reyes, I. B., Anthony, J., and Chernoff, J. J. (2017). The impact of Transitional Kindergarten on California students: Final report from the study of California's Transitional Kindergarten program. https://files.eric.ed.gov/fulltext/ED609085.pdf.

Manship, K., Quick, H., Holod, A., Mills, N., Ogut, B., Chernoff, J. J., Blum, J., Hauser, A., Anthony, J., and Gonzalez, R. (2015). Impact of California's Transitional Kindergarten program, 2013–14. Technical report, American Institutes for Research, San Mateo, CA.

Markowitz, A. and Bassok, D. (2025). ECE workforce. In Harris, D., editor, *Live Handbook of Education Policy Research*. Association for Education Finance and Policy.

Maryland State Department of Education (2019). The 2018-2019 kindergarten readiness assessment technical report.

Masten, A. S. (2010). Developmental cascades. *Development and Psychopathology*, 22:491–495.

Mattera, S. K., Jacob, R., MacDowell, C., and Morris, P. A. (2021). Long-term effects of enhanced early childhood math instruction: The impacts of Making Pre-K Count and High 5s on third-grade outcomes. Technical report, MDRC.

McCormick, M., Pralica, M., Weiland, C., Hsueh, J., Moffett, L., Guerrero-Rosada, P., Weissman, A., Zhang, K., Maier, M., Snow, C., and Davies, E. (2022). Does kindergarten instruction matter for sustaining the prekindergarten (prek) boost? Evidence from individual-and classroom-level survey and observational data. *Developmental Psychology*, 58.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., Yang, R., Koepp, A., and Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher*, 46(8):474–487.

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

Mercado-Garcia, D., Manship, K., Quick, H. E., Miller, M., and Holod, A. (2014). Transitional Kindergarten programs in California: Comparing Transitional Kindergarten and kindergarten classrooms). https://www.air.org/sites/default/files/downloads/report/Transitional%20Kindergarten%20vs%20Kindergarten_Oct%202014%20brief.pdf.

National Academies of Science and Medicine (2024). A new vision for high quality preschool curriculum. Technical report, National Academies Press, Washington, DC.

Office of Innovative Projects (2018). Determination of eligibility for Michigan mandatory special education for infants and toddlers, birth to age three early childhood developmental delay guidance. https://eotta.ccresa.org/Files/Uploads/New/188/ECDD_MMSE_Guidance.pdf.

Pages, R., Lukes, D. J., Bailey, D. H., and Duncan, G. J. (2020). Elusive longer-run impacts of Head Start: Replications within and across cohorts. *Educational Evaluation and Policy Analysis*, 42(4):471–492.

Perlman, M., Fletcher, B., Falenchuk, O., Brunsek, A., McMullen, E., and Shah, P. S. (2017). Child–staff ratios in early childhood education and care settings and child outcomes: A systematic review and meta-analysis. *PLOS One*, 12.

Phillips, D., Gormley, W., and Anderson, S. (2016). The effects of Tulsa's CAP Head Start program on middle-school academic outcomes and progress. *Developmental Psychology*, 52(8).

Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., Duncan, G., Dynarski, M., Magnuson, K., and Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects - a consensus statement. Technical report, The Brookings Institution, Washington, DC.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., and Downer, J. (2012). Third grade follow-up to the Head Start impact study: Final report. Technical Report OPRE Report 2012-45, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Ricks, M. D. (2024). Strategic selection around policy recommendations: The case of kindergarten entry. Working Paper.

Sameroff, A. (2009). *The transactional model of development: How children and contexts shape each other*. American Psychological Association.

Shapiro, A., Berne, J., Garcia, K. C., Jacob, B., Musaddiq, T., Owusu, S., and Weiland, C. (2023). Michigan Transitional Kindergarten: A first look at program reach and features. Technical report, Education Policy Initiative, Ann Arbor, MI.

Shapiro, A., Steiner, E. D., Woo, A., Cannon, J. S., Doss, C. J., Karoly, L. A., and Kassan, E. B. (2025). Instructional resources in public school–based pre-k. https://www.rand.org/pubs/research_reports/RRA3279-3.html.

Unterman, R. and Weiland, C. (2024). Quantifying and predicting variation in the medium-term effects of an oversubscribed prekindergarten program. *Journal of Research on Educational Effectiveness*, pages 1–28.

Weiland, C., Chaudry, A., Shapiro, A., Berne, J., Hyland, K., Hamp, N., and Taylor, A. (2023). An evidence-based path to expanding high-quality pre-k in Michigan. Technical report, Education Policy Initiative, Ann Arbor, MI.

Weiland, C., McCormick, M., Duer, J., Friedman-Krauss, A., Pralica, M., Xia, S., Nores, M., and Mattera, S. (2024). The mixed-delivery pre-k opportunity gap? Differences in demographics, quality, and children's gains in community-based versus public school programs across five large-scale systems. *Early Childhood Research Quarterly*, 68:247–259.

Weiland, C. and Rosada, P. G. (2025). Quality in ECE. In Harris, D., editor, *Live Handbook of Education Policy Research*. Association for Education Finance and Policy.

Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., and Martin, E. (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. *Child Development*, 91(5):1401–1422.

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

WestEd (2015). Ready for kindergarten: Kindergarten readiness assessment technical report addendum.

What Works Clearinghouse (2022). Procedures and standards handbook, version 5.0. Technical report, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Wu, J. H., Herbowicz, T., Miller, S. R., Egeren, L. A. V., and Akaeze, H. O. (2023). Great Start Readiness Program state evaluation 2021-22 annual report. https://cep.msu.edu/upload/gsrp/GSRP%20Annual%20Report%202021-22.pdf.

Yoshikawa, H., Weiland, C., and Brooks-Gunn, J. (2016). When does preschool matter? *Future of Children*, pages 21–35.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., and Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. Technical report, Society for Research in Child Development, Ann Arbor, MI.

Appendix A Supplemental Tables and Figures

Table A1. Characteristics of Districts and Charters with TK in SY 2021-22 for Which We Do and Do Not Have Survey Responses

	Districts		Charter Schools	
	No Survey Response	Have Survey Response	No Survey Response	Have Survey Response
	Panel A. Student Characteristics			
White	75%	69%	36%	45%
Black	8%	12%	40%	25%
Hispanic	7%	9%	14%	13%
Asian	4%	4%	2%	8%
Other race	6%	7%	8%	10%
English learner	8%	6%	11%	12%
Economically disadvantaged	48%	52%	73%	60%
Special education	16%	16%	14%	13%
Students (N)	192,904	160,035	10,002	9,032
	Panel B. District Characteristics			
Average regular K enrollment	244.9	227.1	78.9	96.9
K students who attended state-funded pre-K	29%	29%	29%	25%
City	8%	5%	28%	21%
Suburb	35%	41%	38%	63%
Town	22%	20%	3%	8%
Rural	35%	34%	31%	8%
Average 3rd grade math score	0.00 SD	0.01 SD	−0.30	0.01
	(0.36)	(0.36)	(0.39)	(0.47)
Average 3rd grade ELA score	0.01 SD	0.01 SD	−0.21	−0.01
	(0.33)	(0.32)	(0.40)	(0.42)
Districts (N)	160	147	33	24

Note: Figures in this table were estimated using administrative data from school year 2021-22. Analysis is restricted to districts designated as having TK in 2021-22 based on administrative records and primary data collection. The statistics in Panel A and the urbanicity shares in Panel B are tabulated using data on students in grades 1-5. "State-funded Pre-K" refers to students who enrolled in Michigan's Great Start Readiness Program (GSRP) or a GSRP/Head Start blend program. 3rd grade test scores are scores on the Michigan Student Test of Educational Progress (M-STEP). The numbers in parentheses below the mean test scores are standard deviations.

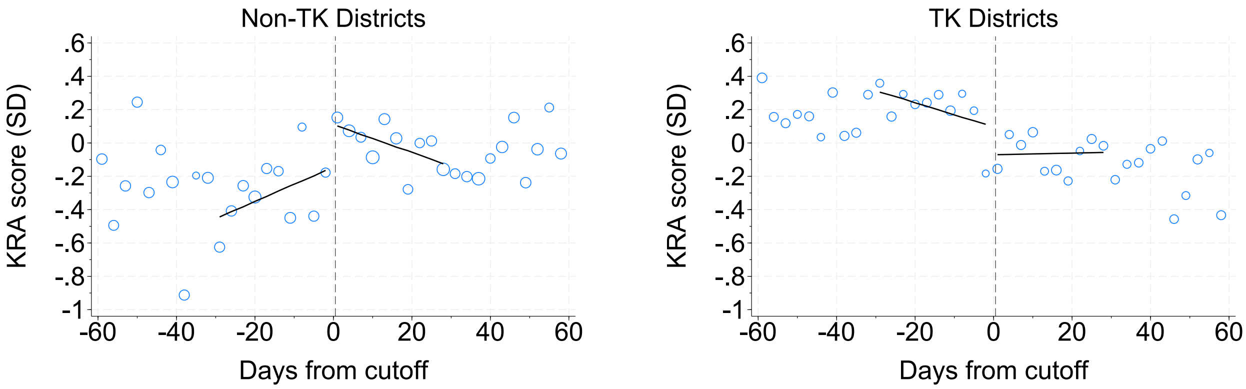
Table A2. First Stage Effect for Each Second Stage Outcome

	TK Districts				Non-TK Districts	
	EK		TK		EK	
	Point Estimate	F-stat.	Point Estimate	F-stat.	Point Estimate	F-stat.
KRA scores	0.065	8.4	0.388	68.1	0.168	11.2
SPED in K	0.176	288.9	0.360	441.9	0.349	347.8
SPED in G1	0.172	289.7	0.381	516.7	0.349	342.1
SPED in G2	0.170	300.3	0.382	563.0	0.345	347.3
SPED in G3	0.171	295.1	0.381	533.3	0.349	347.7
SPED in K-G2	0.171	295.1	0.381	533.3	0.349	347.7
SPED in K-G3	0.171	295.1	0.381	533.3	0.349	347.7
3rd Grade Math score	0.171	294.2	0.381	539.3	0.349	341.9
3rd Grade ELA score	0.169	279.8	0.382	536.0	0.348	345.0

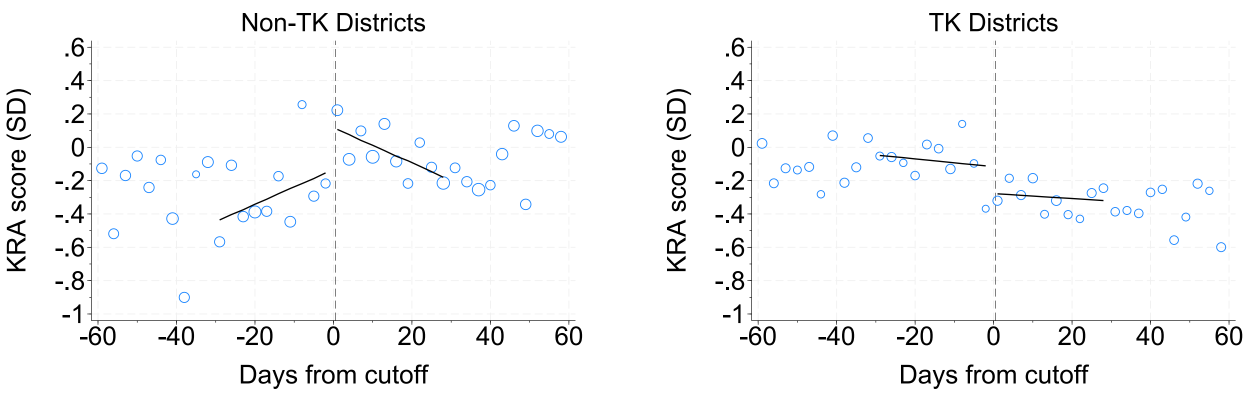
Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. F-statistics are computed analytically, with clustering on the running variable. We do not present separate estimates for the KRA overall score and subscores because the samples (and thus first stages) are identical. The slight differences in the first stages between various special education and 3rd grade test score outcomes reflect minor sample differences stemming from missing outcome data.

Figure A1. Intent-to-Treat Effects on Kindergarten Readiness Scores I

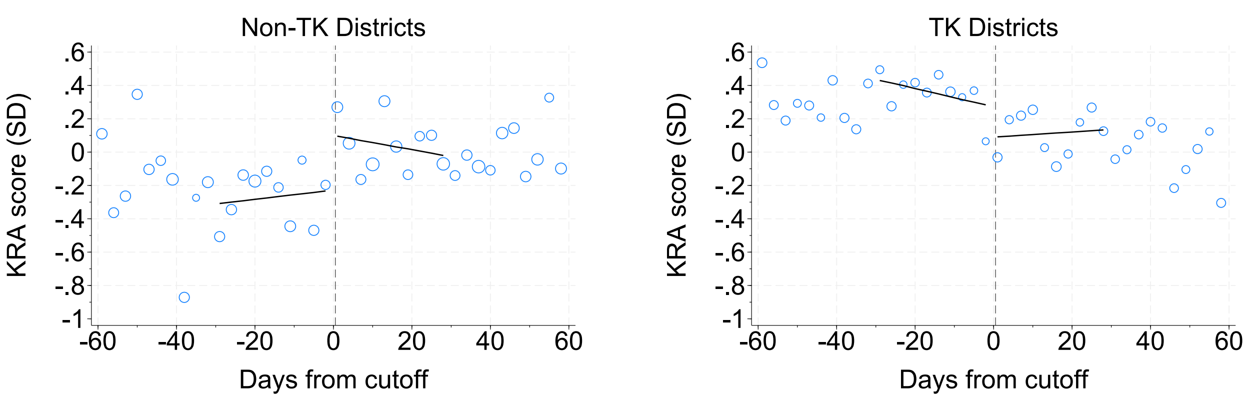
(a) Overall



(b) Social Foundations



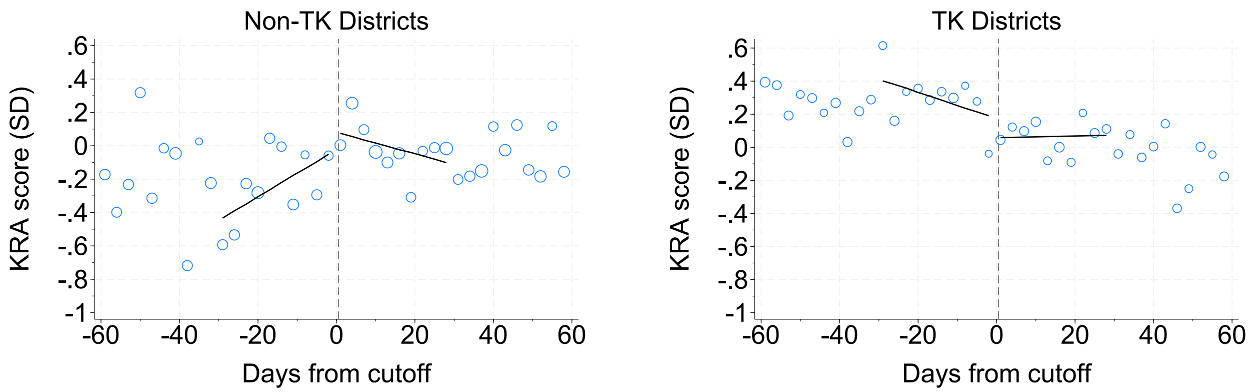
(c) Language and Literacy



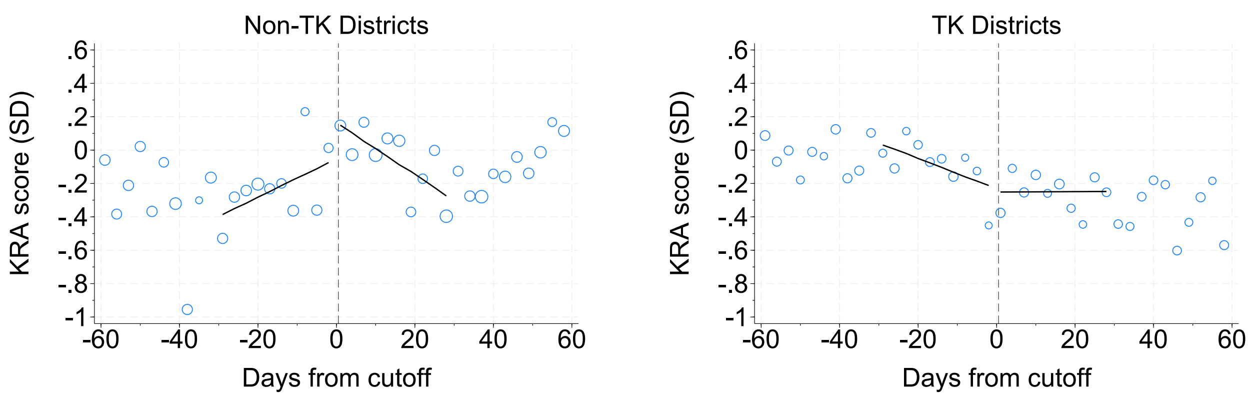
Note: Each dot gives the average Kindergarten Readiness Assessment (KRA) score among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). Panel (a) in this figure is a reprint of panel (a) in Figure 3.

Figure A2. Intent-to-Treat Effects on Kindergarten Readiness Scores II

(a) Mathematics



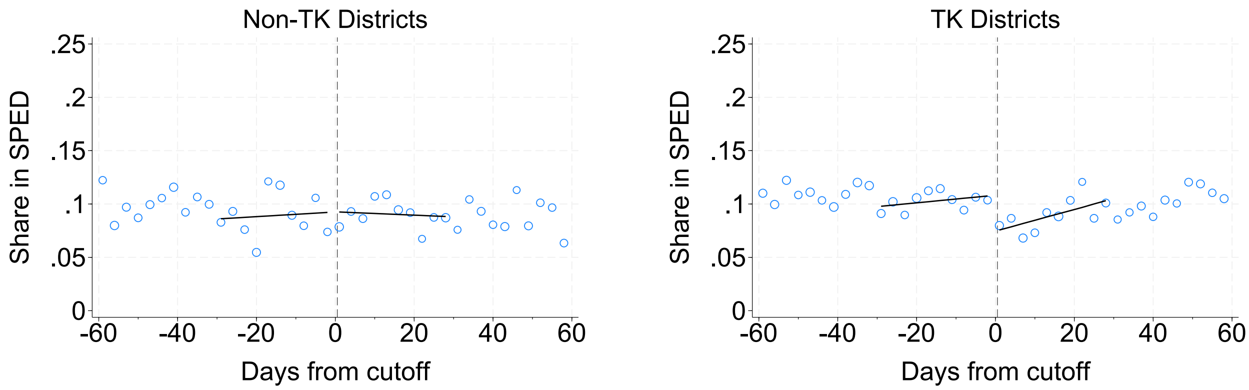
(b) Physical and Motor Development



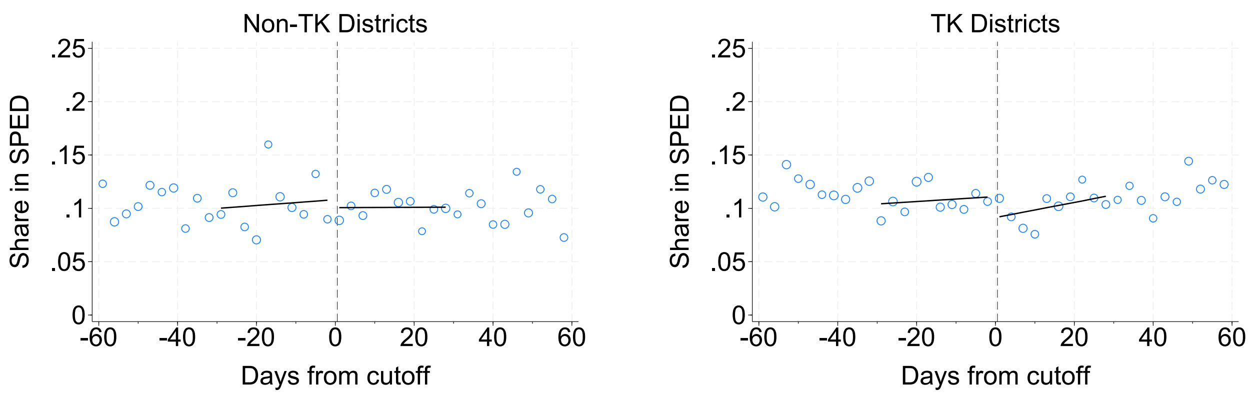
Note: Each dot gives the average Kindergarten Readiness Assessment (KRA) score among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1).

Figure A3. Intent-to-Treat Effects on Special Education Outcomes I

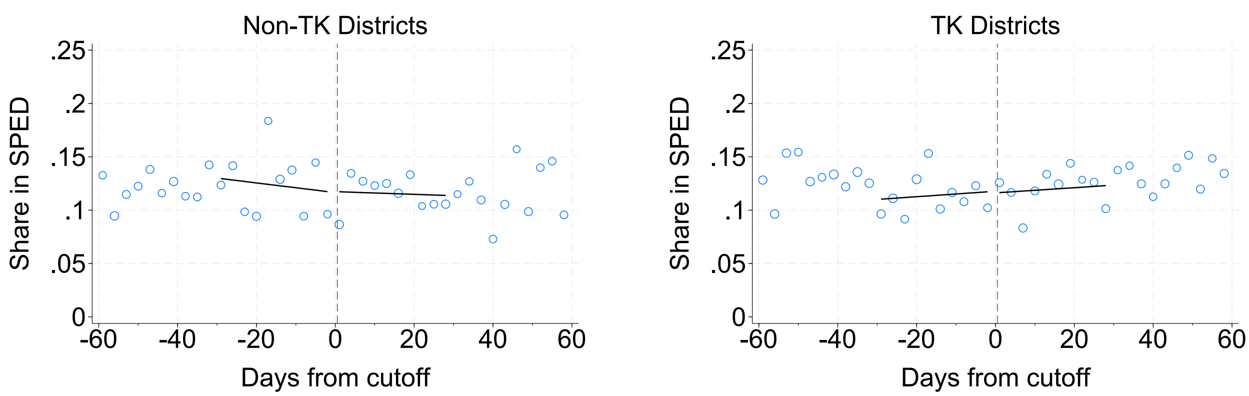
(a) Special Education in Kindergarten



(b) Special Education in 1st Grade



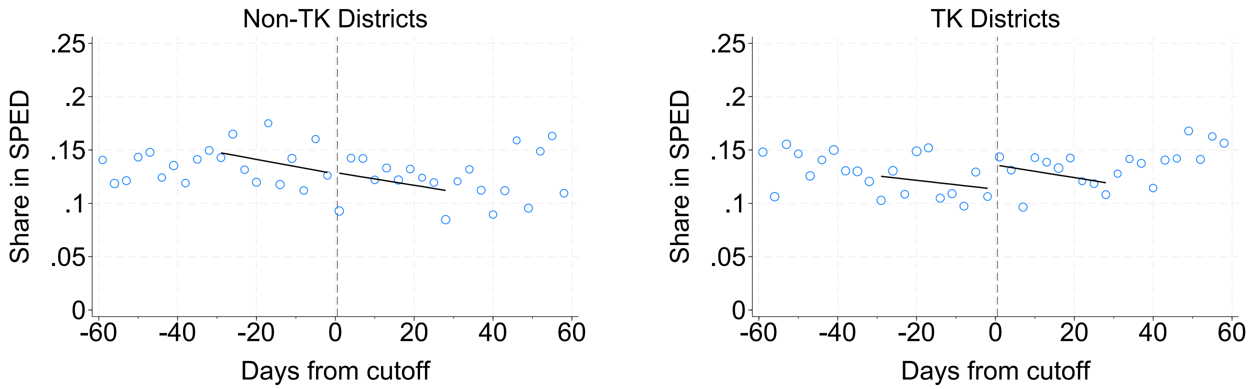
(c) Special Education in 2nd Grade



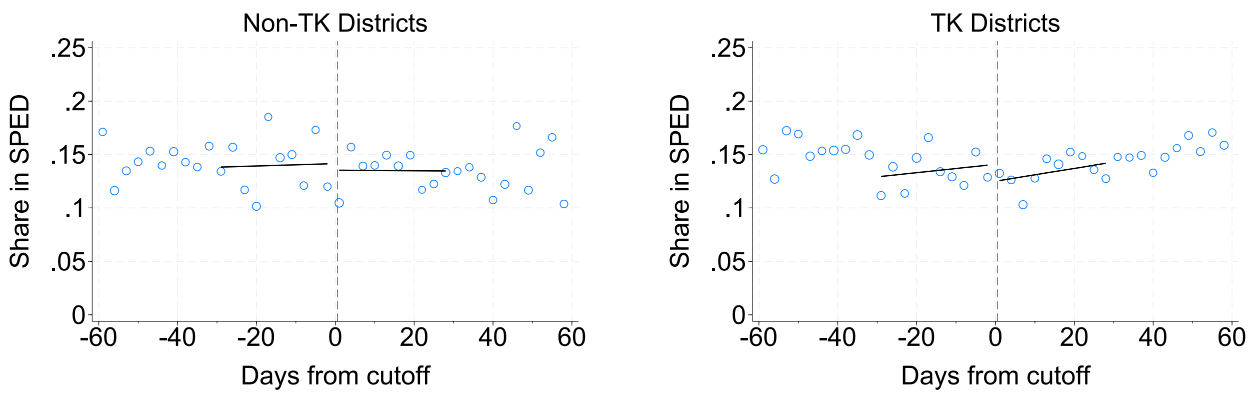
Note: Each dot gives the share of students receiving special education services among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1).

Figure A4. Intent-to-Treat Effects on Special Education Outcomes II

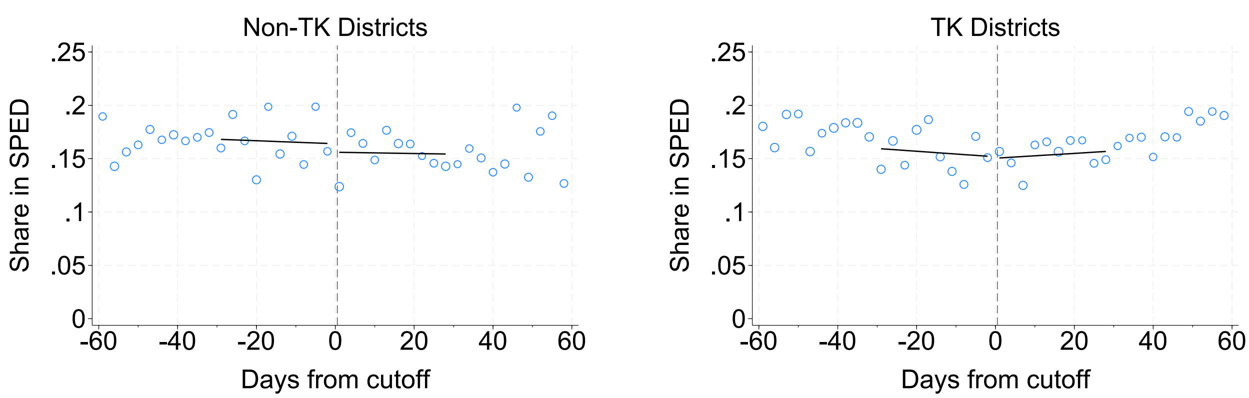
(a) Special Education in 3rd Grade



(b) Special Education in K, G1, or G2

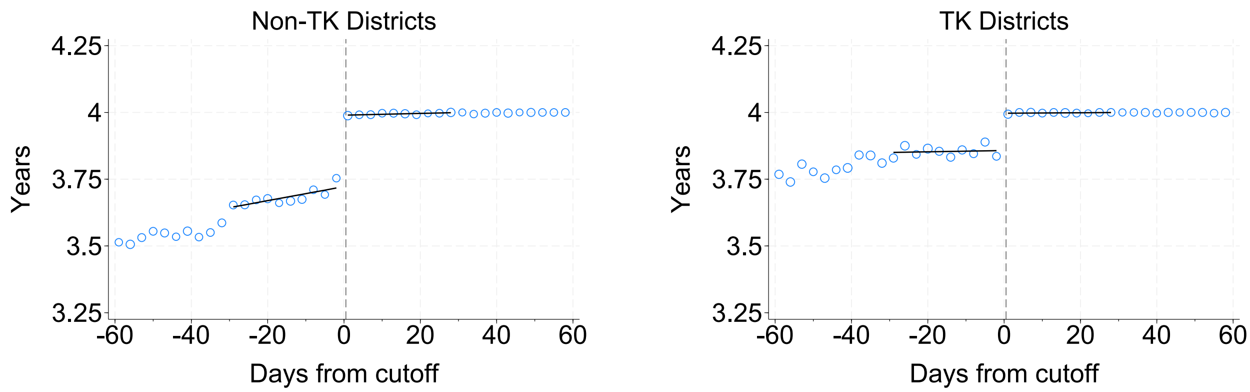


(c) Special Education in K, G1, G2, or G3



Note: Each dot gives the share of students receiving special education services among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). Panel (a) in this figure is a reprint of panel (c) in Figure 3. In panels (b) and (c), students are coded as being in special education if they are observed as having an IEP in *any* of the listed grades.

Figure A5. Effects on Number of Years Between One's Pre-K Year and 3rd Grade



	Complier Control Mean	Baseline	Relaxed Assumptions
$LATE_{TK}$	4.019	−0.036**	−0.031
[P-value]		[0.019]	[0.447]
$LATE_{EK}$	3.944	−0.731***	−0.743**
[P-value]		[0.000]	[0.022]
Observations		15,680	15,680

*** p<0.01, ** p<0.05, * p<0.1

Note: In the plots, each dot gives the average number of years that pass between one's pre-K year and the first time they are in third grade, among children born in a three-day birthday range. Dot size is proportional to the number of students in a birthday range. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1). In the table, estimates are obtained via RD models as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. Complier control means are estimated by subtracting (baseline approach) impact estimates from observed outcomes at the RD cutoff.

Appendix B Sample and Variable Construction

B.1 Identifying TK and Non-TK District×Cohorts

In the longitudinal student administrative data, all kindergartners and TK students are marked as being in grade 0. A separate variable is used to indicate if the student is enrolled in TK. However, because TK students are funded as regular kindergartners, districts were advised but not required to report on the separate TK flag during the time of our study. As a result, in many districts we cannot tell whether a student in grade 0 is in TK or regular kindergarten. This complicates identification of TK offering districts and charter schools in a particular year.

To address concerns of measurement error in TK enrollment, we restrict our sample to districts and cohorts where we we feel confident about individual-level reporting on TK enrollment. When a district reports 10 or more TK students in the data in a given year, we are reasonably confident the district offered TK that year for two reasons. First, districts are unlikely to report 10 or more students as being in TK erroneously. Second, as explained below, we check for grade progression patterns in all districts that report TK students. The patterns match our expectations in districts that meet our data reliability standards, but not in districts that don't. We acknowledge the trade-off in using this criterion of reporting 10 or more TK students in a year; we may be dropping districts with particularly small TK programs, but we are increasing the likelihood that districts included as offering TK are categorized accurately.

Throughout the paper, when we refer to TK districts, we are referring to district×years that report 10+ TK students. It is possible that a given district may report 10+ TK students in one year but not another year. TK districts are identified at the district×year level in our sample.

We are also reasonably confident that a certain set of districts never offered TK. We use two data sources and conditions to define districts that never offer TK. First, a district must not have reported a single TK student in the student administrative data in any year. Second, we rely on an extensive data triangulation effort our team undertook in spring 2022, where we reviewed district websites and communicated with district staff via email and phone calls to make a determination for every district about whether they offered TK in that school year. Based on this data collection effort, we identify districts as non-TK district if they report not having had TK in school year 2021-22. Throughout the paper, when we refer to non-TK districts, we are referring to districts that meet these two conditions.

After identifying TK and non-TK districts, we impose an additional district-level sample restriction based on our broader project evaluating Michigan TK in which we are interested in estimating treatment effect heterogeneity across districts. We focus on districts with a positive, precise discontinuity in TK enrollment at the birth date cutoff (December 1). This can be either when we consider the district on their own or when we pool observably similar districts. This condition excludes very small districts and larger districts with no obvious discontinuity at the cutoff. Overall, this restriction only drops 3.2% of students from the TK district sample.

Ultimately, our analysis sample contains 292 district×cohorts from 205 TK districts and 696 district×cohorts from 376 non-TK districts. Note that the number of non-TK *district×cohorts* is not exactly twice the number of non-TK *districts* because some small districts do not have students born within the 30-day bandwidth in both cohorts.

Once TK and non-TK districts are identified using these criteria, we then define the sample of students in TK districts at the district×cohort level. We define the relevant sample of students as students enrolled in districts that offered TK the year before one's "scheduled" kindergarten year (based on students' birthdays and the statewide kindergarten cutoff).

We examine the grade progression patterns of TK and early K students as a check on our sample selection. The results provide reassurance that our TK and non-TK districts are categorized correctly and that student-level program enrollment is accurate. 77% of early kindergarten students in non-TK districts and 82% of early kindergarten students in districts we confidently identify as TK districts move on to 1st grade the year after waiving into kindergarten. On the other hand, in districts that don't meet our requirements for reliable reporting, only 51% of students who appear to have waived into kindergarten early move on to 1st grade the following year. We believe it is likely that many of these students who appear to have waived in early K were actually enrolled in TK, which would explain why a relatively low share advances to 1st grade the next year. Further, we also see 98% of students enrolled in TK, in our TK offering districts, move on to regular kindergarten the following year.

B.2 Student-Level Sample Restrictions

We impose three main student-level restrictions on the sample. First, we use data on enrollment dates to drop students who attended neither TK nor kindergarten in a Michigan public school for at least 20 days. The purpose of this restriction is to ensure that "treated" students in our sample experienced at least some treatment practically. This 20-day restriction excludes 0.6% of all students (relative to a no restriction on days of enrollment).

Second, we drop students who attended neither TK nor kindergarten in a Michigan public school. This condition excludes Michigan students who appear in the Michigan student administrative data in later grades but attended early grades in private schools, home schools, or schools outside Michigan. We impose this condition because we are interested in the effects of early learning programs and students who do not attend early learning programs in Michigan are not relevant for our sample. This is not a large group of students. For example, such students constituted roughly 8% of Michigan third-graders in school year 2018-19.

Third, in line with our regression discontinuity research design that relies on student birthdates and the December 1 cutoff, we drop students with invalid birthday information. This includes students with no birthday information, multiple listed birthdays across the years, and birthdays that seem implausible.[10] we drop less than 0.1% of students due to birthday-related reasons.

B.3 Constructing Grade and District Variables

Once we define our set of TK and non-TK districts with reliable information, it is straightforward to identify TK, EK, and "on-schedule" kindergarten students. The grade level for all three types of students is grade 0 in the data. In addition, a separate variable identifies whether the student is enrolled in a TK program. Among the "grade 0" students who are identified as *not* attending TK, we differentiate early and on-schedule kindergarten students using a three pieces of information : students' birthdays, their year of kindergarten enrollment in the data, and institutional knowledge of Michigan's kindergarten cutoff dates.

We also clean the data so that each student is assigned to a single grade×school in a given school year. Some students appear in the data multiple times in the same school year. This is mainly either because they enrolled in different grades in the same year or different schools in Michigan in the same year. If we observe a student in the same grade multiple times in a year, we use the attendance and enrollment records to keep the observation with the most days attended and drop all other observations of the student. We use the same criterion When a student is observed in multiple grades kindergarten or above in the same year. However, we also observer students in more than one of TK, K, or an early childhood program in the same year in our data. For these cases, we use a more nuanced procedure that uses a combination of days in the program and grade progression to arrive at a single most relevant grade/program for the student. We provide details below.

[10] We consider a birthday implausible if it implies a student was born outside the two-year window that would be expected based on the year a student first enrolls in TK or K. The exact length of the window varies slightly by cohort to accommodate Michigan's changing kindergarten entry policies, but each window accounts for kindergarten redshirting and early entry. We also add a one-month cushion to the front and back ends of each window to account for non-compliance.

Among all TK students in our analysis sample, 3% are also enrolled in kindergarten in the same year. We use information on days attended and grade progression to determine whether to keep the student's TK or K observation. Specifically, we use the following algorithm:

- For students who attend TK first, followed by K, in year t:
  - If they attend TK in year $t+1$, we keep the TK observation in year t.
  - If they attend K in year $t+1$, we keep the observation in year t with more days attended. If a year t observation has a missing value for days attended, we do not keep that observation.
  - If they attend 1st or 2nd grade in year $t+1$, we keep the K observation in year t.
- For students who attend K first, followed by TK, in year t:
  - If they attend TK or K in year $t+1$, we keep the TK observation in year t.
  - If they attend 1st or 2nd grade in year $t+1$, we keep the K observation in year t.

Next, we reconcile students enrolled in TK or K in the same year as an early childhood program. We observe 0.4% of all TK students and 0.4% of all K students as also enrolled in an early childhood program in the same year. We cannot use days *attended* for this reconciliation because attendance information is not available in our data for early childhood programs. For these students we therefore compare the number of days students were *enrolled* in the programs for, keeping the observation with the higher number. For TK and K, we have data on the number of school days enrolled. For early childhood programs, the enrollment variable includes weekends, so we multiply by 5/7 to make it comparable with the TK and K variable.

Once the data is unique at the student×year level, we assign each student to the district they are first observed in. This allows us to capture the district a student *could* have enrolled in when they were on the margin of age-eligibility for TK and early K entry. Students who participate in TK are assigned to their TK district, and other students are assigned to their kindergarten district.

B.4 Summary Statistics

Table B1 presents summary statistics for the full third-grade sample. Unlike Table 1 in the paper, this table does not restrict the sample to students with test scores in third grade. The differences are minor and the overall takeaways are much the same.

Table B1. Summary Statistics for Sample, Including Students With Missing Outcome Data

	All Students		TK Students	Early K Students	
	TK Districts	Non-TK Districts	TK Districts	TK Districts	Non-TK Districts
Has 3rd grade test scores (%)	84	87	87	89	90
2015 Pre-K cohort (%)	50	51	37	45	30
2019 Pre-K cohort (%)	50	49	63	55	70
Female (%)	50	56	50	57	50
White (%)	48	33	74	63	77
Black (%)	36	51	12	18	11
Hispanic (%)	10	9	7	7	7
Asian American (%)	4	6	6	11	5
Other race (%)	2	1	1	1	1
Economically disadvantaged (%)	70	76	46	56	38
Prior state pre-K enrollment (%)	4	13	3	13	7
English learner (%)	11	15	10	21	6
Neighborhood White share (%)	64	51	85	81	86
Neighborhood poverty share (%)	18	21	9	13	8
Neighborhood unemployment rate (%)	13	13	20	19	20
Neighborhood BA attainment rate (%)	13	13	20	19	20
Neighborhood median household income ($)	49,666	45,951	66,494	61,976	69,120
School is in a city (%)	38	47	20	33	19
School is in a suburb (%)	31	34	50	46	50
School is in a town (%)	7	5	12	8	12
School is in a rural area (%)	23	14	18	13	20
Magnet school (%)	18	20	8	7	7
School enrollment (%)	443	479	444	441	433
School pupil:teacher ratio (%)	17.8	18.5	17.2	16.9	16.4
School FRL share (%)	67	73	43	49	41
Charter school (%)	30	46	3	4	5
District is in a city (%)	40	49	21	35	19
District is in a suburb (%)	29	32	54	49	57
District is in a town (%)	8	5	12	9	11
District is in a rural area (%)	23	13	13	7	13
District free- and reduced-price lunch share (%)	64	70	40	43	38
District English learner share (%)	11	11	7	12	6
District avg. 3rd grade math M-STEP score (SD)	−0.26	−0.34	0.25	0.19	0.23
Observations	8,410	1,689	9,902	923	2,043

Note: This table uses the third-grade sample described in Section 7.2.1, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Appendix C Derivation of Equation 1

In this section, we derive the expression for the intent-to-treat (ITT) effect shown in Equation 1 in the paper. The equation shows that the ITT effect is a weighted average of the TK and EK local average treatment effects (LATEs). For the sake of readability in this appendix, we'll use slightly different notation than in the paper:

- L_i is an indicator for being born to the left of December 1st (i.e., on or before).
- Treatment status, D_i, may take on values TK for TK, EK for waiving into K early, and 0 for doing neither TK nor waiving into K.
- $D_i(1)$ is the treatment a student would choose if they're to the left of the cutoff; $D_i(0)$ is what they would choose if they're to the right.
- Ω_x is the share of students who would participate in treatment x when eligible for all treatments, where x takes on values TK, EK, and 0 (neither TK nor EK).
- Y_i is a student's observed outcome and $Y_i(D)$ is their potential outcome under treatment D.

Now let's derive Equation 1. Focusing only on district×cohorts with TK, the ITT effect of being to the left of the cutoff can be written as:

$$ITT = E[Y_i|L_i = 1] - E[Y_i|L_i = 0]$$

We can break this equation apart by program complier types:

$$\begin{aligned} ITT = &\underbrace{\Omega_{TK}E[Y_i|D_i(1) = TK, L_i = 1] + \Omega_{EK}E[Y_i|D_i(1) = EK, L_i = 1] + \Omega_0 E[Y_i|D_i(1) = 0, L_i = 1]}_{\text{Left of cutoff}} \\ &- \underbrace{\Omega_{TK}E[Y_i|D_i(1) = TK, L_i = 0] - \Omega_{EK}E[Y_i|D_i(1) = EK, L_i = 0] - \Omega_0 E[Y_i|D_i(1) = 0, L_i = 0]}_{\text{Right of cutoff}} \end{aligned}$$

The IV exclusion restriction implies $E[Y_i|D_i(1) = 0, L_i = 1] = E[Y_i|D_i(1) = 0, L_i = 0]$ because outcomes depend on treatment, not treatment eligibility. These terms cancel out and we have:

$$\begin{aligned} ITT = &\Omega_{TK}E[Y_i|D_i(1) = TK, L_i = 1] + \Omega_{EK}E[Y_i|D_i(1) = EK, L_i = 1] \\ &- \Omega_{TK}E[Y_i|D_i(1) = TK, L_i = 0] - \Omega_{EK}E[Y_i|D_i(1) = EK, L_i = 0] \end{aligned}$$

Rearranging and substituting in potential outcomes, we have:

$$ITT = \Omega_{TK}\{E[Y_i|D_i(1) = TK, L_i = 1] - E[Y_i|D_i(1) = TK, L_i = 0]\} \\ + \Omega_{EK}\{E[Y_i|D_i(1) = EK, L_i = 1] - E[Y_i|D_i(1) = EK, L_i = 0]\}$$

$$ITT = \Omega_{TK}\{E[Y_i(TK)|D_i(1) = TK] - E[Y_i(0)|D_i(1) = TK]\} \\ + \Omega_{EK}\{E[Y_i(EK)|D_i(1) = EK] - E[Y_i(0)|D_i(1) = EK]\}$$

$$ITT = \Omega_{TK} \underbrace{E[Y_i(TK) - Y_i(0)|D_i(1) = TK]}_{LATE_{TK}} + \Omega_{EK} \underbrace{E[Y_i(EK) - Y_i(0)|D_i(1) = EK]}_{LATE_{EK}}$$

Appendix D RD Validity Checks

D.1 Attrition

Intuitively, the RD analysis requires that students born just before and just after December 1 be similar in ways other than treatment eligibility. Sample attrition poses a potential threat to this fundamental assumption. If the type of students who leave the sample are systematically different on either side of the cutoff, the resulting sample may not be continuous in observable or unobservable characteristics through the cutoff.

Attrition may occur for two reasons in our context. First, students may exit our data because they leave the Michigan public school system before 3rd grade. Second, students enrolled in a Michigan public school in 3rd grade may not have test score information in the data. Note that conditional on enrolling in a Michigan public school, there is no missingness in our special education outcomes.

Table D1. Attrition Estimates for the Third-Grade Sample

	Control Mean	Estimate	Standard Error	P-value
	Panel A. Non-TK Districts			
Ever observed in 1st grade	0.978	−0.002	0.007	0.742
Ever observed in 2nd grade	0.957	−0.010	0.011	0.352
Ever observed in 3rd grade	0.921	−0.001	0.012	0.962
Number of grades observed in between 1st and 3rd	2.856	−0.013	0.024	0.589
Has a 3rd grade math test score	0.856	−0.002	0.015	0.912
Has a 3rd grade ELA test score	0.857	−0.003	0.016	0.830
	Panel B. TK Districts			
Ever observed in 1st grade	0.979	−0.009	0.008	0.233
Ever observed in 2nd grade	0.951	−0.005	0.010	0.636
Ever observed in 3rd grade	0.907	0.004	0.013	0.785
Number of grades observed in between 1st and 3rd	2.837	−0.010	0.025	0.687
Has a 3rd grade math test score	0.877	−0.015	0.017	0.394
Has a 3rd grade ELA test score	0.876	−0.014	0.017	0.398

Note: These estimates are obtained via intent-to-treat RD models analogous to the LATE models described in Section 8. Inference is conducted analytically with clustering on the running variable. A student is coded as not having a 3rd grade test score if they are not observed in 3rd grade or if they are observed in 3rd grade but do not have a test score in their first year in 3rd grade. Special education placement is never missing in our data for those enrolled in a Michigan public school.

Table D1 shows that there does not appear to be differential attrition at the cutoff. In TK and non-TK districts, 91%-92% of students in our sample born near December 1 are also observed in 3rd grade in a later year. The difference in this likelihood at the cutoff is small and statistically insignificant. When we account for missing test score data, the probability of remaining in a Michigan public school and having 3rd grade test score data is around 86% for TK and non-TK districts. Again, there is no evidence that this attrition occurs differentially at the cutoff.

D.2 Density Manipulation

As another check of the RD assumptions necessary for causal inference, we investigate whether the density of our sample is continuous through the cutoff. In our context, it seems unlikely that families manipulate the running variable (i.e., children's birthdays) in order to gain access to TK or EK—either through misreporting or birth timing. The more plausible concern is that families with children born between September 2 and December 1 may relocate to districts that offer TK to gain access to the program. If this were the case, the children to the left of the cutoff may be systematically different than those to the right within district type.

We check for potential discontinuities in sample density using two approaches. First, we use the McCrary (2008) test with bandwidths of 5 and 10 days from the cutoff. Second, we use the Cattaneo et al. (2020) test that uses a mean squared error minimizing selection procedure to determine an optimal bandwidth. Our results are shown visually in Figure D1 and summarized in Table D2.

In TK districts, the McCrary tests do not find a statistically significant discontinuity in density. The estimate from the Cattaneo, Jansson, and Ma test is statistically significant at the 10% level, but the magnitude of the discontinuity is quite small. On the other hand, our tests do find a potential discontinuity in non-TK districts. All of our tests find that the density is lower in non-TK districts to the left of the cutoff, i.e., for children who are age-eligible for TK and EK.

What might explain the density discontinuity in non-TK districts? As mentioned before, it's possible families with children born between September 2 and December 1 move from non-TK districts to TK districts to gain access to TK. Indeed, this would result in the density being lower to the left of the cutoff. However, if this were the story, the density would likely be lower for the entirety of the left-of-cutoff sample, whereas Figure A1 shows that the density is lower only for those near the cutoff. Moreover, if this were the story, we would expect a corresponding *increase* in density to the left of the cutoff in districts with TK. Figure A1 shows that this is not the case. More generally, it is difficult explain why the density appears to dip close to the cutoff but not further away. Perhaps the birthday cutoff is most salient for families with children closest to the cutoff, but on the other hand, it is

Table D2. Tests for Density Manipulation in the Third-Grade Sample

		Non-TK Districts		TK Districts	
Test	Bandwidth	T-statistic	P-value	T-statistic	P-value
McCrary (2008)	5	−2.37	0.012	0.44	0.622
McCrary (2008)	10	−3.78	0.000	−1.03	0.288
Cattaneo, Jansson, and Ma (2020)	22.5	−4.09	0.000	−1.82	0.069

Note: The bandwidths for the McCrary (2008) tests are user-specified. The bandwidth for the Cattaneo, Jansson, and Ma (2020) tests is determined via a mean squared error minimizing selection procedure. The procedure selects 22.5 for non-TK and TK districts.

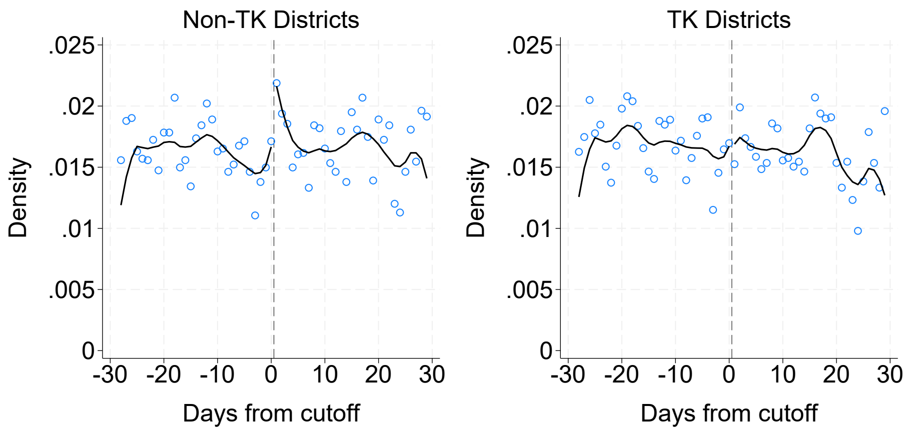
the older children who are more likely to be prepared for TK or EK.

A discontinuity in density is only problematic for our analysis insofar as it reflects differences between students to the left and right of the cutoff that are systematically related to test scores. In the next section, we explore whether observable student characteristics are continuous through the RD cutoff, which is a simple test of whether the density discontinuity is non-random. In short, we find no evidence that the density discontinuity reflects systematic differences between students.

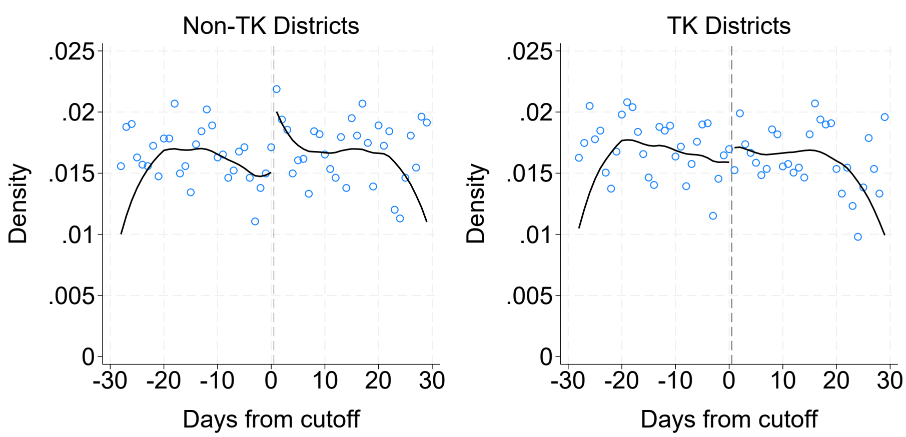
The density discontinuity in non-TK districts warrants some caution. However, given the lack of a coherent theoretical explanation for the non-TK discontinuity, the continuity of observable characteristics through the cutoff, the potential for the non-TK discontinuity to be driven by noise, and the lower level of concern about density discontinuity in TK districts, we view this issue as unlikely to bias our estimates.

Figure D1. Density by Birthdate in the Third-Grade Sample

(a) Panel A. Five-Day Bandwidth



(b) Panel B. Ten-Day Bandwidth



Notes: Each dot gives the density of students born on a particular day. The smoothed lines in each panel are estimated using the McCrary (2008) density test. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1).

D.3 Covariate Continuity

One of the key assumptions underlying our RD analysis is that student characteristics unrelated to treatment are continuous, on average, through the RD cutoff. We evaluate this assumption by estimating RD models with student characteristics as the outcome variable. Consistent with our main impact models, we use a bandwidth of ±30 days and specify linear relationships between the running variable and outcome variable that may vary on either side of the cutoff. Our results are presented in Table D3.

Overall, Table D3 provides strong evidence that baseline student characteristics are balanced at the cutoff. The vast majority of estimates are small and indistinguishable from 0. Only a small number of estimates are statistically significant at conventional levels, as would be expected to happen by chance when testing so many hypotheses.

To facilitate a summary test of covariate continuity through the cutoff, we combine the observable student characteristics into a single summary statistic. Specifically, we construct a measure of predicted 3rd grade math scores and then estimate RD models using the predicted score as the outcome variable. To obtain the relationship between student characteristics and test scores, we estimate a linear regression with 3rd grade math scores as the outcome variable and the following variables as predictors: student demographics (sex, race, and economic disadvantage); neighborhood characteristics (White share, poverty share, BA attainment rate, and log of median household income); district characteristics (share of students eligibile for free- or reduced-price lunch, log of student enrollment, and urbanicity level); and region of Michigan. In both TK and non-TK districts, the estimated discontinuity in predicted test scores is small and indistinguishable from 0 (see the first row of each panel in Table D3).

The only characteristic with a statistically significant discontinuity in TK and non-TK districts is "prior state Pre-K enrollment." By this we mean enrollment in Michigan's income-targeted state Pre-K program the year before a student is on the margin of being age-eligible for TK or early kindergarten. This discontinuity is expected given our knowledge of the Michigan Pre-K landscape. Michigan's Pre-K program, called the Great Start Readiness Program (GSRP), is intended for students who turn 4 years old by September 1. However, children who turn 4 between September 2 and December 1 are sometimes eligible to enroll in GSRP as 3-year-olds when space is available after initial enrollment. Therefore, GSRP enrollment has the same birthday cutoff as TK and EK, although it applies two years before "on-schedule" kindergarten enrollment instead of one year before.

Consistent with our understanding of the GSRP age-eligibility rules, Table D3 shows that students born after December 1 do not enroll in GSRP before their Pre-K year. On the other hand, 8.8% of students at the cutoff in non-TK districts and 5.7% of students at the cutoff in TK districts enroll in GSRP before their Pre-K year.

If GSRP enrollment as a 3-year-old has a non-zero impact on student outcomes in 3rd grade, the discontinuity in enrollment at the RD cutoff could bias our estimates. In our main impact models, our omission of GSRP as a treatment option implicitly assumes that enrolling in GSRP before one's Pre-K year does not have an effect on test scores that persists to 3rd grade. We view this assumption as a reasonable benchmark. For one, the discontinuity in 3-year-old GSRP enrollment is not particularly large, meaning GSRP's impact would have to be especially large to affect our estimates. Second, compared to TK and EK, the curriculum used in GSRP is typically less focused on academics and its teacher workforce is paid substantially less, making it plausible that test score impacts do not persist through 3rd grade. Third, the impacts of GSRP as a 3-year-old would have to persist conditional on the various child care and preschool arrangements children experience the following year, i.e., in their Pre-K year. Assuming that potential test score impacts do not persist through 3rd grade is consistent with RCT evidence from the federal Head Start Impact Study, which found that cognitive impacts from 3-year-old Head Start enrollment did not persist through kindergarten (Puma et al., 2012).

Returning to our discussion on density manipulation from the previous section, this analysis of covariate continuity provides strong reassurance that the density discontinuity in non-TK districts is not particularly concerning. We observe several important and predictive characteristics that feed into our predicted 3rd grade math scores, and we find no discontinuity in this measure. If the density discontinuity reflects unobservable differences between students across the cutoff, these differences would have to be orthogonal to all the observable characteristics we account for, which seems highly unlikely.

Table D3. Covariate Continuity Through the Cutoff for the Third-Grade Sample

	Control Mean	Estimate	Standard Error	P-value
	Panel A. Non-TK Districts			
Predicted 3rd grade math score (SD)	−0.095	−0.015	0.019	0.434
Female	0.490	−0.006	0.019	0.755
White	0.491	−0.045	0.022	0.045
Black	0.355	0.022	0.020	0.256
Hispanic	0.096	0.011	0.010	0.254
Asian American	0.039	0.004	0.008	0.610
Other race	0.019	0.007	0.006	0.224
Economically disadvantaged	0.697	0.017	0.017	0.331
Prior state Pre-K enrollment	0.000	0.088	0.007	0.000
Neighborhood White share	0.640	−0.012	0.014	0.396
Neighborhood poverty share	0.175	−0.004	0.008	0.580
Neighborhood BA attainment rate	0.133	0.004	0.006	0.528
Neighborhood median HH income	49,213	−634	1,155	0.585
School is in a city	0.381	0.035	0.024	0.156
School is in a suburb	0.304	0.001	0.017	0.956
School is in a town	0.075	−0.024	0.012	0.046
School is in a rural area	0.240	−0.012	0.019	0.544
Magnet school	0.176	0.029	0.014	0.036
Log(school enrollment)	5.941	−0.007	0.016	0.692
School pupil:teacher ratio	17.8	0.0	0.2	0.806
School FRL share	0.668	0.005	0.012	0.717
	Panel B. TK Districts			
Predicted 3rd grade math score (SD)	0.272	0.020	0.017	0.242
Female	0.511	−0.021	0.020	0.291
White	0.732	0.005	0.024	0.831
Black	0.124	−0.009	0.012	0.439
Hispanic	0.075	−0.004	0.012	0.711
Asian American	0.057	0.005	0.014	0.730
Other race	0.011	0.003	0.004	0.354
Economically disadvantaged	0.469	−0.026	0.019	0.166
Prior state Pre-K enrollment	0.000	0.057	0.005	0.000
Neighborhood White share	0.849	0.005	0.008	0.521
Neighborhood poverty share	0.089	0.004	0.004	0.261
Neighborhood BA attainment rate	0.197	0.001	0.004	0.863
Neighborhood median HH income	66,425	−11	1,099	0.992
School is in a city	0.190	0.009	0.015	0.540
School is in a suburb	0.505	0.005	0.019	0.812
School is in a town	0.122	0.001	0.016	0.940
School is in a rural area	0.184	−0.015	0.016	0.348
Magnet school	0.078	−0.006	0.012	0.594
Log(school enrollment)	6.056	−0.062	0.017	0.000
School pupil:teacher ratio	17.3	−0.4	0.1	0.018
School FRL share	0.429	0.007	0.008	0.437

Notes: These estimates are obtained via intent-to-treat RD models analogous to the LATE models described in Section 8, using the covariates as the outcomes. Inference is conducted analytically with clustering on the running variable. Predicted 3rd grade math scores are the predicted values from regressing observed scores on the covariates in this table. "Prior state pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Appendix E Subgroup Estimates in the "Relaxed Assumptions" Approach

Table E1 shows the subgroup estimates and weights that feed into our "relaxed assumptions" estimation approach for third-grade test scores. For each outcome domain, we estimate eight EK LATEs, one for each group defined by sex × race × economic disadvantage status. Most of the subgroup estimates are negative, as we expect, although some of the subgroups with small samples have imprecise positive estimates. We use the shares in the "Share in TK Districts" column as weights to aggregate the subgroup LATEs into a single $LATE_{EK}$ estimate. The shares in the "Share in Non-TK Districts" column are provided as a comparison point.

Our "relaxed assumptions" estimate for $LATE_{EK}$ is greater in magnitude than our baseline estimate because demographic cells with large $LATE_{EK}$ estimates are a larger fraction of all EK compliers in TK districts than in non-TK districts. In particular, female students who are White or Asian (regardless of economic disadvantage status) have large negative estimates and receive much more weight in TK districts than in non-TK districts. Recall from Table 1 in the paper that students in districts with TK are substantially more likely to be White.

Table E1. Subgroup EK LATE Estimates and EK Complier Shares (Third-Grade Test Scores)

Sex	White or Asian	Economically Disadvantaged	$LATE_{EK}$	Share in TK Districts	Share in Non-TK Districts
			Panel A. 3rd Grade Math		
Male	No	No	0.260	.02	.03
Male	No	Yes	−0.298	.09	.25
Male	Yes	No	−0.543	.15	.06
Male	Yes	Yes	0.362	.16	.09
Female	No	No	1.036	.03	.04
Female	No	Yes	−0.571	.12	.28
Female	Yes	No	−0.366	.22	.10
Female	Yes	Yes	−0.879	.21	.13
			Panel B. 3rd Grade ELA		
Male	No	No	0.386	.02	.03
Male	No	Yes	−0.075	.09	.25
Male	Yes	No	−0.162	.15	.06
Male	Yes	Yes	0.237	.16	.09
Female	No	No	−1.558	.03	.04
Female	No	Yes	−0.208	.11	.28
Female	Yes	No	−0.107	.22	.10
Female	Yes	Yes	−0.638	.21	.13

Note: The point estimates in the $LATE_{EK}$ column are estimated using models analogous to Equations 2 and 3 from the paper, but with covariates excluded because the subgroups are defined based on the covariates. The last two columns are the share of all EK students (within our 30 day bandwidth) in TK and non-TK districts who belong to each demographic cell. The shares do not always sum to 1 due to rounding. The shares are slightly different for math and ELA due to small differences in missing test score data by domain.

Appendix F Details on the KRA Analysis

To estimate the effect of TK enrollment on kindergarten readiness, we use Kindergarten Readiness Assessment (KRA) data shared with us by the Washtenaw Intermediate School District (ISD). Note that the data contain assessments from throughout Michigan despite being maintained by a single ISD. The KRA is a psychometrically validated assessment with a composite score and four subscale scores: (1) social foundations; (2) language and literacy; (3) mathematics; and (4) physical and motor development. For our analysis, we convert scale scores into standardized scores using means and standard deviations from the universe of KRA-takers in Ohio (see Table F1).[11] We do not have any KRA data for our 2014-2015 cohort. For the 2018-2019 cohort, not all districts administered the KRA, and only some students within KRA districts were assessed.

Estimating impacts for our 2018-2019 cohort requires KRA data from two years; children who enrolled in kindergarten early were assessed in fall 2018, whereas TK and other students were assessed in fall 2019. A greater share of districts administered assessments in fall 2019 than fall 2018, but there is a sufficient number of districts that administered assessments in both years for our analysis to be viable. For our primary sample, we drop all districts that assessed less than 20% of regular kindergartners in fall 2018 and in fall 2019. With a 30-day badwidth, this sample consists of 58 non-TK districts and 61 TK districts. As Table F6 shows, this sample is very similar to the broader set of districts we use to estimate impacts on special education and third-grade test scores (see Table 1).

F.1 Internal Validity Checks

The fact that not all students within KRA districts were assessed poses a potential validity threat to our analysis. Recall that our RD identification strategy assumes that students in our sample born just before the TK birthday cutoff are equivalent, on average, to children born just after the cutoff. If assessment rates change discontinuously at the cutoff, and for non-random reasons, the children on each side of the cutoff may be systematically different.

There are at least two reasons assessed children born before the cutoff may be different than assessed children born after. First, assessment rates tended to be higher in fall 2019 than fall 2018. Thus, EK students (who enrolled in regular kindergarten in fall 2018) may be less likely to have been assessed than TK and other students (who enrolled in regular kindergarten in fall 2019). Second, although state policy was to either assess all kindergartners or to randomly select students to assess, our analysis of the KRA data suggests that districts

[11] For students who waive into kindergarten early, we standardize scores using means and standard deviations from fall 2018. For TK and other students, we use means and standard deviations from fall 2019.

did not select students randomly. As Table F2 shows, in our primary KRA sample, White students and non-economically disadvantaged students were more likely to be assessed than other students. As hypothesized, EK students were less likely to be assessed.

This non-random selection into assessment may bias our RD estimates. To examine the robustness of our results, we create three additional samples intended to safeguard, in some ways, against non-random assessment:

1. **80%+ Coverage.** The first robustness sample drops all districts that assessed less than 80% of regular kindergartners in fall 2018 and in fall 2019.

2. **90%+ Coverage.** The second robustness sample drops all districts that assessed less than 90% of regular kindergartners in fall 2018 and in fall 2019.

3. **Imputing Missing Outcomes.** The third robustness sample begins with the primary sample (20%+ coverage) but imputes KRA scores for students who were not assessed. The imputed scores are predicted values from a regression of KRA scores on the following predictor variables: TK/EK enrollment, sex, race, economic disadvantage status, and district fixed effects.

When at least 80% or 90% of students are assessed, there is much less scope—although still some scope—for differential assessment at the cutoff. For the third robustness sample, we eliminate differential assessment rates by filling in missing outcome data with plausible values. RD estimates may still be biased if non-assessed students are not assessed due to unobservable (to the researcher) reasons, which is likely, but accounting for non-assessed students with plausible outcome values should be an improvement over not accounting for them at all. Tables F6, F7, F8, and F9 give summary statistics for the primary KRA sample and for the three robustness samples.

Table F3 formally confirms that there may be differential attrition in terms of having a KRA score at the RD cutoff. The point estimates, though not all statistically significant, suggest that fewer students to the left of the cutoff have KRA scores. Nevertheless, Table F5 shows that students' observable characteristics are generally quite smooth through the cutoff. Out of the 21 characteristics we examine, very few appear to have discontinuities—no more than would be expected by chance.

F.2 Robustness Checks

Tables F10, F11, F12, F13, and F14 present RD results for the overall and subscore KRA outcomes. Each table includes results for the primary and robustness samples using 30-, 60-, and 90-day bandwidths. Across all samples and bandwidths, we generally find very large negative effects from EK and large positive effects from TK. The exact values vary somewhat, but they are qualitatively similar across approaches.

Table F15 presents impact estimates using our primary (20%+) sample with and without covariates. Our results are robust to excluding covariates. Figure F2 presents placebo checks that estimate impacts at false (i.e., placebo) birthday cutoffs. Reassuringly, with the possible exception of the mathematics subdomain, the estimates at the true RD cutoff are extreme relative to the placebo cutoffs.

Table F16 conducts a final stress test against the attrition issue already discussed. In the spirit of a Manski bounding exercise, we fill in missing KRA scores with extreme values meant to produce ultra-conservative impact estimates. In non-TK districts, for students born to the left (right) of the RD cutoff, we replace missing scores with scores at the 95th (5th) percentile of the observed distribution in non-TK districts. This attenuates our EK LATE estimate and, thus, the eventual TK LATE estimate. In TK districts, for students born to the left (right) of the RD cutoff, we replace missing scores with scores at the 5th (95th) percentile of the observed distribution in TK districts. This attenuates our TK LATE estimate. Table F16 shows the resulting RD estimates. Reassuringly, the TK LATEs are still positive and large (though much attenuated, as expected).[12] The implication is that attrition almost certainly cannot be producing a large enough bias to qualitatively overturn our results.

F.3 Consistency with Third-Grade Analysis

Lastly, we conduct an additional analysis to ensure that the results from our KRA and third-grade test score analyses are comparable. Columns 1, 2, and 3 of Table F17 give TK and EK LATE estimates for KRA outcomes using the 20%+ coverage sample and 30- and 90-day bandwidths. Columns 4 and 5 show analogous TK and EK LATE estimates but for third-grade math and ELA. The sample in these columns is the subset of students in columns 1, 2, and 3 who also have third-grade test scores. The estimates in columns 4 and 5 are very similar to the estimates from our main third-grade analysis (see Table 6), suggesting that our KRA LATE estimates may generalize to districts that did not administer the KRA.

Columns 6 and 7 give TK and EK LATE estimates for third-grade test scores among a slightly different sample. We continue to use districts in the 20%+ KRA coverage sample, but we use all students from these districts who have third-grade test scores *regardless* of whether they have KRA scores. This analysis provides a useful test of the potential bias in the KRA analysis. If the KRA estimates were solely driven by differential KRA assessment,

[12] We do not conduct this exercise for the 20%+ coverage sample because so many students are missing outcome data and imputing them all with worst-case values would yield meaningless results.

we would not expect to see any effects in columns 6 and 7. In practice, the estimates are slightly smaller but still positive and large. This may suggest that the KRA estimates are slightly upwardly biased. However, it would be highly unlikely to see impacts on third-grade outcomes in columns 6 and 7 if there were not large impacts on kindergarten readiness.

Table F1. KRA Scale Scores of Children in Ohio

	Overall	Social Foundations	Language and Literacy	Mathematics	Physical/Motor Development
	Panel A. Fall 2018				
Mean	266.6	274.6	265.2	264.4	271.5
Standard deviation	13.7	21.1	14.5	14.3	18.4
Observations	119,309	119,295	119,297	119,296	119,296
	Panel B. Fall 2019				
Mean	266.7	275.1	265	264.3	272
Standard deviation	13.7	21.1	14.5	14.4	18.4
Observations	119,339	119,343	119,345	119,343	119,343

Note: This table provides statistics on the scores of all Ohio kindergartners who took the KRA in the fall of 2018 and 2019, provided to us by the Ohio Department of Children and Youth and published here with permission. As discussed in Section 7.3, we use these statistics to standardize Michigan kindergartners' KRA scale scores.

Table F2. Non-Randomness of KRA Administration

	District Fixed Effects				District-Year Fixed Effects			
	Unconditional Means				Unconditional Means			
	Has KRA	No KRA	Strata-Adjusted Difference	P-value	Has KRA	No KRA	Strata-Adjusted Difference	P-value
Predicted score (SD)	−0.025	−0.006	0.051	0.008	−0.025	−0.006	0.064	0.000
TK last year	0.171	0.214	−0.013	0.481	0.171	0.214	0.017	0.243
EK this year	0.111	0.148	−0.064	0.026	0.111	0.148	−0.017	0.243
Neither TK nor EK	0.717	0.639	0.077	0.015	0.717	0.639	0.000	1.000
Female	0.488	0.485	0.013	0.343	0.488	0.485	0.016	0.296
Economically disadvantaged	0.556	0.536	−0.055	0.000	0.556	0.536	−0.067	0.000
White	0.697	0.667	0.043	0.002	0.697	0.667	0.048	0.001
Black	0.149	0.135	−0.032	0.000	0.149	0.135	−0.033	0.001
Hispanic	0.118	0.146	−0.007	0.603	0.118	0.146	−0.011	0.392
Asian American	0.025	0.041	−0.006	0.247	0.025	0.041	−0.006	0.223
Other race	0.010	0.011	0.001	0.601	0.010	0.011	0.001	0.513
Observations	5,988	2,927			5,984	2,910		

Note: This table assesses the randomness of the KRA administration within our primary KRA sample (20%+ coverage) among children born within 90 days of the TK/EK birthday cutoff. A student is coded as having a KRA score if we observe a score at the beginning of regular kindergarten (with or without a waiver) the first time a student is in kindergarten. Predicted scores are the predicted values from regressing observed scores on the characteristics in this table. To obtain the estimates in the "strata-adjusted difference" column, we regress each student characteristic on a binary indicator for having a KRA score in kindergarten. We also include fixed effects for each district or for each district-year based on the year a student was first in kindergarten. Inference is conducted analytically with clustering either at the district level or the district-year level.

Table F3. Attrition Estimates for the KRA Samples

	Non-TK Districts				TK Districts			
	Control Mean	Estimate	Standard Error	P-value	Control Mean	Estimate	Standard Error	P-value
					Panel A. 20%+ Coverage			
Observed in K	1.000	0.000	0.000	-	0.996	−0.019	0.009	0.033
Has a KRA score	0.806	−0.146	0.051	0.005	0.632	−0.045	0.038	0.241
					Panel B. 80%+ Coverage			
Observed in K	1.000	0.000	0.000	-	0.999	0.000	0.012	0.987
Has a KRA score	0.987	−0.074	0.049	0.137	0.949	−0.059	0.033	0.079
					Panel C. 90%+ Coverage			
Observed in K	1.000	0.000	0.000	-	0.992	0.015	0.014	0.290
Has a KRA score	0.945	0.098	0.066	0.143	0.927	−0.013	0.037	0.719

Note: These estimates are obtained via intent-to-treat RD models analogous to the LATE models described in Section 8. Inference is conducted analytically with clustering on the running variable.

Table F4. Tests for Density Manipulation in the KRA Sample (20%+ Coverage)

		Non-TK Districts		TK Districts	
Test	Bandwidth	T-statistic	P-value	T-statistic	P-value
McCrary (2008)	5	−0.14	0.922	−0.82	0.446
McCrary (2008)	10	−0.78	0.420	−1.93	0.036
Cattaneo, Jansson, and Ma (2020)	22.5	−0.30	0.762	−2.50	0.012

Note: The bandwidths for the McCrary (2008) tests are user-specified. The bandwidth for the Cattaneo, Jansson, and Ma (2020) tests is determined via a mean squared error minimizing selection procedure. The procedure selects 22.5 for non-TK and TK districts.

Table F5. Covariate Continuity Through the Cutoff for the KRA Sample (20%+ Coverage)

	Control Mean	Estimate	Standard Error	P-value
	Panel A. Non-TK Districts			
Predicted KRA score	0.024	−0.035	0.025	0.165
Female	0.550	−0.025	0.030	0.399
White	0.501	−0.075	0.037	0.048
Black	0.335	0.059	0.034	0.084
Hispanic	0.112	0.014	0.018	0.445
Asian American	0.040	−0.004	0.011	0.719
Other race	0.013	0.006	0.007	0.365
Economically disadvantaged	0.721	0.027	0.022	0.220
Prior state Pre-K enrollment	0.000	0.132	0.012	0.000
Neighborhood White share	0.649	−0.033	0.021	0.120
Neighborhood poverty share	0.177	−0.006	0.011	0.561
Neighborhood BA attainment rate	0.133	0.004	0.007	0.616
Neighborhood median HH income	49,492	−2,197	1,862	0.243
School is in a city	0.364	0.055	0.034	0.112
School is in a suburb	0.297	−0.003	0.026	0.902
School is in a town	0.080	−0.014	0.016	0.399
School is in a rural area	0.259	−0.038	0.023	0.102
Magnet school	0.154	0.027	0.019	0.161
Log(school enrollment)	5.918	0.012	0.022	0.575
School pupil:teacher ratio	18.0	0.1	0.4	0.823
School FRL share	0.690	0.018	0.016	0.253
	Panel B. TK Districts			
Predicted KRA score	0.175	0.019	0.018	0.308
Female	0.492	−0.023	0.025	0.371
White	0.700	0.020	0.031	0.535
Black	0.155	−0.008	0.018	0.648
Hispanic	0.084	−0.013	0.013	0.329
Asian American	0.056	−0.003	0.016	0.871
Other race	0.005	0.004	0.004	0.348
Economically disadvantaged	0.512	−0.039	0.023	0.091
Prior state Pre-K enrollment	0.000	0.060	0.007	0.000
Neighborhood White share	0.833	0.006	0.009	0.510
Neighborhood poverty share	0.090	0.003	0.005	0.527
Neighborhood BA attainment rate	0.194	0.001	0.005	0.889
Neighborhood median HH income	65,533	−732	1,109	0.512
School is in a city	0.206	−0.016	0.019	0.415
School is in a suburb	0.506	0.014	0.022	0.539
School is in a town	0.102	0.014	0.017	0.429
School is in a rural area	0.186	−0.012	0.021	0.578
Magnet school	0.082	−0.010	0.013	0.419
Log(school enrollment)	6.076	−0.030	0.019	0.114
School pupil:teacher ratio	17.1	0.0	0.2	0.804
School FRL share	0.459	0.000	0.010	0.975

Notes: These estimates are obtained via intent-to-treat RD models analogous to the LATE models described in Section 8, using the covariates as the outcomes. Inference is conducted analytically with clustering on the running variable. Predicted KRA scores are the predicted values from regressing observed scores on the covariates in this table. "Prior state pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Table F6. Summary Statistics for KRA Sample With 20%+ Coverage

	Non-TK Districts		TK Districts		
	All Students	Early K Students	All Students	Early K Students	TK Students
2015 Pre-K cohort (%)	0	0	0	0	0
2019 Pre-K cohort (%)	100	100	100	100	100
Female (%)	50	58	51	57	51
White (%)	58	47	77	75	77
Black (%)	20	26	10	20	9
Hispanic (%)	20	24	8	5	9
Asian American (%)	1	3	3	0	4
Other race (%)	1	0	1	0	1
Economically disadvantaged (%)	74	71	48	55	44
Prior state-funded pre-K enrollment (%)	10	31	4	16	6
English learner (%)	10	17	4	5	4
Neighborhood White share (%)	78	74	87	82	87
Neighborhood poverty share (%)	15	19	8	10	7
Neighborhood unemployment rate (%)	14	13	19	19	19
Neighborhood BA attainment rate (%)	14	13	19	19	19
Neighborhood median household income ($)	45,807	44,392	62,760	60,099	63,953
School is in a city (%)	44	53	12	18	8
School is in a suburb (%)	5	7	46	45	51
School is in a town (%)	15	15	17	11	13
School is in a rural area (%)	36	25	25	27	29
Magnet school (%)	7	7	15	13	11
School enrollment	365	352	440	390	424
School pupil:teacher ratio	16.4	16.9	17	16.5	16.4
School FRL share (%)	71	73	46	53	43
Charter school (%)	7	11	1	0	2
District is in a city (%)	45	56	10	18	5
District is in a suburb (%)	4	6	55	57	62
District is in a town (%)	14	21	14	7	12
District is in a rural area (%)	37	18	20	18	21
District free- and reduced-price lunch share (%)	67	68	42	46	39
District English learner share (%)	12	12	4	5	3
District avg. 3rd grade math M-STEP score (SD)	−0.25	−0.25	0.20	0.14	0.21
Observations	650	72	1,293	56	281

Note: This table uses the KRA sample with 20%+ assessment coverage, described in Sections 7.2.2 and F, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Table F7. Summary Statistics for KRA Sample With 80%+ Coverage

	Non-TK Districts		TK Districts		
	All Students	Early K Students	All Students	Early K Students	TK Students
2015 Pre-K cohort (%)	0	0	0	0	0
2019 Pre-K cohort (%)	100	100	100	100	100
Female (%)	50	52	49	57	47
White (%)	69	56	80	71	80
Black (%)	16	32	11	25	10
Hispanic (%)	12	4	6	4	8
Asian American (%)	2	8	2	0	1
Other race (%)	2	0	1	0	1
Economically disadvantaged (%)	69	68	54	68	47
Prior state-funded pre-K enrollment (%)	10	48	4	11	9
English learner (%)	2	8	1	4	1
Neighborhood White share (%)	83	71	88	80	89
Neighborhood poverty share (%)	15	22	10	14	8
Neighborhood unemployment rate (%)	11	9	17	15	16
Neighborhood BA attainment rate (%)	11	9	17	15	16
Neighborhood median household income ($)	45,486	39,447	57,122	55,616	57,834
School is in a city (%)	20	28	15	32	8
School is in a suburb (%)	9	20	33	21	39
School is in a town (%)	21	20	27	14	21
School is in a rural area (%)	50	32	25	32	32
Magnet school (%)	9	8	23	25	15
School enrollment	376	328	427	366	403
School pupil:teacher ratio	16.6	17	17	16.9	16.3
School FRL share (%)	65	70	49	59	48
Charter school (%)	6	8	1	0	3
District is in a city (%)	21	32	16	32	8
District is in a suburb (%)	8	16	33	25	42
District is in a town (%)	17	20	23	7	19
District is in a rural area (%)	53	32	28	36	30
District free- and reduced-price lunch share (%)	62	67	45	50	43
District English learner share (%)	3	3	2	2	2
District avg. 3rd grade math M-STEP score (SD)	−0.26	−0.26	0.06	−0.08	0.08
Observations	331	25	750	28	165

Note: This table uses the KRA sample with 80%+ assessment coverage, described in Sections 7.2.2 and F, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Table F8. Summary Statistics for KRA Sample With 90%+ Coverage

	Non-TK Districts		TK Districts		
	All Students	Early K Students	All Students	Early K Students	TK Students
2015 Pre-K cohort (%)	0	0	0	0	0
2019 Pre-K cohort (%)	100	100	100	100	100
Female (%)	59	67	49	73	46
White (%)	87	89	86	91	83
Black (%)	1	0	5	9	7
Hispanic (%)	5	0	6	0	8
Asian American (%)	3	11	2	0	1
Other race (%)	3	0	1	0	2
Economically disadvantaged (%)	63	67	50	45	39
Prior state-funded pre-K enrollment (%)	12	33	2	0	7
English learner (%)	1	11	0	0	0
Neighborhood White share (%)	95	95	93	89	94
Neighborhood poverty share (%)	9	10	8	8	7
Neighborhood unemployment rate (%)	11	12	17	19	17
Neighborhood BA attainment rate (%)	11	12	17	19	17
Neighborhood median household income ($)	47,410	45,604	59,218	67,279	59,162
School is in a city (%)	9	11	13	27	6
School is in a suburb (%)	0	0	38	18	48
School is in a town (%)	37	44	28	27	18
School is in a rural area (%)	54	44	21	27	28
Magnet school (%)	0	0	21	45	7
School enrollment	468	388	463	470	426
School pupil:teacher ratio	17.5	18.6	17.8	18.2	16.9
School FRL share (%)	59	62	44	41	43
Charter school (%)	14	11	1	0	3
District is in a city (%)	9	11	13	27	6
District is in a suburb (%)	0	0	38	27	51
District is in a town (%)	37	44	23	9	16
District is in a rural area (%)	54	44	26	36	27
District free- and reduced-price lunch share (%)	56	57	39	35	39
District English learner share (%)	3	3	1	1	1
District avg. 3rd grade math M-STEP score (SD)	−0.12	−0.12	0.16	0.20	0.16
Observations	94	N<10	548	11	123

Note: This table uses the KRA sample with 90%+ assessment coverage, described in Sections 7.2.2 and F, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Table F9. Summary Statistics for KRA Sample With 20%+ Coverage and Imputed Outcome Data

	Non-TK Districts		TK Districts		
	All Students	Early K Students	All Students	Early K Students	TK Students
2015 Pre-K cohort (%)	0	0	0	0	0
2019 Pre-K cohort (%)	100	100	100	100	100
Female (%)	50	53	51	49	54
White (%)	57	53	75	63	77
Black (%)	20	22	11	27	11
Hispanic (%)	20	22	9	6	8
Asian American (%)	1	3	4	5	4
Other race (%)	1	0	1	0	1
Economically disadvantaged (%)	76	75	49	61	41
Prior state-funded pre-K enrollment (%)	10	27	4	16	7
English learner (%)	11	14	5	8	4
Neighborhood White share (%)	78	76	87	82	88
Neighborhood poverty share (%)	15	17	8	10	7
Neighborhood unemployment rate (%)	13	14	20	19	20
Neighborhood BA attainment rate (%)	13	14	20	19	20
Neighborhood median household income ($)	46,130	44,391	64,811	58,916	67,258
School is in a city (%)	0	0	0	0	0
School is in a suburb (%)	4	4	51	53	53
School is in a town (%)	18	21	13	10	12
School is in a rural area (%)	34	28	26	22	30
Magnet school (%)	7	6	12	12	11
School enrollment	365	342	432	387	425
School pupil:teacher ratio	16.3	15.9	16.9	16.2	16.4
School FRL share (%)	72	73	44	51	41
Charter school (%)	7	9	1	0	1
District is in a city (%)	45	49	8	14	4
District is in a suburb (%)	3	3	62	61	69
District is in a town (%)	17	24	12	8	11
District is in a rural area (%)	35	23	19	17	17
District free- and reduced-price lunch share (%)	68	68	40	44	37
District English learner share (%)	13	12	4	5	4
District avg. 3rd grade math M-STEP score (SD)	−0.24	−0.24	0.24	0.14	0.27
Observations	856	116	1,997	104	454

Note: This table uses the KRA sample with 20%+ assessment coverage and imputed outcome data, described in Sections 7.2.2 and F, restricted to students born within 30 days of the TK cutoff. All statistics are calculated at the student level. "Early K" students are those who use a waiver to enroll in regular kindergarten because they turn five after the kindergarten birthday cutoff. "Prior state-funded pre-K enrollment" refers to enrollment in Michigan's Great Start Readiness Program (GSRP) prior to one's Pre-K year. "School FRL share" refers to the share of students in a school who are eligible for free- and reduced-price lunch. Neighborhood characteristics are measured at the block group level in the 2010 census.

Table F10. Impacts on KRA Scores (Overall)

	30-Day Bandwidth		60-Day Bandwidth		90-Day Bandwidth	
	Complier Control Mean	LATE	Complier Control Mean	LATE	Complier Control Mean	LATE
			Panel A. 20%+ Coverage			
TK	−0.360	0.914**	−0.167	0.841***	−0.191	0.771***
[P-value]		[0.027]		[0.000]		[0.000]
EK	1.628	−1.841*	1.508	−1.753***	1.454	−1.811***
[P-value]		[0.079]		[0.005]		[0.001]
Observations	1,943		3,912		5,988	
			Panel B. 80%+ Coverage			
TK	−1.140	1.590**	−0.353	1.014***	−0.417	0.979***
[P-value]		[0.027]		[0.001]		[0.000]
EK	2.588	−2.837	1.551	−1.727	1.855	−2.060
[P-value]		[0.318]		[0.165]		[0.105]
Observations	1,081		2,248		3,421	
			Panel C. 90%+ Coverage			
TK	−0.487	0.949**	0.020	0.672***	−0.225	0.809***
[P-value]		[0.020]		[0.003]		[0.000]
EK	2.339	−2.430	1.820	−1.655	2.128	−1.922
[P-value]		[0.401]		[0.204]		[0.182]
Observations	642		1,323		2,008	
			Panel D. 20%+ Coverage (Imputing Missing Outcomes)			
TK	−0.158	0.770***	−0.087	0.768***	−0.148	0.764***
[P-value]		[0.000]		[0.000]		[0.000]
EK	0.698	−1.025**	0.720	−1.097***	0.697	−1.125***
[P-value]		[0.023]		[0.001]		[0.000]
Observations	2,853		5,816		8,915	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Across the columns, we re-estimate models using either a 30-, 60-, or 90-day bandwidth. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

Table F11. Impacts on KRA Scores (Social Foundations)

	30-Day Bandwidth		60-Day Bandwidth		90-Day Bandwidth	
	Complier Control Mean	LATE	Complier Control Mean	LATE	Complier Control Mean	LATE
			Panel A. 20%+ Coverage			
TK	−0.674	0.934**	−0.509	0.862***	−0.492	0.758***
[P-value]		[0.037]		[0.001]		[0.000]
EK	1.585	−2.127*	1.176	−1.596**	1.375	−1.931***
[P-value]		[0.072]		[0.018]		[0.001]
Observations	1,943		3,912		5,988	
			Panel B. 80%+ Coverage			
TK	−1.647	1.905**	−0.713	1.133***	−0.797	1.085***
[P-value]		[0.021]		[0.001]		[0.000]
EK	4.615	−5.051	1.591	−2.033	2.557	−2.933**
[P-value]		[0.119]		[0.107]		[0.045]
Observations	1,081		2,248		3,421	
			Panel C. 90%+ Coverage			
TK	−0.561	0.918**	−0.384	0.896***	−0.574	0.969***
[P-value]		[0.032]		[0.000]		[0.000]
EK	2.075	−1.765	1.689	−1.227	2.719	−2.454
[P-value]		[0.478]		[0.389]		[0.165]
Observations	642		1,323		2,008	
			Panel D. 20%+ Coverage (Imputing Missing Outcomes)			
TK	−0.446	0.739***	−0.408	0.752***	−0.444	0.730***
[P-value]		[0.001]		[0.000]		[0.000]
EK	0.610	−1.131**	0.540	−1.025***	0.642	−1.183***
[P-value]		[0.023]		[0.006]		[0.000]
Observations	2,853		5,816		8,915	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Across the columns, we re-estimate models using either a 30-, 60-, or 90-day bandwidth. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

Table F12. Impacts on KRA Scores (Language and Literacy)

	30-Day Bandwidth		60-Day Bandwidth		90-Day Bandwidth	
	Complier Control Mean	LATE	Complier Control Mean	LATE	Complier Control Mean	LATE
	Panel A. 20%+ Coverage					
TK	−0.231	0.915**	−0.036	0.781***	0.037	0.658***
[P-value]		[0.018]		[0.001]		[0.000]
EK	1.979	−2.079*	2.043	−2.088***	1.734	−1.951***
[P-value]		[0.067]		[0.002]		[0.001]
Observations	1,943		3,912		5,988	
	Panel B. 80%+ Coverage					
TK	−0.778	1.350*	−0.221	0.921***	−0.160	0.845***
[P-value]		[0.097]		[0.007]		[0.000]
EK	1.634	−1.858	1.872	−1.896	1.515	−1.623
[P-value]		[0.571]		[0.184]		[0.215]
Observations	1,081		2,248		3,421	
	Panel C. 90%+ Coverage					
TK	−0.346	0.882*	0.231	0.459*	0.060	0.615***
[P-value]		[0.050]		[0.062]		[0.000]
EK	1.932	−1.942	1.910	−1.691	1.475	−1.156
[P-value]		[0.575]		[0.230]		[0.430]
Observations	642		1,323		2,008	
	Panel D. 20%+ Coverage (Imputing Missing Outcomes)					
TK	-0.004	0.714***	0.067	0.681***	0.067	0.644***
[P-value]		[0.000]		[0.000]		[0.000]
EK	0.899	−1.091**	1.032	−1.224***	0.864	−1.148***
[P-value]		[0.027]		[0.001]		[0.000]
Observations	2,853		5,816		8,915	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Across the columns, we re-estimate models using either a 30-, 60-, or 90-day bandwidth. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

Table F13. Impacts on KRA Scores (Mathematics)

	30-Day Bandwidth		60-Day Bandwidth		90-Day Bandwidth	
	Complier Control Mean	LATE	Complier Control Mean	LATE	Complier Control Mean	LATE
			Panel A. 20%+ Coverage			
TK	−0.108	0.638*	0.014	0.658***	−0.053	0.606***
[P-value]		[0.076]		[0.002]		[0.000]
EK	0.924	−0.848	1.312	−1.285**	1.324	−1.305***
[P-value]		[0.299]		[0.022]		[0.005]
Observations	1,943		3,912		5,988	
			Panel B. 80%+ Coverage			
TK	−0.681	1.050	−0.130	0.749**	−0.192	0.669***
[P-value]		[0.101]		[0.015]		[0.001]
EK	0.884	−0.812	1.203	−1.039	1.098	−1.081
[P-value]		[0.712]		[0.328]		[0.309]
Observations	1,081		2,248		3,421	
			Panel C. 90%+ Coverage			
TK	−0.538	0.875*	0.142	0.472*	−0.081	0.509**
[P-value]		[0.091]		[0.098]		[0.011]
EK	2.238	−2.748	1.386	−1.564	1.265	−1.291
[P-value]		[0.395]		[0.238]		[0.321]
Observations	642		1,323		2,008	
			Panel D. 20%+ Coverage (Imputing Missing Outcomes)			
TK	−0.003	0.596***	0.067	0.608***	−0.010	0.608***
[P-value]		[0.001]		[0.000]		[0.000]
EK	0.451	−0.585	0.610	−0.793**	0.638	−0.822***
[P-value]		[0.134]		[0.014]		[0.001]
Observations	2,853		5,816		8,915	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Across the columns, we re-estimate models using either a 30-, 60-, or 90-day bandwidth. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

Table F14. Impacts on KRA Scores (Physical and Motor Development)

	30-Day Bandwidth		60-Day Bandwidth		90-Day Bandwidth	
	Complier Control Mean	LATE	Complier Control Mean	LATE	Complier Control Mean	LATE
	Panel A. 20%+ Coverage					
TK	−0.394	0.544	−0.111	0.492**	−0.297	0.613***
[P-value]		[0.219]		[0.049]		[0.000]
EK	1.130	−1.645	0.321	−0.972	0.563	−1.156**
[P-value]		[0.173]		[0.117]		[0.021]
Observations	1,943		3,912		5,988	
	Panel B. 80%+ Coverage					
TK	−1.327	1.476*	−0.339	0.775**	−0.613	0.934***
[P-value]		[0.056]		[0.023]		[0.000]
EK	2.777	−3.223	0.248	−0.856	1.708	−2.106
[P-value]		[0.251]		[0.494]		[0.138]
Observations	1,081		2,248		3,421	
	Panel C. 90%+ Coverage					
TK	−0.634	0.917	−0.113	0.645**	−0.470	0.857***
[P-value]		[0.104]		[0.038]		[0.000]
EK	1.487	−1.340	−0.011	−0.084	2.240	−2.122
[P-value]		[0.592]		[0.957]		[0.252]
Observations	642		1,323		2,008	
	Panel D. 20%+ Coverage (Imputing Missing Outcomes)					
TK	−0.282	0.554***	−0.145	0.543***	−0.297	0.649***
[P-value]		[0.008]		[0.000]		[0.000]
EK	0.332	−0.851*	0.030	−0.652*	0.198	−0.771***
[P-value]		[0.094]		[0.058]		[0.005]
Observations	2,853		5,816		8,915	

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Across the columns, we re-estimate models using either a 30-, 60-, or 90-day bandwidth. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff.

Table F15. Impacts of TK and EK on KRA Scores, With and Without Covariates (20%+ Coverage)

	Overall	Social Foundations	Language and Literacy	Math	Physical/Motor Development
	Panel A. With Covariates				
$LATE_{TK}$	0.914**	0.934**	0.915**	0.638*	0.544
[P-value]	[0.027]	[0.037]	[0.018]	[0.076]	[0.219]
$LATE_{EK}$	−1.841*	−2.127*	−2.079*	−0.848	−1.645
[P-value]	[0.079]	[0.072]	[0.067]	[0.299]	[0.173]
	Panel B. With Covariates				
$LATE_{TK}$	0.860**	0.865*	0.878**	0.609	0.480
[P-value]	[0.049]	[0.063]	[0.028]	[0.112]	[0.290]
$LATE_{EK}$	−1.825	−2.134*	−2.073*	−0.819	−1.636
[P-value]	[0.115]	[0.078]	[0.091]	[0.406]	[0.176]
TK-complier control mean	−0.360	−0.674	−0.231	−0.108	−0.394
EK-complier control mean	1.628	1.585	1.979	0.924	1.130
Observations	1,943	1,943	1,943	1,943	1,943

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff. When included, the covariates are sex, race, and economic disadvantage status. District fixed effects are always included.

Table F16. Impacts of TK and EK on KRA Scores, Filling in Missing Outcomes With Extreme Values to Generate Conservative Estimates

	Overall	Social Foundations	Language and Literacy	Math	Physical/Motor Development
	Panel B. 80%+ Coverage				
$LATE_{TK}$	0.379	0.508	0.266	0.161	0.271
[P-value]	[0.398]	[0.264]	[0.605]	[0.712]	[0.532]
$LATE_{EK}$	−0.440	−1.425	0.097	0.559	−0.820
[P-value]	[0.663]	[0.160]	[0.939]	[0.554]	[0.398]
Observations	1,168	1,168	1,168	1,168	1,168
	Panel C. 90%+ Coverage				
$LATE_{TK}$	0.233	0.232	0.230	0.262	0.271
[P-value]	[0.572]	[0.578]	[0.570]	[0.592]	[0.595]
$LATE_{EK}$	−2.405	−1.235	−1.998	−2.691	−1.029
[P-value]	[0.256]	[0.559]	[0.415]	[0.258]	[0.591]
Observations	686	686	686	686	686

*** p<0.01, ** p<0.05, * p<0.1

Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Before estimating the models, we fill in missing outcome data with extreme values designed to make the EK and TK LATEs conservative. In non-TK districts, we replace missing values to the left (right) of the cutoff with scores at the 95th (5th) percentile of the distribution in non-TK districts (within a 30-day bandwidth). In TK districts, we replace missing values to the left (right) of the cutoff with scores at the 5th (95th) percentile of the distribution in TK districts (within a 30-day bandwidth). We do this separately for each of the 20%+, 80%+, and 90%+ samples.

Table F17. Impacts of TK and EK on Kindergarten and Third-Grade Test Scores, in Districts with 20%+ KRA Coverage

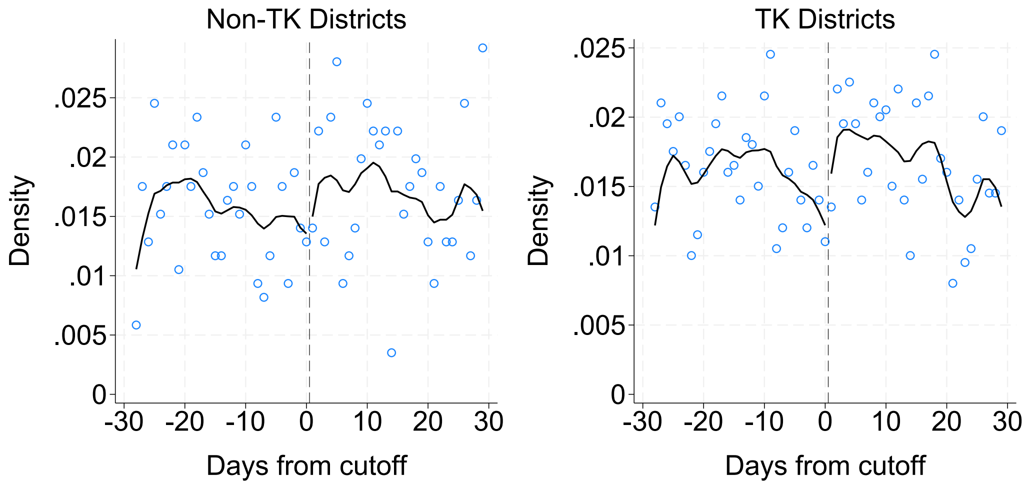
	Students with KRA Scores			Students with KRA and G3 Scores		Students in KRA Districts With G3 Scores	
	KRA Overall (1)	KRA Math (2)	KRA LL (3)	3rd Grade Math (4)	3rd Grade ELA (5)	3rd Grade Math (6)	3rd Grade ELA (7)
	Panel A. 30-Day Bandwidth						
$LATE_{TK}$	0.914**	0.638*	0.915**	0.308	0.317	0.259	0.185
[P-value]	[0.027]	[0.076]	[0.018]	[0.228]	[0.376]	[0.147]	[0.420]
$LATE_{EK}$	−1.841*	−0.848	−2.079*	−0.703	0.762	−0.184	0.678
[P-value]	[0.079]	[0.299]	[0.067]	[0.322]	[0.285]	[0.730]	[0.159]
TK-complier control mean	−0.360	−0.108	−0.231	0.249	0.284	0.297	0.416
EK-complier control mean	1.628	0.924	1.979	1.020	−0.481	0.501	−0.397
Observations	1,943	1,943	1,943	1,719	1,722	2,470	2,472
	Panel B. 90-Day Bandwidth						
$LATE_{TK}$	0.771***	0.606***	0.658***	0.316**	0.272*	0.220**	0.175
[P-value]	[0.000]	[0.000]	[0.000]	[0.013]	[0.067]	[0.026]	[0.121]
$LATE_{EK}$	−1.811***	−1.305***	−1.951***	−0.993**	−0.402	−0.477*	−0.182
[P-value]	[0.001]	[0.005]	[0.001]	[0.028]	[0.404]	[0.091]	[0.527]
TK-complier control mean	−0.191	−0.053	0.037	0.191	0.182	0.287	0.279
EK-complier control mean	1.454	1.324	1.735	1.274	0.639	0.759	0.419
Observations	5,988	5,988	5,988	5,258	5,266	7,613	7,614

*** p<0.01, ** p<0.05, * p<0.1

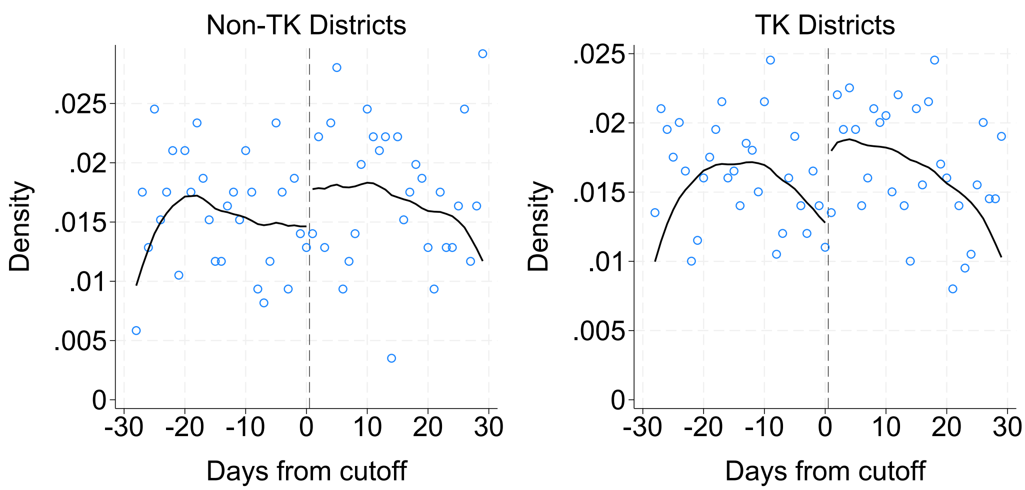
Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. We standardize KRA scale scores using means and standard deviations from the universe of KRA test-takers in Ohio in fall 2018 and fall 2019 (Appendix Table F1). Complier control means are estimated by subtracting impact estimates from observed outcomes at the RD cutoff. The first columns restate the impact estimates from Tables F10, F13, and F12. The middle columns estimate impacts on 3rd grade test scores among students in the 20%+ coverage sample who have KRA scores *and* 3rd grade test scores. The final columns estimate impacts for all students in the 20%+ coverage sample who have 3rd grade test scores *regardless* of whether they also have KRA scores in kindergarten.

Figure F1. Density by Birthdate in the KRA Sample

(a) Panel A. Five-Day Bandwidth

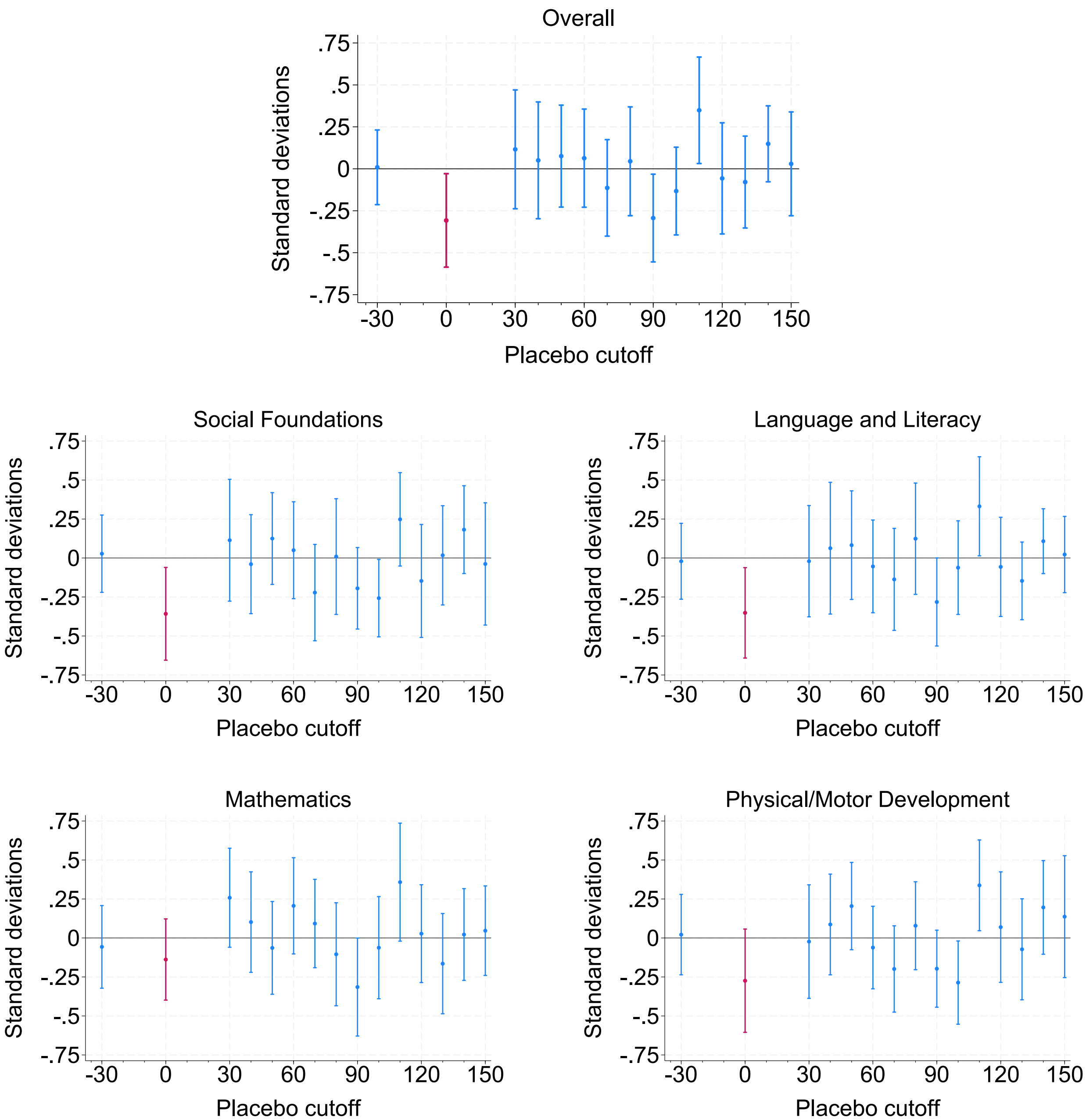


(b) Panel B. Ten-Day Bandwidth



Notes: Each dot gives the density of students born on a particular day. The smoothed lines in each panel are estimated using the McCrary (2008) density test. "Days from cutoff" gives the number of days between a child's birthday and the TK/EK cutoff date (December 1).

Figure F2. ITT Estimates for KRA Scores at Placebo Cutoffs in Non-TK Districts



Note: Each estimate in these plots comes from an RD regression using the associated x-axis value as the (potentially placebo) birthday cutoff. The vertical bars indicate 95% confidence intervals. Our primary estimates using the true cutoff (0) are shown in red. We do not estimate regressions with placebo cutoffs within 30 days of 0 so that outcomes near the true cutoff do not influence the placebo estimates. We do not estimate regressions with placebo cutoffs less than −30 to avoid discontinuities associated with the kindergarten birthday cutoff. For each placebo model, we limit the sample to observations within 30 points of the placebo cutoff. Standard errors are clustered at the running variable level.

Appendix G Robustness Checks

G.1 Models With and Without Covariates

Table G1 shows that our results are robust to the inclusion or exclusion of covariates in the impact models. The differences between the estimates are small and statistically insignificant.

In the "relaxed assumptions" approach, we always exclude covariates when estimating EK LATEs because the demographic subgroups are defined by the covariates. Hence, the estimates of the EK LATEs are identical, by construction, for the with- and without-covariate results shown in the table. The TK LATEs, however, do change slightly when we exclude covariates from the estimation of the other pieces involved in backing out the TK LATE (i.e., ITT, Ω_{TK}, and Ω_{EK}). The estimates for math and ELA both increase slightly, but the differences are not statistically significant. Overall, the inclusion of covariates hardly matters for our estimates.

Table G1. 3rd Grade Test Score Impacts With and Without Covariates

	Math		ELA	
	Baseline	Relaxed Assumptions	Baseline	Relaxed Assumptions
	Panel A. With Covariates			
$LATE_{TK}$	0.212*	0.294	0.097	0.191
[P-value]	[0.051]	[0.111]	[0.401]	[0.293]
$LATE_{EK}$	−0.366***	−0.557*	−0.219*	−0.435
[P-value]	[0.000]	[0.092]	[0.078]	[0.181]
	Panel B. Without Covariates			
$LATE_{TK}$	0.252**	0.331*	0.123	0.209
[P-value]	[0.046]	[0.088]	[0.321]	[0.253]
$LATE_{EK}$	−0.378***	−0.557*	−0.240*	−0.435
[P-value]	[0.000]	[0.092]	[0.061]	[0.181]
TK-complier control mean	0.219		0.310	
EK-complier control mean	0.609		0.447	
Observations	15,680		15,669	

Note: These estimates are obtained via RD models as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. In the "relaxed assumptions" approach, we always exclude controls when estimating EK LATEs because the demographic subgroups are defined by the covariates. Hence, the EK estimates in the with- and without-controls columns are identical by construction. Complier control means are estimated by subtracting (baseline approach) impact estimates from observed outcomes at the RD cutoff. When included, the covariates are sex, race, and economic disadvantage status. District-cohort fixed effects are always included.

Table G2. Impacts of TK and EK on Special Education Placement, With and Without Covariates (Baseline Estimation Approach)

	SPED in K-G2	SPED in K-G3	SPED in K	SPED in G1	SPED in G2	SPED in G3
	Panel A. With Covariates					
$LATE_{TK}$	0.028	−0.013	0.097***	0.032	−0.007	−0.070
[P-value]	[0.455]	[0.787]	[0.000]	[0.270]	[0.856]	[0.106]
$LATE_{EK}$	0.016	0.018	0.009	0.029	0.001	−0.003
[P-value]	[0.714]	[0.730]	[0.786]	[0.415]	[0.978]	[0.941]
	Panel B. Without Covariates					
$LATE_{TK}$	0.030	−0.012	0.098	0.034	−0.006	−0.070
[P-value]	[0.417]	[0.783]	[0.002]	[0.276]	[0.885]	[0.136]
$LATE_{EK}$	0.017	0.018	0.010	0.029	0.002	−0.002
[P-value]	[0.708]	[0.719]	[0.782]	[0.446]	[0.965]	[0.963]
TK-complier control mean	0.101	0.144	−0.001	0.064	0.109	0.151
EK-complier control mean	0.043	0.057	0.030	0.014	0.031	0.058
Observations	15,704	15,704	15,530	15,614	15,589	15,704

*** p<0.01, ** p<0.05, * p<0.1

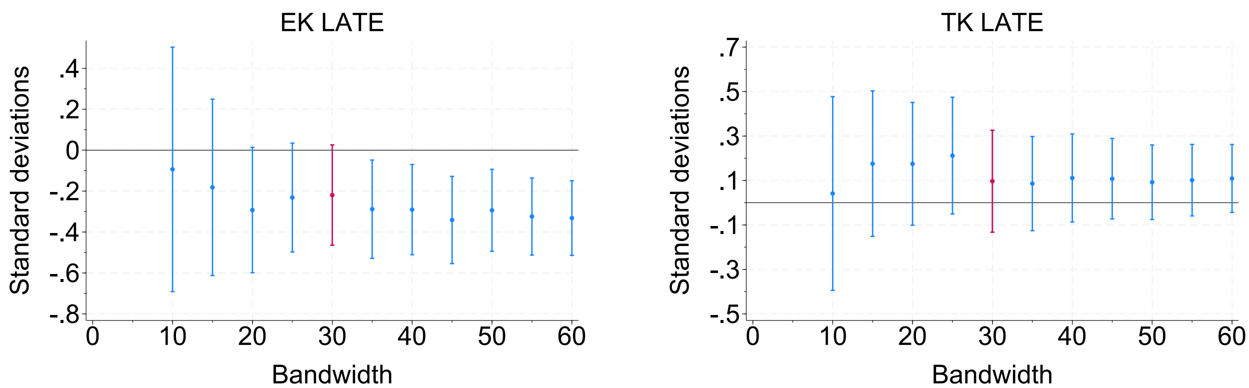
Note: These estimates are obtained via RD models (baseline approach) as described in Section 8. Inference is conducted via bootstrap, with clustering on the running variable. Complier control means are estimated by subtracting with-covariates impact estimates from observed outcomes at the RD cutoff. As a reference point, 9.6% of TK students at the birthday cutoff are in special education while in TK. When included, the covariates are sex, race, and economic disadvantage status. District-cohort fixed effects are always included.

G.2 Alternative Bandwidths

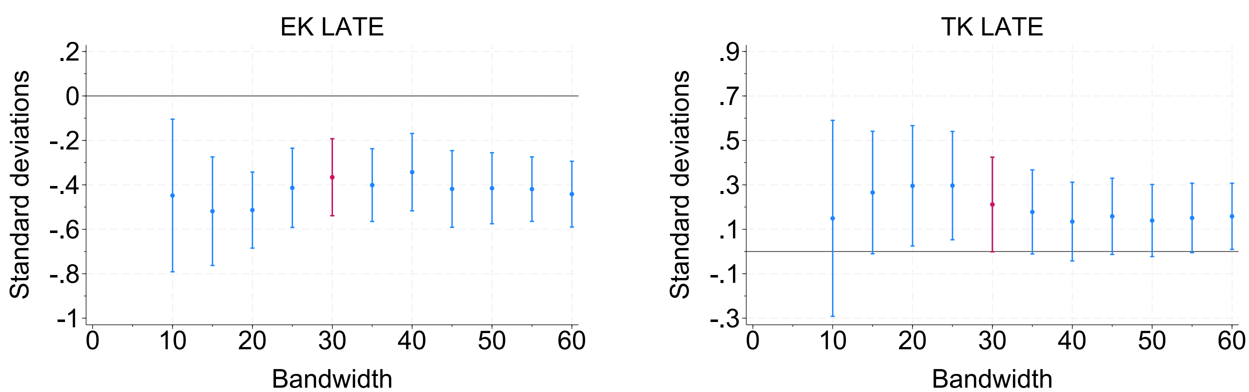
Tables G1, G2, and G3 show how our third-grade impact estimates change as the choice of bandwidth changes. For bandwidths ranging from 10 to 60 days, our estimates vary somewhat quantitatively but are highly similar qualitatively. The one exception is that estimates coming from the smallest bandwidths are sometimes qualitatively different than the other estimates. However, these estimates are highly imprecise as they come from the smallest samples.

Figure G1. Robustness of 3rd Grade Test Score Impacts to Alternative Bandwidths

(a) 3rd Grade ELA Scores



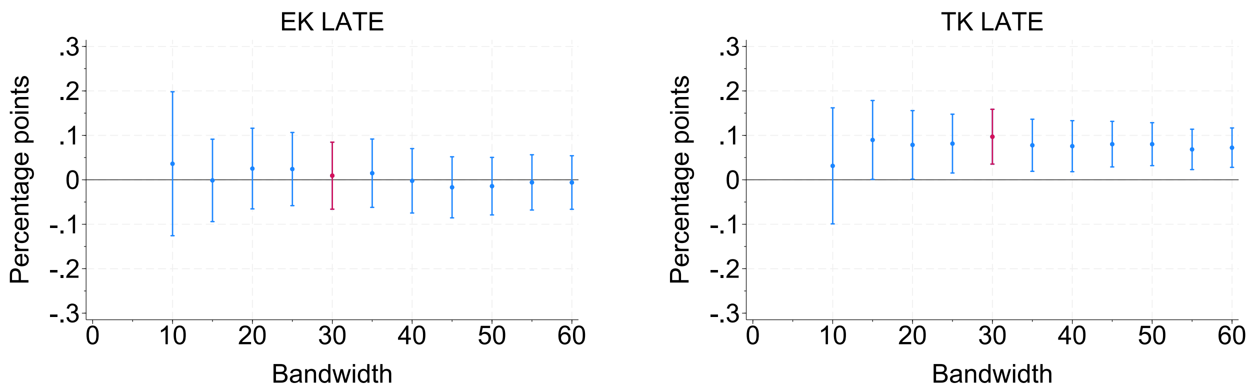
(b) 3rd Grade Math Scores



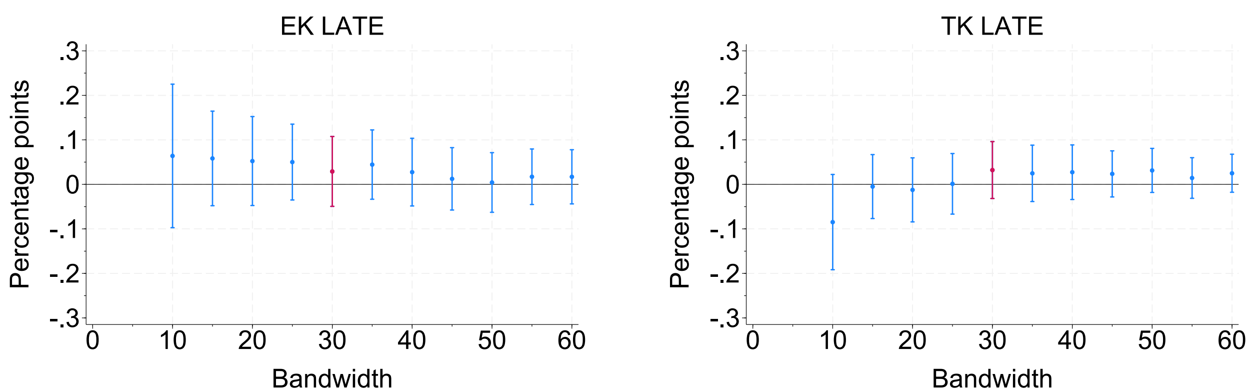
Note: Each estimate in these plots comes from an RD regression using an alternative bandwidth. For each value X on the x-axis, we re-estimate impacts using the baseline estimation procedure described in Section 8, except that we impose a bandwidth of (-X,X). Standard errors are clustered at the running variable level. The vertical bars indicate 95% confidence intervals. Our primary estimates, which use a ±30-day bandwidth, are shown in red.

Figure G2. Robustness of SPED Impacts to Alternative Bandwidths I

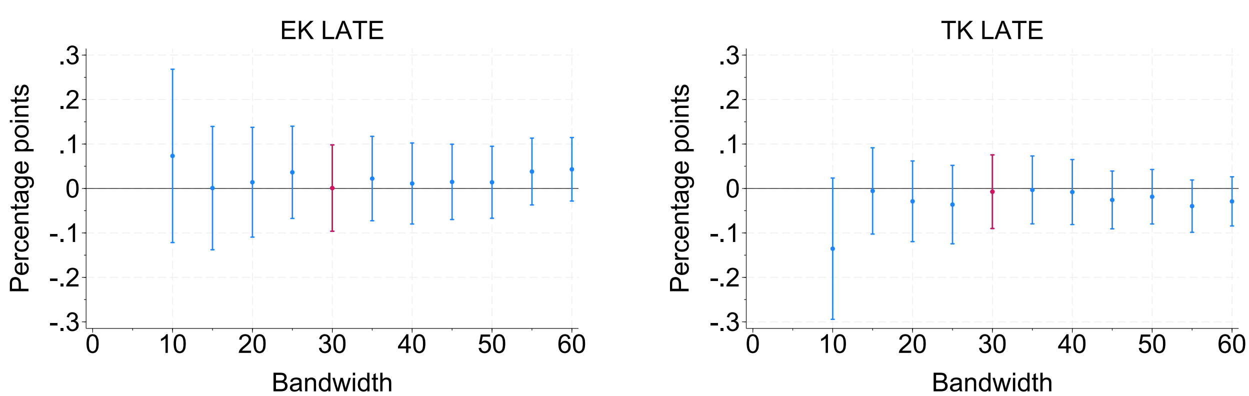
(a) Special Education in Kindergarten



(b) Special Education in 1st Grade



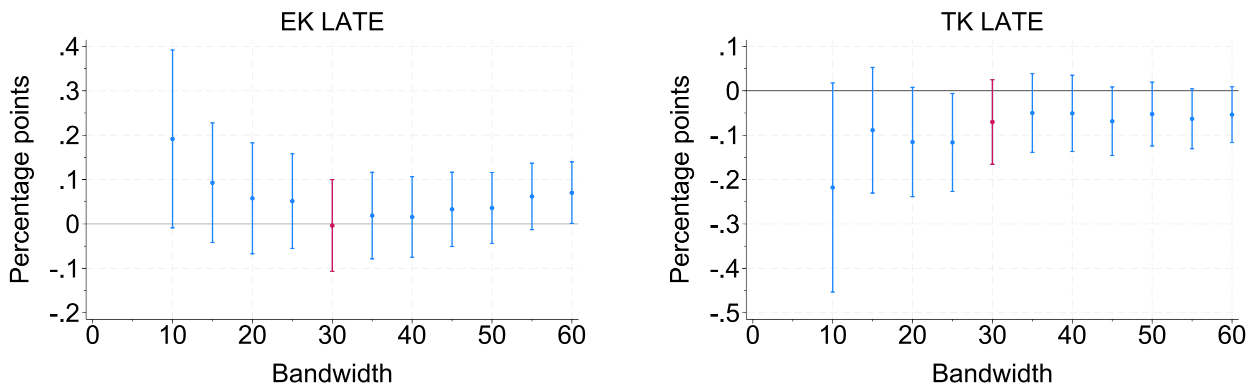
(c) Special Education in 2nd Grade



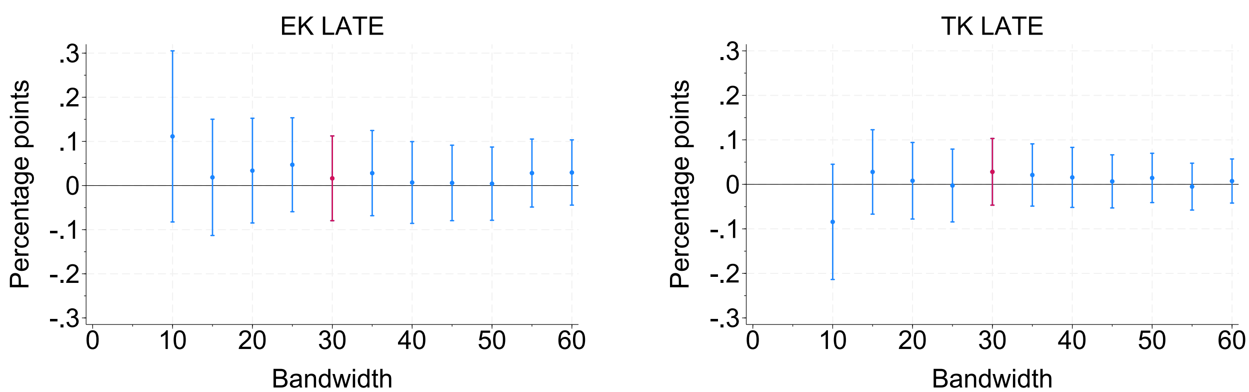
Note: Each estimate in these plots comes from an RD regression using an alternative bandwidth. For each value X on the x-axis, we re-estimate impacts using the baseline estimation procedure described in Section 8, except that we impose a bandwidth of (-X,X). Standard errors are clustered at the running variable level. The vertical bars indicate 95% confidence intervals. Our primary estimates, which use a ±30-day bandwidth, are shown in red.

Figure G3. Robustness of SPED Impacts to Alternative Bandwidths II

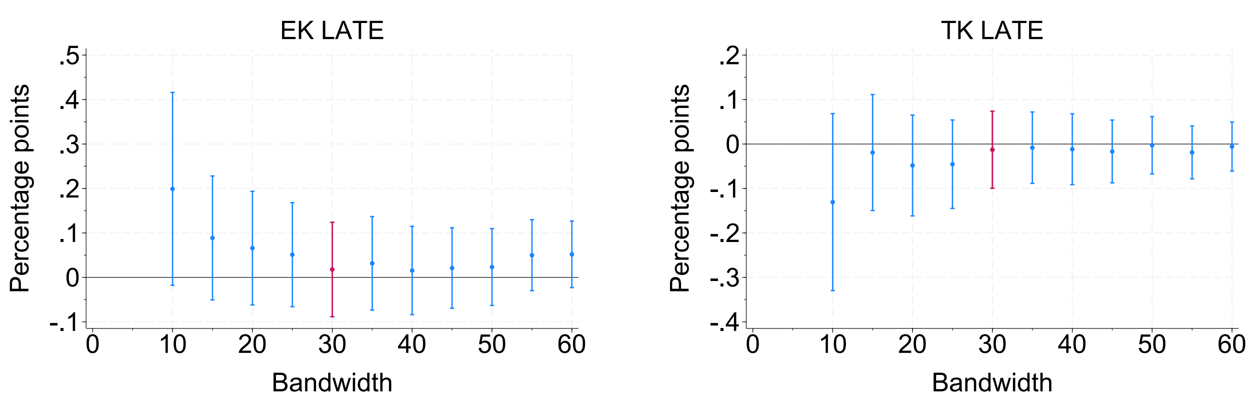
(a) Special Education in 3rd Grade



(b) Special Education in K, G1, or G2



(c) Special Education in K, G1, G2, or G3



Note: Each estimate in these plots comes from an RD regression using an alternative bandwidth. For each value X on the x-axis, we re-estimate impacts using the baseline estimation procedure described in Section 8, except that we impose a bandwidth of (-X,X). Standard errors are clustered at the running variable level. The vertical bars indicate 95% confidence intervals. Our primary estimates, which use a ±30-day bandwidth, are shown in red.

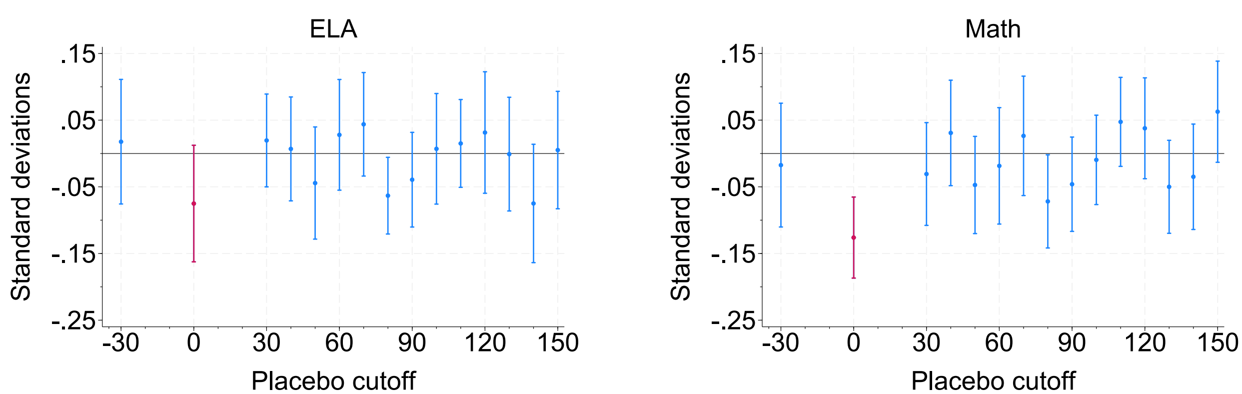
G.3 Placebo Cutoffs

Here, we conduct placebo checks to gauge the likelihood of finding non-zero impacts simply by chance. We do this using our baseline estimation strategy and re-estimating models at false (i.e., placebo) birthday cutoffs. Small and statistically insignificant impact estimates at the placebo cutoffs would suggest that we are unlikely to estimate large effects the true cutoff if there is not, in fact, an impact.

Placebo checks are less informative in TK districts because an ITT effect may be zero because EK and TK effects offset each other. For this reason, we focus on ITT effects in non-TK districts. We also focus on placebo cutoffs more than 30 days away from the true December 1 cutoff so that outcomes near the true cutoff do not influence the placebo estimates. Finally, we do not use placebo cutoffs less than -30 to avoid discontinuities associated with the birthday cutoff for regular kindergarten.

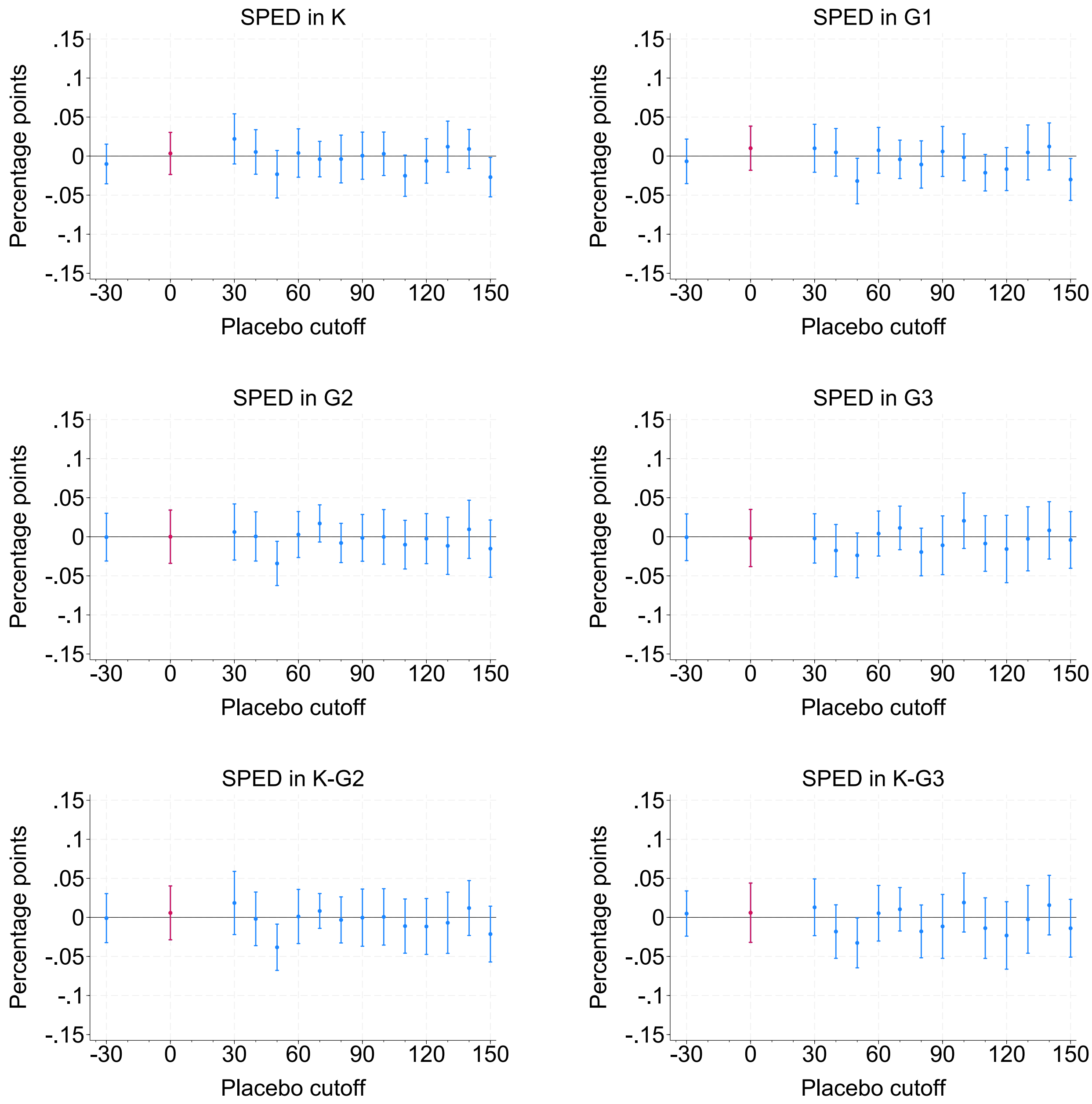
Figures G4 and G5 show our results. Reassuringly, we find find null effects at nearly every placebo cutoff. The special education impacts are also null at the true cutoff, consistent with the LATE estimates reported in Table 5. On the other hand, for math and reading test scores, the estimates at the true cutoff are extreme relative to the placebo estimates—lending credibility to their internal validity.

Figure G4. ITT Estimates for 3rd Grade Test Scores at Placebo Cutoffs in Non-TK Districts



Note: Each estimate in these plots comes from an RD regression using the associated x-axis value as the (potentially placebo) birthday cutoff. The vertical bars indicate 95% confidence intervals. Our primary estimates using the true cutoff (0) are shown in red. We do not estimate regressions with placebo cutoffs within 30 days of 0 so that outcomes near the true cutoff do not influence the placebo estimates. We do not estimate regressions with placebo cutoffs less than -30 to avoid discontinuities associated with the kindergarten birthday cutoff. For each placebo model, we limit the sample to observations within 30 points of the placebo cutoff. Standard errors are clustered at the running variable level.

Figure G5. ITT Estimates for SPED Outcomes at Placebo Cutoffs in Non-TK Districts



Note: Each estimate in these plots comes from an RD regression using the associated x-axis value as the (potentially placebo) birthday cutoff. The vertical bars indicate 95% confidence intervals. Our primary estimates using the true cutoff (0) are shown in red. We do not estimate regressions with placebo cutoffs within 30 days of 0 so that outcomes near the true cutoff do not influence the placebo estimates. We do not estimate regressions with placebo cutoffs less than −30 to avoid discontinuities associated with the kindergarten birthday cutoff. For each placebo model, we limit the sample to observations within 30 points of the placebo cutoff. Standard errors are clustered at the running variable level.

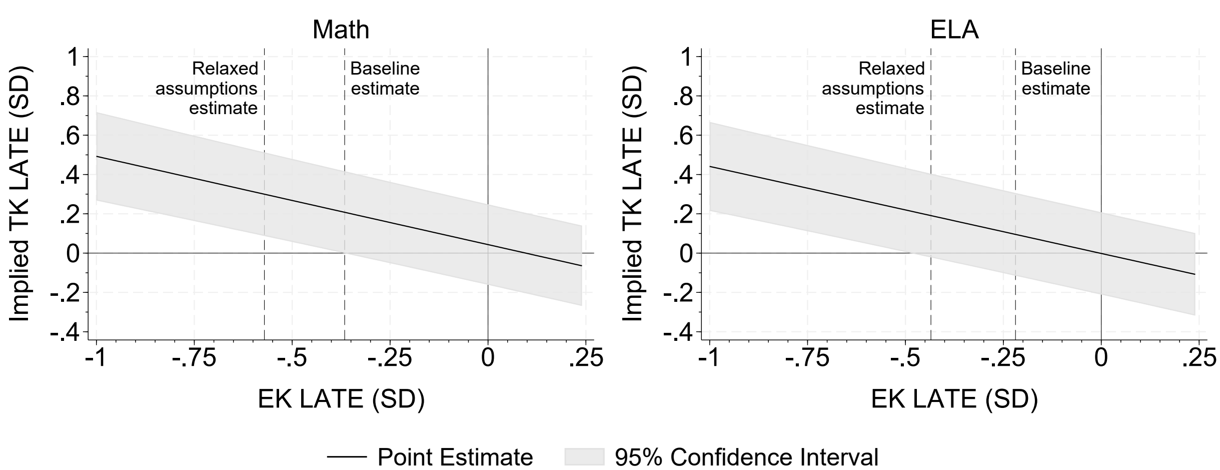
G.4 Sensitivity Analysis

Relative to the baseline approach, our "relaxed assumptions" approach relaxes the assumption of EK treatment effect homogeneity. Specifically, the approach assumes EK treatment effects may differ *across* the eight demographic groups, but *not* within demographic group across TK and non-TK districts. For example, it assumes the treatment effect of EK for white or Asian females who are not economically disadavantaged is identical in TK and non-TK districts.

In this section, we take a different approach to investigating the robustness of our TK LATE estimates. We explore how large or small the true EK LATEs would have to be to imply substantially different TK LATE estimates. Figure G6 shows the implied $LATE_{TK}$ estimate for every value of $LATE_{EK}$ ranging from -1.0 to 0.25 using Equation 1 from the paper and our primary estimates of ITT, Ω_{TK}, and Ω_{EK}.

For both outcome domains, we estimate a larger TK LATE when the EK LATE is larger. Because the intent-to-treat discontinuity is slightly larger for math than ELA, the implied TK LATE is also slightly larger for math for any given EK LATE. For us to estimate that the TK LATEs for math and ELA are roughly 0, the true EK LATEs would have to be as small as 0.08 and -0.02 standard deviations, respectively. The math TK LATE would lose statistical significance at the 95% confidence level if the true EK LATE is less than -0.36 standard deviations, which is approximately our baseline EK LATE estimate. The EK LATE for ELA would have to be more negative than -0.5 standard deviations for us to estimate a statistically significant positive TK LATE for ELA.

Figure G6. Range of Possible TK Test Score Impacts



Note: This figure plots the implied TK LATE estimates for given values of the EK LATE. To obtain the estimates, we use Equation 1 from the paper and our primary estimates of ITT, Ω_{TK}, and Ω_{EK}.

G.5 Alternative Identification Assumptions

Disentangling the TK and EK treatment effects requires some restriction on treatment effect heterogeneity. For instance, our baseline approach assumes the EK LATE is the same in TK and non-TK districts; our "relaxed assumptions" approach assumes the EK LATE is the same in TK and non-TK districts, but only within demographic groups. In this section, we discuss a third assumption that, in theory, would also allow us to disentangle the two treatment effects. In practice, however, this approach is uninformative for us.

The third potential assumption is that the EK impact is homogeneous across demographic groups, although it may differ across districts with and without TK. For example, with demographic groups defined based on sex, the assumption would be that boys and girls in TK districts have the same EK LATE. We view this assumption as complementary to our "relaxed assumptions" approach; each assumption relaxes treatment effect homogeneity in one dimension while enforcing it in another. Our "relaxed assumptions" approach relaxes homogeneity across student type, whereas this approach relaxes homogeneity across district type. Caetano et al. (2023) develop an identification argument and estimation techniques using this assumption.

Unfortunately, the third approach was not informative in our setting. Using the Caetano et al. (2023) estimator and only data from TK districts, in various specifications we defined demographic groups based on one of sex, race, economic disadvantage status, and cohort. The resulting estimates were too noisy for us to draw any conclusions.

Another student characteristic we considered using within the Caetano et al. (2023) framework was distance from one's neighborhood to the nearest in-district school that offers TK. As in other settings, distance likely affects program take-up and plausibly does not separately influence academic outcomes. However, two issues prevented us from using distance with this approach. The first issue was power. In around half of all TK districts, every school with kindergarten students also has a TK program. In these districts, distance would not cause differential take-up of TK and EK. The second issue was the non-random placement of TK programs. In the other half of TK districts—the ones that offer TK in some but not every building with kindergartners—TK programs are more likely to be in schools that serve more economically disadvantaged children. They are also more likely to be placed in elementary schools that have (non-TK) Pre-K programs or whose highest grade is not higher than 3rd grade. These observable differences make it less likely that a homogeneous treatment effect assumption would hold between students who live different distances from TK programs.

Appendix H Inference Strategy

We conduct inference via bootstrap because our "relaxed assumptions" approach to identification requires a multi-step estimation procedure. For consistency, we conduct bootstrap inference in our baseline approach too, although we get nearly identical results using the standard parametric approach.

Specifically, we implement a "Bayesian bootstrap" that creates new samples by reweighting rather than resampling. The procedure is stratified across TK and non-TK districts and clustered on the RD running variable (i.e., birthday). All observations within a cluster share a single replication weight, drawn randomly from an exponential distribution. The replication weights are normalized so that the sum of each cluster weight equals the number of clusters in a strata. We draw 1,000 sets of weights.

In the baseline approach, we estimate our two-stage least squares model with bootstrap weights 1,000 times. In the "relaxed assumptions" approach, we re-estimate every part of the multi-step procedure 1,000 times. Doing a full bootstrap accounts for uncertainty in the 8 subgroup EK LATEs in non-TK districts; the 8 EK complier shares in TK districts; and the ITT, Ω_{TK}, and Ω_{EK} in TK districts.

Throughout the paper, we summarize our inference results using p-values rather than standard errors. We omit standard errors because they are uninformative in our "relaxed assumptions" approach. The bootstrap distributions are highly non-normal in our "relaxed assumptions" approach in that they contain some extreme outliers. These outliers likely exist because we split the sample into small subgroups and estimate a large number of parameters, which creates several opportunities for sampling variation to produce extreme outcomes. Consequently, the outliers drive up the TK and EK LATE standard errors, making them uninformative about variation throughout most of the distributions.

Instead of standard errors, we present p-values from two-tailed hypothesis tests. We calculate p-values in two steps. First, for each estimate, we enforce a null hypothesis that there is no effect by subtracting the mean of the bootstrap distribution from each bootstrap estimate. Second, we calculate the share of demeaned estimates that are greater (in absolute value) than our primary point estimate. This share is the p-value.

Appendix I Counterfactual Enrollment in Michigan

In this appendix, we provide more detail on our construction of program enrollment statistics for Michigan 4-year-olds (Table 7). Our methodology draws from several sources, combining publicly available data with administrative student records from the Michigan Education Research Institute (MERI). As necessary, we make data-driven assumptions to fill in gaps in the data. Ultimately, our methodology produces enrollment estimates for every major category of licensed child care in the state.

As with all estimates, ours are inherently uncertain. The first source of uncertainty is measurement error. All the underlying data we draw from contain some measurement error due to misreporting, data entry mistakes, etc. The second source of uncertainty stems from assumptions we make for some care types due to data limitations. The TK and "other licensed child care" estimates are more uncertain than the GSRP and Head Start estimates because they require more assumptions to estimate enrollment.

Generally, our strategy is to estimate the share of students who enroll in the following early learning programs during their Pre-K year: 1) Transitional Kindergarten, 2) Kindergarten (with a waiver), 3) Michigan's state-funded Pre-K program (called GSRP), 4) Head Start, and 5) other licensed child care programs. To align with our primary analysis, we focus on the 2014-2015 and 2018-2019 Pre-K cohorts and districts that offered TK programs in those years and that have reliable individual-level data on TK enrollment. Below, we describe how we estimate enrollment in each program for these districts.

Number of Children by District. We begin by estimating the number of children in each district-year. Using the Michigan administrative student data, for every district, we identify all children enrolled in kindergarten (in any year) whose birthdays put them in the 2014-2015 and 2018-2019 Pre-K cohorts. Then, we inflate the number of children in each district-year by 10/9 to account for children unobserved in the data because they enroll in private school.

Transitional Kindergarten and Early Kindergarten Entry. Given the dataset we constructed for our main analysis, estimating TK and EK enrollment is straightforward. We simply count the number of TK and EK students in each of the relevant districts in 2014-2015 and 2018-2019.

Great Start School Readiness Program (GSRP). Estimating GSRP enrollment in each district-year is also straightforward. The Michigan administrative data records whether each child enrolled in a GSRP program in a given year, including whether the program was a straight GSRP program or a GSRP/Head Start blend program. To assign enrollment to a particular district, we use the district associated with a child's primary residential address.

Head Start. Estimating Head Start enrollment by district requires greater imputation than the previous programs. Our primary sources are the 2014-2015 and 2018-2019 Head Start Program Information Reports (PIRs) and the Head Start Center Locator (an online tool run by the U.S. Department of Health and Human Services, accessed in May 2023). The PIR data contain information on enrollment for each program, and the Center Locator contains the zip code of each Head Start Center in Michigan. Note that in Head Start terminology, the "program" level is above the "center" level. Each program may operate several centers.

Beginning with the PIR data, we limit our focus to enrollment in regular Head Start programs, meaning we drop Early Head Start, which serves children younger than our focal age, and other programs like Migrant Head Start that account for a very small fraction of slots in Michigan.

We make one other adjustment to the PIR enrollment counts. For some programs, reported enrollment exceeds reported capacity. This may be partly due to measurement error, but it may also reflect a difference in the way enrollment and capacity are defined. Capacity is defined as a "point-in-time" measure, capturing how many children could enroll at any given time. On the other hand, enrollment includes all children who enroll over the course of a year. If there is enough enrollment turnover throughout the year, reported enrollment may exceed reported capacity. To reduce the influence of turnover on our enrollment estimates, we adjust enrollment down when it exceeds capacity.

Next, we merge the PIR enrollment counts onto the Head Start Locator Data. As previously mentioned, each program may operate several Head Start Centers in different locations. We assume enrollment is spread evenly across centers within a given program. Then, to estimate enrollment at the school district level, we crosswalk between center zip codes and districts. When a zip code spans more than one district, we allocate enrollment to districts based on the share of the zip code's land in each district.

Other licensed child care. Our best data on other licensed child care programs come from Michigan's Department of Licensing and Regulatory Affairs (LARA), which keeps records on all licensed child care providers in the state. Among the universe of child care providers, we restrict our attention to programs that serve 4-year-olds and are not Head Start or GSRP programs. The LARA data has information on capacity by zip code, but it does not contain any enrollment data.

To estimate enrollment, we incorporate information from the American Community Survey (ACS). Using the ACS, we calculate the number of four-year-olds in Michigan that enrolled in some form of preschool in 2015 and 2019. Then, we subtract from those totals the number of four-year-olds who enrolled in TK, kindergarten (with a waiver), GSRP, and

Head Start (across all of Michigan). The result is an estimate for 2015 and 2019 of the number of four-year-olds who enrolled in all other licensed care programs. We then allocate these state-level totals to the zip code level in proportion to our capacity estimates by zip code from the LARA data. Finally, we crosswalk between zip codes and school districts as we did for Head Start.

Some additional detail on the LARA data should be noted. Reported capacity is combined across every age a program serves, so we must make an assumption about the share of a program's capacity that is for four-year-olds. We assume capacity is allocated in proportion to the rates at which Michigan children of each age enroll in child care. For example, if a program serves four- and five-year-old children, and four-year-olds are twice as likely as five-year-olds to enroll in child care in Michigan, then we would allocate twice as much of a program's overall capacity to four-year-olds. We obtain relative rates of enrollment across ages from a publicly available table titled "FY 2020 Preliminary Data Table 9 - Average Monthly Percentages of Children In Care By Age Group," published by the Office of Child Care (a unit within the U.S. Department of Health and Human Services). The table combines enrollment for children ages 6 to 12, so we make a further assumption about relative enrollment rates within that age range. Another limitation of this table is that it only accounts for enrollment funded by the Child Care and Development Fund and so may not be representative of all child care enrollment. The distribution we ultimately use is:

Age	Assumed Share
0	5%
1	10%
2	13%
3	14%
4	13%
5	10%
6	6%
7	6%
8	5%
9	5%
10	5%
11	4%
12	4%

Combining the estimates. To convert our estimates into shares, we divide our estimate of the number of children enrolled in each program by our estimate of the total number of children, for each district-year. The difference between 1 and the sum of these shares gives our estimate for the share of children not enrolled in any program. These results are in Table I1. Each share is the weighted mean across districts, using the number of children in each district as weights.

Estimating counterfactuals. Finally, we use the estimates in Table I1 to estimate enrollment for children born just after the TK birthday cutoff. To do this, we assume the ratios of enrollment between GSRP, Head Start, other licensed child care programs, and no formal child care programs are the same regardless of TK eligibility. In other words, for programs other than TK and early kindergarten, we assume the ratio of the share of children who enroll in one program versus another is the same for children born just before and just after the TK cutoff. Given this assumption, we estimate enrollment in each program by scaling up the estimates in Table I1 to account for the fact that children born just after the cutoff are not eligible for TK and EK. Because we estimate that TK and EK constitute 13.5% of all program enrollment, we scale up enrollment in each of the other options by $1/(1 - 0.135)$. The results are in Table 7.

Table I1. Estimated Proportion of Michigan 4-Year-Olds Enrolled in Different Care Settings in 2015 and 2019

	Transitional Kindergarten	Early Kindergarten Entry	GSRP	Head Start	Other Licensed Child Care	Residual Care Arrangements
2015 Cohort	5%	5%	20%	10%	13%	46%
2019 Cohort	11%	6%	20%	8%	14%	42%
Both Cohorts	8%	5.5%	20%	9%	13.5%	44%

Note: This table presents program enrollment estimates for districts that offered TK in 2014-2015 and 2018-2019 and that reliably reported TK enrollment. The estimates combine population and enrollment data from a variety of sources, requiring substantial imputation. For more information on our estimation approach, see Appendix I. GSRP stands for the Great Start Readiness Program, which is Michigan's income-targeted state-funded Pre-K program.