Online Appendix (Rucker C. Johnson, 2024 *AEA Papers & Proceedings*)

***Empirical Approach & Econometric Model Specification*.**

As with other public preK programs across the country, California's TK program is voluntary for eligible children. This means that children enrolled in TK may differ from those whose parents choose other care setting options. An analysis that does not take into account the endogeneity of TK enrollment and the difference in children's age by design is likely to lead to biased results. I exploit quasi-experimental variation to avoid this endogeneity, based on the timing of the introduction and rollout of the TK program and the rules governing eligibility in the year the child turned 5, which used a strict December 2nd birthdate cutoff. Eligibility for TK is limited to students in a very specific age range, which means that a regression discontinuity (RD) design can be used to approximate the rigor and credibility of random assignment without actually randomly assigning children. The birthdate age eligibility cutoff enables one to use a regression discontinuity (RD) design, combined within a school-by-kindergarten cohort fixed effects model (framework), to compare outcomes for children from the same school who were part of the same kindergarten cohort, but who differed in their TK eligibility/attendance the year before kindergarten due to differences in age by only weeks (within 60 days on either side of the December 2 cutoff date to enter TK) – specifically, this involves comparisons of children who are just a little too young to be eligible for TK by virtue of turning 5 sometime after December 2 through January 31 vs children who are just old enough to be eligible for TK due to turning 5 between October 1 – December 2. This study takes advantage of this birthdate cutoff and limited age range, and it employs the RD design. The analysis focuses on students born between October 1 and January 31/February 2 (within 60 days on either side of the December 2 cutoff date to enter TK). I then compare the academic achievement outcomes in elementary school of students who attended TK with the corresponding achievement outcomes of those who did not, as determined by the birthdate cutoff. I focus on math and reading achievement, 3 and 4 years after kindergarten.

Let Y_{itsc} represent the student-level end-of-year outcome *t* years after kindergarten of student *i* from school *s* and kindergarten cohort *c*, including math and reading achievement, where *s* indexes kindergarten school attended, *c* indexes kindergarten cohort, *t* indexes years since kindergarten, *i* indexes student. All models include dummy indicator variables for child family socioeconomic disadvantage (SES_i), child gender and race/ethnicity ($demog_i$), and years since kindergarten fixed effects (η_t), years since kindergarten fixed effects interacted with child family SES ($\eta_t * SES_i$), and a complete set of school-by-kindergarten cohort fixed effects (θ_{sc}). All models are estimated separately for students whose parents are English speakers and for those whose parents are non-English speakers.

One of the endogenous treatment measures, $ppe_{d(s)ct}^{3yravg}$, is average per-pupil spending during the three immediately preceding years (*t*-2 through *t*) in district *d* and kindergarten cohort *c*. (e.g., for 4th grade achievement, this corresponds with spending in grades 2-4 (ages 7-9)). Average district per-pupil spending during school years *t*-2 through *t* is inflation-adjusted using the CPI-U deflator (in real 2015 dollars) and then expressed in thousands (and centered around 12). This is done to facilitate interpretation of marginal effects, so that the estimated effects are in the range we observe LCFF-induced variation in per-pupil spending, and so that the average TK treatment effects are evaluated at the average school spending (~$12,000) experienced for exposed cohorts.

The simulated instrumental variables for school funding reform (based on the funding formula, pre- & post-LCFF) are represented by $\widehat{Sim\Delta LCFF}_{d(s)ct(t-1)(t-2)(t-3)}$: LCFF-induced increase in funding from state (relative to pre-LCFF) for district *d*, kindergarten cohort *c*, at *t*, *t*-1, *t*-2, *t*-3 years since kindergarten (holding district socioeconomic disadvantage, property values, fixed at 2013 levels). The inclusion of school-by-kindergarten cohort fixed effects ensures that sources of identifying variation rely on comparisons of students from the same school and same kindergarten cohort across successive years after kindergarten, which exploits differences in the duration of school-age years of exposure and the intensity of dosage of reform-induced spending changes. The two-stage least squares instrumental variables (2SLS-IV) models isolate exogenous variation in LCFF-induced increases in per-pupil spending (experienced over multiple years) on student achievement in elementary school. The student outcomes include math and reading standardized achievement (NAEP-normed adjusted in grade-level equivalent units), three and four years after kindergarten. The analysis is restricted to students observed in kindergarten or earlier, and followed thereafter.

Formally, using the student-level longitudinal data, I estimate the following system of equations by 2SLS (where the first three equations comprise the 1st stage, and the fourth equation below represents the 2nd stage):

Model Specification.

(1) 1st stage (Prob(TK participation)):

$$\begin{aligned}\widehat{TKattend}_{isc} = {} & \alpha_{1,1}(TKelig_{isc'} * SES_i) + f(Birthdate_i) * SES_i + \pi_{1,1}\left(Sim\widehat{\Delta LCFF}_{d(s)c't'} * SES_i\right) \\ & + \pi_{2,1}\left(Sim\widehat{\Delta LCFF}_{d(s)c't'-1} * SES_i\right) + \pi_{3,1}\left(Sim\widehat{\Delta LCFF}_{d(s)c't'-2} * SES_i\right) \\ & + \pi_{4,1}\left(Sim\widehat{\Delta LCFF}_{d(s)c't'-3} * SES_i\right) \\ & + \pi_{5,1}\left(Sim\widehat{\Delta LCFF}_{d(s)c't'(t'-1)(t'-2)(t'-3)} * TKelig_{isc'} * SES_i\right) \\ & + \gamma_{1,1}SES_i + \gamma_{2,1}Demog_i + (\eta_{t,1} * SES_i) + \theta_{sc',1}\end{aligned}$$

where $Birthdate_i$ represents the student's birthdate centered at the TK age eligibility cutoff (December 2nd), and $f(Birthdate_i)$ is a flexible non-parametric (semi-parametric, polynomial function) specification for the birthdate running variable centered around the TK eligibility cutoff; $TKelig_{isc}$ is an dummy indicator for whether the student is TK eligible. Let c' represent kindergarten cohort year based upon child's birth date and kindergarten state eligibility law that applied for that particular school year when the child turned 5 (which used in constructing exogenous instruments; and thus $t' = year - c'$); while c represents the kindergarten cohort year child was observed actually entering kindergarten (which is used in the instrumented key measures; and thus $t = year - c$). Because the timing parents choose to have their child first enter kindergarten is partly endogenous and potentially an outcome of grade progression influenced by the TK treatment, it is important to use c in the construction of the instruments to ensure exogeneity, while accounting for this via 2SLS to address non-compliance issues. For example, it is found that more affluent parents are often more likely to "redshirt" their children on the margin. The indicator for child family socioeconomic disadvantage is fully interacted with all the TK and LCFF instrumental variables throughout (the models include the main effects but some of the terms are suppressed in the equations for notational ease), to enable tests of differential treatment effects for non-poor children vs children from low-income families.

(2) 1st stage (per-pupil spending):

$$\widehat{ppe}^{3yravg}_{d(s)ct} = \alpha_{1,2}(TKelig_{isc'} * SES_i) + f(Birthdate_i) * SES_i + \pi_{1,2}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'} * SES_i\right)$$
$$+ \pi_{2,2}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-1} * SES_i\right) + \pi_{3,2}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-2} * SES_i\right)$$
$$+ \pi_{4,2}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-3} * SES_i\right)$$
$$+ \pi_{5,2}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'(t'-1)(t'-2)(t'-3)} * TKelig_{isc'} * SES_i\right) + \gamma_{1,2} \cdot SES_i$$
$$+ \gamma_{2,2} Demog_i + (\eta_{t,2} * SES_i) + \theta_{sc',2}$$

(3) 1st stage (TK*spending):

$$T\widehat{Kattend}_{isc} * \widehat{ppe}^{3yravg}_{d(s)ct}$$
$$= \alpha_{1,3}(TKelig_{isc'} * SES_i) + f(Birthdate_i) * SES_i$$
$$+ \pi_{1,3}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'} * SES_i\right) + \pi_{2,3}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-1} * SES_i\right)$$
$$+ \pi_{3,3}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-2} * SES_i\right) + \pi_{4,3}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'-3} * SES_i\right)$$
$$+ \pi_{5,3}\left(Si\widehat{m\Delta LC}FF_{d(s)c't'(t'-1)(t'-2)(t'-3)} * TKelig_{isc'} * SES_i\right)$$
$$+ \gamma_{1,3} SES_i + \gamma_{2,3} Demog_i + (\eta_{t,3} * SES_i) + \theta_{sc',3}$$

(4) 2nd stage (TOT):

$$Y_{itsc} = \beta_{1t}(T\widehat{Kattend}_{isc} * SES_i) + \beta_2\left(\widehat{ppe}^{3yravg}_{dsct} * SES_i\right) + \beta_3(T\widehat{Kattend}_{isc} * \widehat{ppe}^{3yravg}_{dsct} * SES_i)$$
$$+ f(Birthdate_i) * SES_i + \gamma_{1,4} \cdot SES_i + \gamma_{2,4} Demog_i + (\eta_{t,4} * SES_i) + \theta_{sc',4} + \varepsilon_{itsc}$$

(4') 2nd stage (Reduced-Form "ITT" for TK eligibility):

$$Y_{itsc} = \beta_{1t,4'}(TKelig_{isc} * SES_i) + \beta_{2,4'}\left(\widehat{ppe}_{dsct}^{3yravg} * SES_i\right) + \beta_{3,4'}(TKelig_{isc} * \widehat{ppe}_{dsct}^{3yravg} * SES_i)$$
$$+ f(Birthdate_i) * SES_i + \gamma_{1,4'} \cdot SES_i + \gamma_{2,4'}Demog_i + (\eta_{t,4'} * SES_i) + \theta_{sc',4'}$$
$$+ \varepsilon_{itsc}$$

where ε_{itsc} is a stochastic error term. Standard errors are robust to heteroscedasticity and clustered at the district level.

The purpose of the 2SLS-RD models in (4') is to compare students who are eligible for TK with those who are not eligible for TK to estimate effects of offering the program—i.e., the intent-to-treat (ITT) effect. Ignoring noncompliance attenuates the estimated impact of TK attendance, because some of the control students might have attended TK and some parents of treatment eligible students might have chosen not to enroll their child in TK. Therefore, the 2SLS-IV-RD models in (4) aim to estimate the causal impacts of TK attendance (TOT), and in each, examine how the estimated TK effects differ by parental socioeconomic status, parental language, and school resources during subsequent elementary school years (induced by LCFF). The internal validity of the estimated TK impact in the RD design relies on the assumption that, in the absence of the TK program, there would be a smooth relationship (i.e., no discontinuity) between the student achievement outcome and the birthdate running variable. For this reason, any discontinuity observed in the student outcomes at the TK eligibility cutoff is attributable to TK. Therefore, to check the smoothness assumption, I checked for the discontinuity at the cutoff in the birthdate running variable and student and family background characteristics, such as parental socioeconomic status, parental home language, race/ethnicity, among others. The visual inspection of the figures did not reveal any jump around the cutoff for any of these other control variables. Provided that the conditional mean function $E[\varepsilon_{itsc}|Birthdate_i]$ is continuous at the TK eligibility cutoff, the causal impact of TK attendance on a student outcome is given by

$$\beta_1 = \lim_{Birthdate \downarrow cutoff} E[Y_{itsc}|Birthdate_i] \quad - \lim_{Birthdate \uparrow cutoff} E[Y_{itsc}|Birthdate_i]$$

The excluded instruments for per-pupil spending are the pre-post LCFF-induced change in a district's per-pupil funding in each year at ages t, t-1, t-2, t-3. The empirical strategy combines regression discontinuity and difference-in-differences designs to isolate the interactive effects of TK participation and LCFF-induced changes in per-child funding, which mitigate potential sources of bias.

LCFF was a $18 billion commitment of increased funding incremented distributed over a 7-year period. LCFF was a multidimensional reform that included substantial funding increases overall, greater funding explicitly targeted to socioeconomically disadvantaged districts (via concentration grants), elimination of many state categorical programs, and increased local control and local accountability. The empirical approach isolates the effects of increased spending via the pre-post changes in the funding formula, accounting for other coincident changes (see Johnson (2023) for further details). The empirical analysis is complicated by the dynamic nature of student achievement trajectories and how current learning outcomes are influenced by both the history of school resources and resources in the current assessment year. The modeling approach used accounts for the cumulative nature of learning and considers how early learning begets future learning (often in compounding ways). Using population student-level longitudinal administrative data for the full universe of public school students in California, I was able to follow the same students over time.

Measurement. To account for differences in standardized testing regimes over time, we follow the procedures developed in Reardon et al (2016) and convert test scores to a single national scale in three steps. First, we rank each student's scores in the statewide distribution (for a given grade-year-subject). Second, we use data from a national test administered (NAEP) to a sample of students by the US Dept. of Education to convert state-year-grade-subject specific rankings to national scale. Third, we convert the standardized test scores to “grade level” equivalents in both math and reading.

Beyond student outcome data, student administrative records contain demographic characteristics, birthdate, and attendance information. Annual district financial data contains all general ledger financial records (both expenditures and revenues) for all public school districts in a given year.[1] The analyses include both traditional public schools and charter schools by funding type.

I have performed an array of validation checks of my construction of TK attendance used in the study. TK attendance is best constructed by first cleaning and assembling a panel data set from the raw CDE student-level enrollment data files over time, combined with child birth dates, the prevailing K and TK eligibility policy in the year a child turned age 4-5, student-level program participation data files over time, and information collected in the enrollment files on each student’s reported grade progression sequence. Because TK was such a new program, sometimes districts were not sure how to report it for students in the earlier years of the TK rollout as often TK students were combined in kindergarten classrooms for instruction. The measurement procedures used are conducted in a way that is less prone to measurement reporting errors of TK attendance.[2]

A significant component underlying what appear as heterogeneous treatment effects (i.e., differences in the estimated causal impacts of TK attendance) are in fact due to the substantial heterogeneity in the counterfactual care conditions children would have received absent the TK program. These counterfactual settings differ in the quality and affordability of local options (including home care), language modality (bilingual immersion), program funding and resources, and a host of other factors.

[1] The analyses exclude a small portion of districts (less than 1%) with missing financial information and/or for which district annual per-pupil expenditures were not measured reliably.
[2] As a validation check, we have compared our computed school-level TK enrollment counts over time with those reported in the publicly available numbers on the CDE webpage, and we have also compared our overall annual school and district enrollment numbers with those reported in the publicly available numbers on the CDE webpage.

While it is common practice in empirical research to aggregate data, group all students, and present the average treatment effects, this convention masks enormous differences in the effects of TK attendance due to substantial differences at the starting gate by parental SES and between students whose parents are English-speakers and those whose parents are not, as well as differences in the quality of counterfactual care conditions (in the absence of TK).[3] Moreover, there are several factors that are often considered exogenous and independent of TK, but are in fact endogenous outcomes that are directly influenced by the introduction of TK, including grade progression, the age and timing of academic assessments, including initial English Learner (EL) and English proficiency designations, and the subsequent rate of EL reclassifications.[4]

Robustness checks.

In the main analyses, I use a flexible non-parametric (semi-parametric) specification for the running variable with a 60-day bandwidth around the age eligibility cutoff, and present the figures showing the main findings for TK attendance and the impact estimates of both TK eligibility ITT (reduced-form), and TK attendance (TOT), on student achievement up to 4 years after kindergarten, respectively. I also check whether the results are robust to alternative bandwidths: 15 days, 30 days, 90 days, 120 days, outcome-specific IK optimal bandwidth (Imbens and Kalyanaraman, 2012) and CCT optimal bandwidth (Calonico et al., 2014). I also checked whether the main results are robust to different model specifications and conducted a series of sensitivity analyses that tested alternative model specifications, including different functional forms for the birthdate running variable (relative to the age eligibility cutoff). Birthdate relative to the cutoff measured in days is the forcing variable that defines TK program eligibility. I chose 60 days on either side of the eligibility cutoff as the

[3] For example, an insignificant average treatment effect may be masking beneficial effects that are experienced for some sociodemographic groups and negative effects for others.
[4] In a longer companion paper (Johnson, 2024), the full set of results are presented by parental language and SES, and include estimated impacts on math and reading achievement 3 and 4 years after kindergarten, and analyses of impacts on grade repetition and socioemotional development.

optimal bandwidth, which represents students born up to two months before the cutoff and student born up to two months after the cutoff. A formal statistical test for optimal bandwidth (cross-validation method proposed by Ludwig and Miller (2007)), supports this choice of bandwidth. The main results were not very sensitive to the bandwidth selection.