The effect of universal full-day Kindergarten on student achievement*

Jane Friesen
Simon Fraser University

Brian Krauth
Simon Fraser University

Ricardo Meilman Cohn
Social Research and Demonstration Corporation

January 28, 2022

Abstract

We estimate the effect of introducing universal full-day Kindergarten (FDK) on subsequent Grade 4 test scores. The data are from the Canadian province of British Columbia, which moved from mostly half-day to universal full-day Kindergarten between the 2010/11 and 2011/12 academic years. We exploit the staggered timing of the policy implementation using a difference-in-differences research design. Our point estimates for the average effect of FDK on achievement are mostly positive, occasionally statistically significant, and always small. The effect is substantially larger among students who speak English as a second language, a result that is consistent with prior findings.

JEL codes: H75, I20, I28

Keywords: reading and numeracy skills; Kindergarten; student achievement

*Corresponding author, bkrauth@sfu.ca. This research was supported by the Social Sciences and Humanities Research Council of Canada and Simon Fraser University's Community Trust Endowment Fund. All inferences, opinions, and conclusions drawn in this article are those of the authors, and do not reflect the opinions or policies of the Data Steward (the B.C. Ministry of Education).

1 Introduction

Kindergarten classes for five-year old children are a long-standing component of many education systems, and the available evidence suggests that Kindergarten has substantial long-run benefits (Cascio 2009; Dhuey 2011). More recently, Kindergarten in North America has undergone a major transformation as traditional half-day programs have been widely replaced by programs that take up the entire school day. In the United States, the share of Kindergarten students attending full-day programs rose from 28% to 76% in just fifteen years between 1997 and 2012 (Child Trends 2013). In Canada, the population share of provinces with universal full day Kindergarten rose from 5% (two of ten provinces) to 80% (seven provinces) between 1999 and 2016 (McCuaig 2014). Keeping children in class for the full school day is expensive - in many cases this longer version of Kindergarten represents the largest expansion of public investments in early childhood education in decades.

Evidence on the results of these investments is currently limited. Earlier findings of benefits from older, mostly half-day, Kindergarten programs do not necessarily imply that these benefits increase with the length of the school day. In addition to any formal instruction that is provided, a longer day implies more time with teachers relative to other caregivers, more social interaction with peers, and possibly additional stress and fatigue. The net effect may be beneficial or harmful depending on the child's developmental stage as well as the characteristics of the classroom relative to that child's alternative care environment. Unlike the effects of Kindergarten programs in general, there are no studies of the effects of full-day versus half-day Kindergarten on long-run outcomes. Instead, the literature has focused on the short-run effects of FDK on parental labour supply (e.g., Dhuey et al. 2020) and on students' test scores.

This paper contributes to this literature by providing new evidence of the effects of a universal FDK program on reading and numeracy skills in Grade 4. The evidence is based on the Canadian province of British Columbia (B.C.), where province wide universal FDK was introduced over a two-year period. Schools serving approximately half of students began offering universal FDK in fall 2010, with the remainder following in fall 2011. The implementation of universal FDK was not paired with any change to the curriculum or other major policy changes. Additional class time was to be used for learning the existing play-based Kindergarten curriculum at a more relaxed pace and with more individualized teacher attention (British Columbia Ministry of Education 2010, p 4-5). We exploit the staggered rollout of this program to implement a difference-in-differences research design that identifies the effect of FDK by comparing cross-cohort changes in outcomes between early-adopting and late-adopting schools.

This setting allows us to advance the literature in several important ways. First, our study is the first to use population-level data to evaluate the introduction of a universal FDK program. While a small number of previous studies use random-assignment research designs to provide estimates with strong internal validity, their estimates may be highly context specific. For example, Elicker and Mathur's (1997) influential experimental study is based on the random assignment of 179 participating students in a single middle-class community. Gibbs (2014) estimates the effect of FDK on the standardized literacy test scores of almost 1,000 Kindergarten students in five Indiana school districts where students were randomly assigned to oversubscribed FDK programs. The resulting estimates have high internal validity but may not generalize to typical FDK programs even within those districts. Our population-based estimates of the effects of a province-wide universal FDK program provide insights that may be applicable in more general settings.

Several previous studies use quasi-experimental methods to identify the short-run effects of FDK in more general contexts. The closest study to ours in methodological terms is Cannon et al. (2011), who exploit staggered adoption to study the effects of FDK on the academic outcomes of English Language Learners in the Los Angeles Unified School District. Other studies rely solely on cross-sectional variation in Kindergarten programs. An important group of studies uses nationally representative data from the U.S. Early Childhood Longitudinal Study (ECLS-K) (e.g. Cannon et al. 2006; DeCicca 2007; Lee et al. 2006; Votruba-Drzal et al. 2008). The ECLS-K sample is drawn from a wide range of jurisdictions with varying and primarily non-random mechanisms for assigning children to half-day versus full-day Kindergarten, and the reported outcome measures rely on the assessments of teachers who were not themselves randomized across full- and half-day programs. These studies therefore require strong identifying assumptions in combination with various IV and differencing strategies to account for potential non-random selection of students and teachers. Warburton et al. (2012) also use an IV strategy to estimate the effects of an earlier, targeted FDK program in B.C. In contrast to this earlier research, our strategy provides credible estimates of the average effects of FDK in a more general population than in previous experimental studies, while relying on weaker identifying assumptions than those required by previous observational studies.

Previous research consistently finds that students attending FDK programs have stronger academic skills at the end of Kindergarten than those in half-day programs (e.g. Cannon et al. 2006, 2011; DeCicca 2007; Elicker and Mathur 1997; Gibbs 2012; Votruba-Drzal et al. 2008). Gibbs (2014) finds that FDK increases Kindergarten literacy scores on average by 0.31 standard deviations (SD), and by 0.70 SD among Hispanic students. Effect sizes for the average Kindergarten student in the ECLS-K are somewhat smaller: for example, DeCicca (2007) finds

effects of 0.19 SD and 0.17 SD in reading and numeracy among White students, 0.11 SD in both reading and math among Black students, and 0.24 SD and 0.16 SD in reading and math among Hispanics. Cannon et al. (2011) find that FDK increases Kindergarten reading scores among (largely Hispanic) English Language Learners by 0.13 SD. However, in a pattern that has become familiar to researchers evaluating a range of early childhood programs (for reviews see Almond and Currie 2010; Duncan and Magnuson 2013), these promising short-run results largely or fully disappear over subsequent grades. DeCicca (2007) finds that the FDK advantage disappears entirely among Black and Hispanic students by the end of first grade and falls by more than half but remains statistically significant among White students. Votruba-Drzal et al. (2008) estimate that all the advantage of FDK students in math and reading scores fades out completely by the end of third grade. Cannon et al. (2011) find no effect of FDK on reading, math or English language proficiency of English Language Learners in first or second grade. One exception to this pattern of fade-out is Warburton et al. (2012), who find a statistically significant positive effect of 0.14 SD on Grade 4 reading scores, and a statistically insignificant effect of 0.06 SD on Grade 4 numeracy.

We find that universal FDK has a typically positive, occasionally statistically significant, but robustly small effect on Grade 4 test scores in both reading and numeracy. Our point estimates range from 0.01 to 0.03 SD, and confidence intervals rule out effects much larger than that. For comparison, this is roughly the test score increase in our data associated with being one month older. Since our data set does not include measures of academic or cognitive skills at the end of Kindergarten, we cannot determine whether the effects of FDK were initially small or were initially larger but faded out across grades. However, the magnitudes of our estimates are consistent with the general pattern of fade-out documented widely in the literature. This range of

estimates is robust across numerous sample restrictions and alternative modeling choices, though there is some evidence of heterogeneity in effects. Our largest sub-group estimates are for students enrolled in English as a Second Language (ESL) programs: point estimates for this sub-group are 0.06 SD in reading and 0.05 SD in numeracy and they are statistically significant in both cases. These results are somewhat smaller than those of Warburton et al. (2012).

The general pattern of academic fade-out in the literature is puzzling in itself, even more so given that some early childhood education programs that exhibit this pattern nevertheless may deliver long-run benefits (e.g. Chetty et al. 2011; Deming 2009; Garces et al. 2002; Ludwig and Miller 2007). The specific context of our study allows us to shed additional light on some of the mechanisms that have been proposed to explain the fade-out phenomenon. A leading hypothesis is that the initial cognitive gains experienced by children who participate in early childhood education programs may be squandered when they subsequently enrol in low quality schools (e.g. Bitler et al. 2014; Currie and Thomas 1995; 2000; Duncan and Magnuson 2013; Votruba-Drzal et al. 2008). If treated as a separate country, B.C. would rank among the top 10 in 2018 PISA reading results and top 20 in mathematics, both on average and at the lower end of the distribution (see for example O'Grady et al. 2019). It seems likely that B.C.'s schools would be as well or better prepared than most to maintain or build upon any improvements in cognitive skill development in Kindergarten over subsequent grades. This result suggests that something other than poor quality education in Grades 1 through 4 is responsible for the absence of an academic advantage in Grade 4 for the average student who was enrolled in FDK under B.C.'s universal program. However, high quality schools and supplemental funding for ESL programs over subsequent grades may account for the small positive effects of FDK among English language learners in B.C. while these effects are absent in other jurisdictions.

Alternatively, the convergence of mean test scores of treated and untreated students across subsequent grades might reflect a "catch-up" effect among untreated students rather than a "fade-out" effect among treated students. If educators focus on remediating skill gaps, programs that improve cognitive skills among Kindergarten students in the treated group may result in resources being diverted towards untreated students in subsequent grades (Gibbs et al. 2013). Two features of our context largely shut down this potential mechanism. First, while there is some mobility across schools, 88% of grade 4 students who attend an early-adopting (late-adopting) school in our data also attended an early-adopting (late-adopting) school in Kindergarten. The scope for diverting resources from treated to untreated students within schools and classrooms is therefore limited. Second, B.C.'s provincial funding formula (discussed in detail below) limits resource disparities across districts that may lead to diversion of funds across schools.

A third possibility, also discussed by Gibbs et al. (2013), is that early childhood education programs may not be able to generate sustained academic advantages unless they affect important behavioral and socio-emotional skills that are fundamental to cognitive development. Friesen et al. (2017) find that B.C.'s universal FDK program had little effect on children's Kindergarten behavior on average, but reduced child hyperactivity and peer relationship problems among children whose families do not speak English at home. This pattern of evidence hints at social and emotional development resulting from FDK participation as a potential mechanism driving the improvement of later cognitive skills that we find for ESL students.

2 Data and methods

2.1 Full-day Kindergarten in British Columbia

Figure 1 illustrates the time series of FDK participation for the period covered by our study. Through 2009/10, students who were Aboriginal, spoke English as a Second Language (ESL) or had certain disabilities were eligible to enroll in targeted FDK programs. These students accounted for approximately 27% of Kindergarten students, and the take-up rate among eligible students during this period was about 55%. Students outside these targeted groups were almost exclusively in half-day programs. In August 2009, the provincial government announced that all schools would implement universal FDK by the 2011/12 school year, with 2010/11 as a transition year in which approximately half of schools would do so. During the transition year, FDK continued to be available to students in the individually targeted groups. Starting in 2011/12 FDK became universal in all public schools and all but a few private schools.

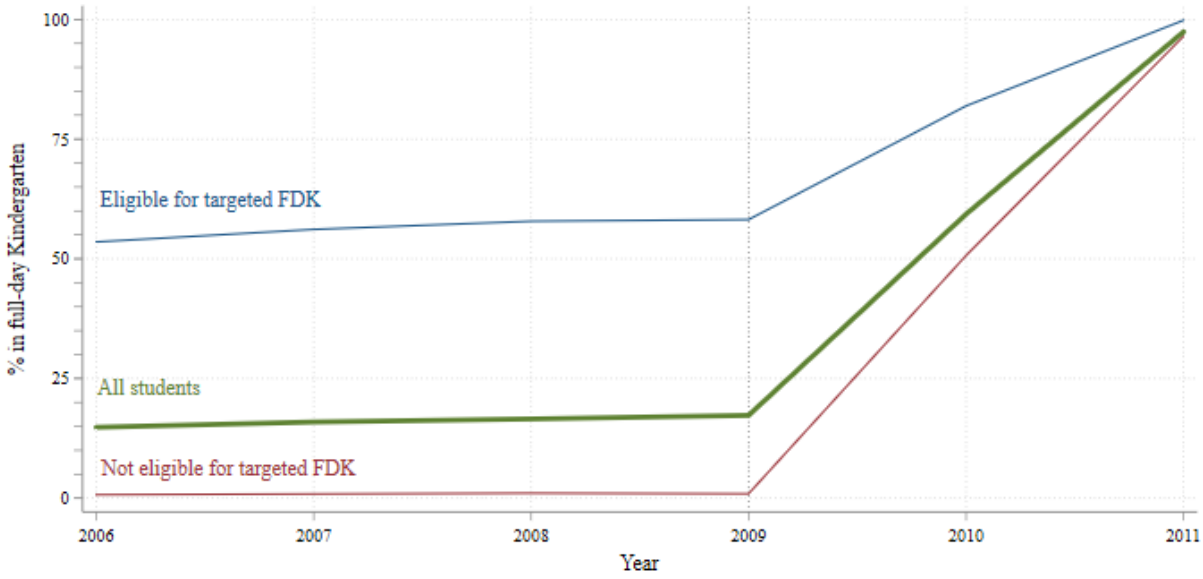

Figure 1: Full-day Kindergarten in British Columbia, 2006/07-2011/12. Students eligible for targeted FDK include Aboriginal, ESL and disabled students.

According to the histogram in Figure 2 below, the school-level FDK participation rate among students who were not already eligible for a targeted FDK program was either 0% or 100% in 2010/11 in all but a few schools. This pattern confirms that B.C. rolled out the new universal FDK program on a school-level basis. We therefore assign all students who were enrolled in Kindergarten in 2010/11 at an "early adopting" school (where the FDK participation rate among non-targeted students was over 50% in 2010/11) to our treatment group and assign all students in "late adopting" schools (where the FDK participation rate among non-targeted students was below 50% in 2010/11) to our comparison group.

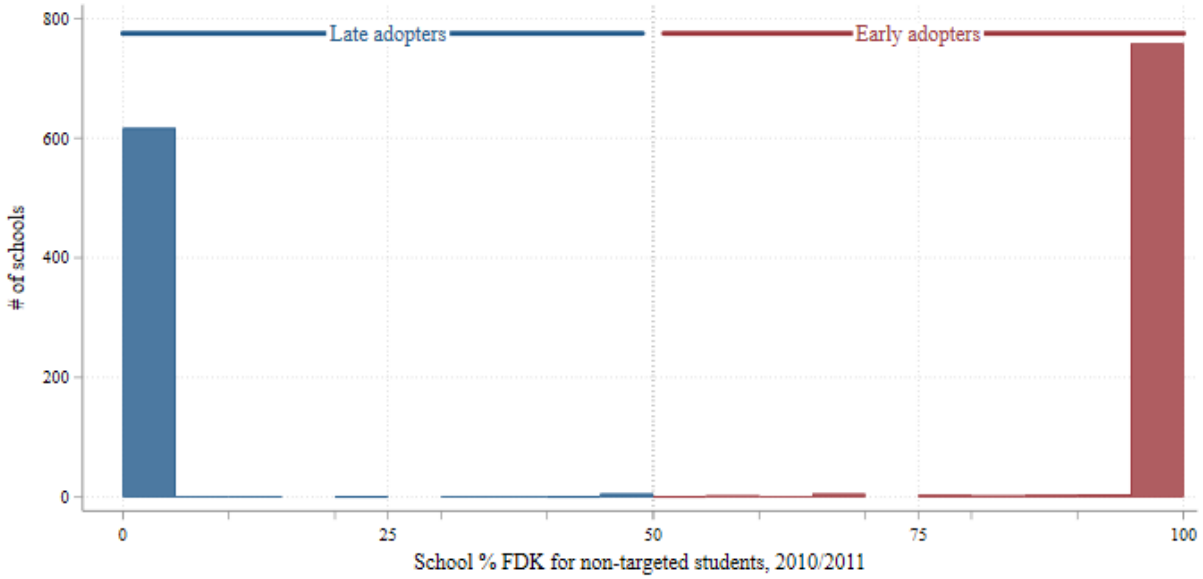

Figure 2: School-level full-day Kindergarten rates for non-targeted students, 2010/11. Schools with FDK rates for non-targeted students above 50% are classified as early adopters of universal FDK and schools below 50% are classified as late adopters.

The institutional context of this policy change has three distinctive features relative to most other North American jurisdictions. First, education in B.C. is funded by the provincial Ministry of Education rather than from local tax revenue. Operating funds are provided using a formula based primarily on full-time equivalent (FTE) enrollment, with supplements for student, school and district special circumstances. Each district's local school board receives these funds and is

responsible for implementation and resource allocation across schools and programs. Second, most private ("independent") schools also receive a per-FTE operating grant and participate in the Ministry's enrollment data collection and standardized testing. As a result, our analysis covers both public and private schools. Third, B.C. operates Francophone schools (serving students with a constitutional right to be educated in French) as well as French Immersion programs (providing instruction conducted in French but aimed primarily at English-speaking students) that are typically co-located with standard English programs. Our analysis treats the English program and French Immersion program in the same school location as distinct "schools."

Table 1 describes in more detail how the roll-out of universal FDK was implemented within the funding model. The provincial funding formula granted half as much funding per student in half-day versus full-day Kindergarten programs (0.5 full-time equivalent (FTE)). Overall transition year funding was provided based on implementation plans submitted by both school districts and independent school authorities. Districts were asked to continue any existing targeted FDK provision, and to propose a list of schools for early adoption of universal FDK. The provincial government asked that implementation plans prioritize schools serving more vulnerable populations. According to interviews with district personnel as well as public statements, most districts considered both the vulnerability of the school population and the availability of teachers and facilities. Although most districts (including all large districts) followed a staggered implementation plan, a few smaller districts committed additional resources to implement universal FDK in all schools in 2010/11 and a few others elected to delay implementation to 2011/12 in all schools. The additional operating cost to the Ministry of Education of funding Kindergarten students on a full-day rather than a half-day basis amounted to approximately

$3,370 per affected student in public school or about $100 million per year in total when the roll-out was complete.

Table 1: Kindergarten funding in British Columbia

Type	Funding	Eligibility		
		Before 2010/11	**2010/11**	**After 2010/11**
Public	1.0 FTE	FDK students from targeted groups (ESL, Aboriginal, disabled)	FDK students from targeted groups FDK students in designated early-adopting schools	All students
	0.5 FTE	All other students	All other students	No students (all are to be in FDK)
Private	1.0 FTE	FDK students from targeted groups	FDK students from targeted groups FDK students (to a pre-determined school-specific maximum) in designated early-adopting schools.	FDK students .
	0.5 FTE	All other students	All other students	All other students

Private schools also received transition-year funds, with 103 schools each awarded a predetermined number of FDK "spots" that could receive 1.0 FTE funding if those students received FDK (Government of British Columbia 2010). Most of these schools implemented universal FDK in 2010/11, as did a few others. In contrast to public schools, a few private schools implemented universal FDK before 2010/11 and a few private schools remained half-day after 2010/11.

2.2 Data

Our data come from merged administrative records of the B.C. Ministry of Education, which include confidential student-level enrollment information and Grade 4 test results (British Columbia Ministry of Education 2020a), as well as publicly available school-level data for all public and private schools in the province (British Columbia Ministry of Education 2020b).

The student-level enrollment data is collected each September and includes longitudinal records through the 2014/15 school year for all students who attended Kindergarten between 2006/07 and 2010/11. These records include birth month and year, gender, current grade, current school, self-reported Aboriginal identity, enrollment in a language program (e.g. ESL, French Immersion, Francophone education), language spoken at home and special needs designation. The school/grade-level data includes school name, type, and location, constructed from publicly available data. Additional school/grade-level variables are constructed by aggregating the confidential student-level data.

Test results are from the Numeracy and Reading portions of the Grade 4 Foundation Skills Assessment (FSA) exams, which are given every spring. Both exams include a mix of multiple choice and open-ended questions and are graded anonymously by certified B.C. teachers. The FSA exams are low-stakes, and do not directly affect student outcomes or personnel decisions. However, results are returned to the student's parents, school-level results are available publicly, and the Fraser Institute (a Vancouver-based policy advocacy organization) uses the school-level results to construct a prominent "report card" ranking all schools in the province. Both exam grades are reported as Item Response Theory (IRT) scale scores, standardized to mean zero and unit variance within our main analysis sample. Finally, Francophone students can take the exams in French. Our data include the French-language exam scores for numeracy but not for reading.

2.3 Empirical framework

Our main empirical framework employs variations on the standard difference-in-differences research design using school-level measures of universal FDK availability. Let $\overline{FDK}_{sc}$ be the proportion in FDK in school s among students in Kindergarten cohort c who are not eligible for

targeted FDK funding (i.e among those who are not Aboriginal, ESL or disabled). We then classify school s as providing "universal FDK" to cohort c if this proportion is greater than 50%:

$$UFDK_{sc} = \begin{cases} 1 & if\ \overline{FDK}_{sc} \geq 0.5 \\ 0 & if\ \overline{FDK}_{sc} < 0.5 \end{cases}$$

We use this school-level indicator of universal FDK status in the transition year of 2010/11 to classify each school as an "early adopter" or "late adopter" of universal FDK:

$$EarlyAdopter_s = UFDK_{s,2010}$$

While the 50% threshold for classifying schools is unavoidably arbitrary, most schools would have the same classification under any reasonable alternative threshold (see Figure 2). Section 3.4 below shows that our results are robust to using a stricter 99% threshold. Our test score data includes four pre-treatment cohorts (2006 – 2009) and one post-treatment cohort (2010).

Our baseline model follows a conventional two-way fixed effects specification. Let *i* index students, *s* index schools, and *c* index entry cohorts or years, and let:

$$y_i = \theta UFDK_{s(i),c(i)} + a_{s(i)} + \delta_{c(i)} + X_i\beta + u_i \qquad \text{(FE1)}$$

where y_i is the outcome of interest, a_s is a school fixed effect, δ_c is a cohort/year fixed effect, and X_i is a vector of student characteristics. The student's cohort *c(i)* is defined as their (first) year in Kindergarten, and their school *s(i)* is defined as the school they attended in that year. We estimate model (FE1) by the standard linear fixed effects ("within") estimator, with standard errors clustered on the school location. A common alternative formulation of model (FE1) would use $EarlyAdopter_s * 1\{Year = 2010\}$ rather than $UFDK_{sc}$. These two variables would be

identical for most schools, but a few private schools have universal FDK before 2010. Section 3.4 below shows that our results are robust to dropping these schools.

The parameter of interest θ can be interpreted as the effect of attending a school with universal FDK, which is closely related to but distinct from the direct individual-level effect of FDK. For public-school students who are not eligible for targeted FDK, the two treatments coincide: such students are in FDK if and only if $UFDK_{sc} = 1$. Public-school students who would be eligible for targeted FDK can be classified conceptually into two groups: those who would have taken up targeted FDK (always-takers) and those who would not have (compliers). For compliers, the implementation of universal FDK represents a move from HDK to FDK, as it does for non-targeted students. For always-takers, the implementation of universal FDK represents a move between two distinct *forms* of FDK and its effect may be quite different in both magnitude and sign than a move from HDK to FDK. The effect of universal FDK is similarly complex for students in private schools, some of whom offered FDK before 2010/11.

This heterogeneity has several implications for the analysis. First, the possibility that universal FDK has a nonzero effect on always-takers rules out its use as an instrument for student-level participation in FDK, as it violates the exclusion restriction. Second, we report estimates for both the full sample and a restricted sample of public-school students who are not eligible for targeted FDK. Identification is cleaner and interpretation is simpler in the restricted sample, but the full sample estimates provide useful information on the effect of FDK among the targeted groups.

The two-way FE model in this setting allows for systematic unobserved differences across students, time and schools but imposes three important restrictions: that the untreated outcomes follow (conditional) parallel paths for early-adopting and late-adopting schools, that the

treatment effect is constant within the treatment group, and that the functional form of the model is correct. Section 3.4 below follows up our baseline two-way FE estimates with other estimation methods that more explicitly account for heterogeneity and relax the functional form assumptions.

3 Results

3.1 Summary statistics

Table 2 below presents summary statistics for our main analysis sample, which consists of all students attending Kindergarten in a standard B.C. public or private school between 2006/07 and 2010/11 who also attend a B.C. school for Grade 4 between 2010/11 and 2014/15. [1] The restriction to students attending a standard school for Kindergarten rules out a few students in distance education or alternative programs for which the school day is not well-defined.

The summary statistics provide additional context for our results. English-language public schools enroll about 76% of the province's Kindergarten students, with private schools and public French Immersion or Francophone programs accounting for roughly equal shares of the rest. Approximately 27% of students are eligible for targeted FDK. ESL students (19%) account for a majority of these students, with Aboriginal students (8%) accounting for most of the rest. These groups overlap: while almost all Aboriginal students in B.C. speak English at home, a substantial number receive "English as a Second Dialect" support under the ESL designation (Battisti et al. 2014), and some Aboriginal and/or ESL students also have special needs. Students self-identify as Aboriginal annually, so the proportion of students who ever identify as Aboriginal is somewhat higher at 11%. Our analysis uses current (Kindergarten) Aboriginal

[1] Students can move in or out of province during the academic year and may be in the September enrollment data but not the spring FSA data or vice versa. We exclude such students from the main analysis sample.

identification for classifying students as eligible for targeted FDK, and "Aboriginal ever" when controlling for student characteristics in regressions. The most common non-English home languages are Punjabi and Chinese/Cantonese/Mandarin, followed in order by Tagalog/Pilipino, Vietnamese, Spanish, Korean, Hindi, French, and Persian. The proportion identified in Kindergarten as gifted or disabled is quite small and dominated by those disability categories that require extensive individual support and imply eligibility for targeted FDK. Learning disabilities, mild intellectual disabilities, and behavior disorders are identified at a much higher rate in later grades. Finally, the average entering Kindergarten student is 61.8 months old on September 1, as one would expect if most children start Kindergarten on schedule. Delayed school entry ("redshirting") and grade repetition are uncommon in B.C.

Column 2 reports summary statistics for the public non-targeted subsample. This sample differs from the full sample as implied by the sample definition, but is broadly similar in gender, age and test results. Although this sample excludes students who are administratively classified in Kindergarten as ESL, Aboriginal, or having certain special needs, it still includes a few students whose reported home language is not English, who self-identify as Aboriginal in later school years, or who have special needs that have not yet been identified.

Table 2: Summary statistics for main analysis sample.

Characteristic	All students	Public non-targeted	By school FDK adoption timing		Exam participants only	
			Early (2010/11)	Late (2011/12)	Numeracy	Reading
# students	188,800	117,300	98,600	81,900	156,000	154,000
# schools	1,556	1,243	774	597	1,549	1,549
Kindergarten school choice						
Public	0.874	1.000	0.910	0.859	0.863	0.861
English	0.759	0.841	0.872	0.657	0.748	0.755
French Immersion	0.102	0.154	0.038	0.177	0.104	0.104
Francophone	0.012	0.005	0.000	0.025	0.011	0.002
Private	0.126	0.000	0.090	0.141	0.137	0.139
English or bilingual	0.125	0.000	0.090	0.139	0.135	0.137
French Immersion	0.001	0.000	0.000	0.003	0.002	0.001
Student characteristics						
Male	0.513	0.506	0.522	0.503	0.504	0.503
Female	0.487	0.494	0.478	0.497	0.496	0.497
Eligible for targeted FDK	0.274	0.000	0.342	0.161	0.258	0.253
ESL	0.191	0.000	0.246	0.106	0.190	0.184
Home language:						
English	0.803	0.944	0.759	0.866	0.798	0.800
Chinese[2]	0.047	0.011	0.058	0.037	0.049	0.050
Punjabi	0.060	0.008	0.076	0.024	0.063	0.064
Other language[3]	0.090	0.037	0.108	0.073	0.089	0.086
Aboriginal (now)	0.084	0.000	0.102	0.045	0.074	0.074
Aboriginal (ever)	0.119	0.050	0.141	0.077	0.107	0.107
Disabled (all)	0.022	0.004	0.026	0.019	0.009	0.010
(targeted)	0.019	0.000	0.022	0.016	0.008	0.008
(non-targeted)	0.003	0.004	0.004	0.003	0.002	0.002
Gifted	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Age on 9/1, months	61.8	61.8	61.7	61.8	61.7	61.7
(std dev)	(3.6)	(3.5)	(3.6)	(3.6)	(3.5)	(3.5)
FDK status						
In FDK	0.250	0.106	0.393	0.053	0.238	0.233
In universal FDK ($UFDK_{s(i)c(i)}$)	0.123	0.106	0.228	0.001	0.117	0.117
School is early adopter ($EarlyAdopter_{s(i)}$)	0.546	0.499	1.000	0.000	0.537	0.542
Grade 4 FSA exam results						
Numeracy participation	0.826	0.833	0.812	0.844	1.000	0.975
Numeracy score	0.000	-0.047	-0.104	0.120	0.000	0.012
(std dev)	(1.000)	(0.945)	(0.988)	(0.984)	(1.000)	(0.998)
Reading participation	0.816	0.827	0.809	0.825	0.963	1.000
Reading score	0.000	0.023	-0.095	0.123	0.009	0.000
(std dev)	(1.000)	(0.973)	(0.992)	(0.992)	(0.998)	(1.000)

Population is students in B.C. Kindergarten during 2006/07-2010/11 who also attended a B.C. school for Grade 4. Sample size is rounded to the nearest 100 to reduce disclosure risk. Student characteristics are based on student's Kindergarten enrollment record.

[2] "Chinese" includes students who reported speaking Cantonese, Mandarin or Chinese at home.

[3] The most common home language in the "other" category is Tagalog/Pilipino (1.2%), followed by Vietnamese, Spanish, Korean, Hindi, French, and Persian (all between 0.4% and 0.9%).

Columns 3 and 4 divide the full sample into early-adopting and late-adopting schools. Note that a few schools cannot be classified as early or late adopters because they were not open or had no non-targeted Kindergarten students in 2010. Early-adopting schools account for a slight majority of students and of classified schools. A comparison across school types confirms that policy makers generally followed the Ministry's directive to prioritize schools with higher populations of vulnerable students. Early adopting schools have relatively more ESL students, more Aboriginal students, and more disabled students, as well as lower Grade 4 test scores. They also have somewhat more male students; this is primarily attributable to the Francophone and French Immersion sectors, both of which have fewer boys and mostly late-adopting schools. Average age does not differ substantially between early and late adopting schools.

Columns 5 and 6 characterize exam participants. Over the period covered by the data, FSA exam participation averages 82.6% for numeracy and 81.6% for reading. Exam participants are more likely to be female and to attend a private school, and less likely to be identified[4] as Aboriginal, ESL or disabled.

Table 3: Grade 4 FSA participation rates by Kindergarten year, main analysis sample

Kindergarten Year	Exam Year	Numeracy Participation	Reading Participation
2006/07	2010/11	0.831	0.823
2007/08	2011/12	0.835	0.825
2008/09	2012/13	0.839	0.830
2009/10	2013/14	0.831	0.823
2010/11	2014/15	0.796	0.780
All years		0.826	0.816

Participation rates are calculated by Kindergarten year, "exam year" is the year in which the student would normally be in Grade 4.

[4] Although private schools submit the same student information as public schools, they may face differential incentives to identify students in these categories. As a result, identification rates may differ between public and private schools.

Table 3 above shows FSA test participation rates over time. Participation dipped in 2014/15 following an anti-FSA campaign by the B.C. Teacher's Federation. We confirm in Section 3.2 that this dip in participation is not related to the school-level introduction of universal FDK.

Figure 3 provides a simple preview of our main results in the form of a time series graph for average FSA scores, according to attendance at an early-adopting or late-adopting school. Almost all schools were untreated (did not have universal FDK) through 2009/10, while the early-adopting schools were treated (had universal FDK) in 2010/11. The figure shows substantial pre-treatment differences between early-adopting and late-adopting schools, with a difference in average test scores around 0.22 SD. Year-to-year variation in pre-treatment test scores is substantially smaller in magnitude. The deviations from parallel paths in the pre-treatment period are small to moderate, implying that our identifying assumption of parallel paths in the post-treatment period is plausible. Finally, the change in average FSA scores for early-adopting schools between 2009/10 and 2010/11falls within the normal range of year-to-year fluctuations during the pre-treatment period. This observation suggests we are unlikely to find large positive or negative effects, and our more detailed analysis below confirms this.

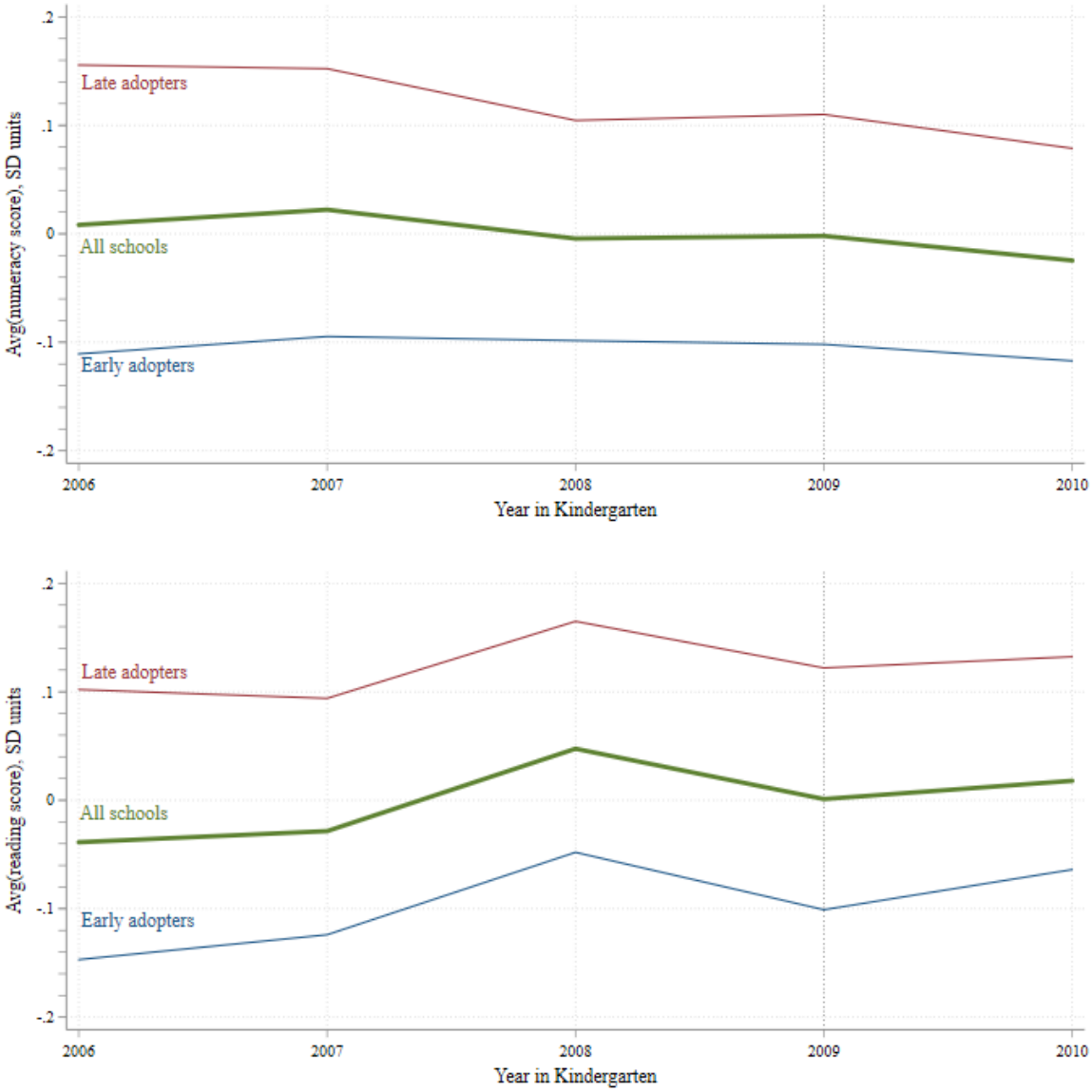

Figure 3: Average Grade 4 FSA scores by early/late adopter status of school. Pre-treatment period is 2006-2009. Early adopters are treated (adopt universal FDK) in 2010, late adopters are untreated.

3.2 Main estimates

Table 4 reports our estimates for the effect of universal FDK on exam participation. As discussed in Section 2.3, we report results both for the full sample and for a restricted sample of public-school students who would not have been eligible for targeted FDK. For this restricted group, the coefficient on universal FDK can be interpreted as the treatment effect of (universal) FDK relative to HDK. As the full sample includes both private-school students and those eligible for

targeted FDK, the coefficient on universal FDK represents the effect of universal FDK relative to a mixture of alternatives including both HDK and targeted FDK. The "public non-targeted" sample is identical to that described in column 2 of Table 2, while the "all students" sample includes all students described in column 1 of Table 2 except roughly 4,100 students whose Kindergarten cohort did not include any non-targeted students and thus do not have a well-defined value for $UFDK_{sc}$. All regressions include school and cohort/year fixed effects, and standard errors are clustered by school location.

We find negligible effects on exam participation in both the restricted and full samples, implying our results for exam scores are not likely to be driven by differential participation. The point estimates indicate that universal FDK raises numeracy exam participation by 0.3 to 0.6 percentage points and reading exam participation by 0.2 to 0.6 percentage points. None of these estimates are statistically significant at any conventional significance level, and the associated 95% confidence intervals rule out anything other than negligible effects on participation.

Table 4: Effect of universal FDK on exam participation, two-way fixed effect models, main analysis sample

Description	Numeracy participation		Reading participation	
	Public non-targeted	All students	Public non-targeted	All students
Universal FDK	0.006	0.003	0.006	0.002
	(0.008)	(0.007)	(0.009)	(0.007)
Number of students	117,300	184,700	117,300	184,700
Number of clusters	1,119	1,403	1,119	1,403

Sample sizes rounded to nearest 100 to reduce disclosure risk. Student-level control variables include gender, Aboriginal identity (ever), disability (2 categories), ESL status, home language (4 categories), and age in months. Fixed effects by school and year. Standard errors (in parentheses) clustered by school location.
Significance levels * = 0.10, ** = 0.05, *** = 0.01

Table 5 reports our baseline two-way FE results for Grade 4 FSA exam scores. Aside from the dependent variable and the restriction of the sample to exam participants, the estimation details are the same as those described for Table 4. Our estimated treatment effects are small and at most marginally statistically significant. For the numeracy exam, our point estimate indicates

that universal FDK raises test scores in the restricted sample by 0.01 SD with a 95% confidence interval of -0.03 to 0.05 SD, and in the full sample by 0.03 SD with a 95% confidence interval of -0.01 to 0.06 SD. For the reading exam, our point estimate indicates that universal FDK raises test scores in the restricted sample by 0.02 SD with a 95% confidence interval of -0.02 to 0.06 SD, and in the full sample by 0.03 SD, with a 95% confidence interval of -0.01 to 0.06 SD.

Table 5: Effect of universal FDK on Grade 4 exam scores, two-way fixed effect models, main analysis sample

Description	Numeracy score		Reading score	
	Public non-targeted	All students	Public non-targeted	All students
Universal FDK	0.012	0.027	0.020	0.029*
	(0.020)	(0.017)	(0.018)	(0.015)
Number of students	97,700	152,700	97,000	150,900
Number of clusters	1,113	1,396	1,107	1,395

Sample sizes rounded to nearest 100 to reduce disclosure risk. Student-level control variables include gender, Aboriginal identity (ever), disability (2 categories), ESL status, home language (4 categories), and age in months. Fixed effects by school and year. Standard errors (in parentheses) clustered by school location.
Significance levels * = 0.10, ** = 0.05, *** = 0.01.

These effects are comparable in magnitude to the association between test score and being one month older, which is typically about 0.02 SD in these same models. For comparison to other FDK studies, estimates of the short-run effect of FDK on Kindergarten math scores are 0.12 SD in Cannon et al. (2006) and 0.11-0.17 SD in DeCicca (2007), while estimates for the effect on Kindergarten reading scores are 0.15 SD in Cannon et al. (2006), 0.11-0.24 SD in DeCicca (2007), and 0.31 SD in Gibbs (2014). Our finding of a small and often statistically insignificant effect is consistent with the results of Cannon et al (2006), DeCicca (2007), and Votruba-Drzal et al. 2008 for later grades. To provide a comparison from outside of the FDK literature, Chetty et al. (2014) find a one standard deviation increase in teacher quality (value-added) raises math test scores by 0.14 SD and English test scores by 0.10 SD. Finally, we can compare these effect sizes to conventional classification schemes. Our effect sizes are clearly small by the commonly used Cohen (1988) scheme, which describes 0.2 SD effects as “small”, 0.5 SD effects as “moderate”

and 0.8 SD effects as "large." Kraft (2020) argues that a more appropriate benchmark for educational interventions is 0.00-0.05 SD for a "small" effect, 0.05-0.20 for a "medium" effect, and 0.20+ SD for a "large" effect. By any criteria, there is little in our main results to indicate anything other than a small overall effect of universal FDK on Grade 4 test scores.

3.3 Heterogeneity

Although our baseline analysis treats the effect of universal FDK as a constant parameter, there are many reasons one might expect it to vary substantially across students. First, Kindergarten itself is structured around the belief that the students are at an age of transition from learning through unstructured play and one-on-one interaction to learning in a more structured setting with more complex social interactions. Children entering Kindergarten will be at various stages along that developmental trajectory and each may respond quite differently to a longer school day. Second, the teacher effects literature has established that classroom effectiveness varies widely across individual teachers. Extra time may be beneficial with a strong teacher but not with a weak teacher. Finally, even if all students and teachers were the same, the alternatives to treatment include formal day care, parental care, and informal care of widely varying qualities and developmental opportunities. It is therefore important both to investigate heterogeneity directly (as is done in this section), and to more explicitly account for it in estimating average effects (as is done in Section 3.5).

Table 6 below reports the results of various direct analyses of heterogeneity estimated by adding an interaction term to the two-way FE model:

$$y_i = (\theta_0 + \theta_1 H_i) UFDK_{s(i),c(i)} + a_{s(i)} + \delta_{c(i)} + X_i\beta + u_i \qquad \text{(FE2)}$$

where H_i represents a student characteristic that is also included in X_i. In most cases, H_i is binary and we report the estimated effect for each group (e.g., θ_0 and $\theta_0 + \theta_1$) as well as the difference between the two (θ_1). Heterogeneous effects are reported for the full sample, and for the restricted sample where applicable.

Our first set of comparisons are between native English speakers and those who speak English as a second language, using both the administrative classification of "ESL" and the self-reported language spoken at home. Prior results consistently suggest that FDK has a stronger positive impact on both test scores (DeCicca 2007, Gibbs 2014) and behavior (Friesen et al. 2017) for ESL students. Our results in the first two panels of Table 6 reinforce these previous findings: the positive impact of universal FDK is two to three times as large for ESL students as for native English speakers, and this gap is statistically significant for numeracy. Breaking the results down by home language, the impact is largest for Punjabi-speaking students in numeracy and for Chinese-speaking students in math, and the non-English/English gap is marginally significant in both subjects. These results are more striking when one notes that they are estimates of the effect of making FDK *universal*: many ESL students were already in targeted FDK so the movement from targeted to universal FDK increased FDK participation for a smaller share of ESL students compared to native English speakers.

Table 6: Heterogeneity analysis, two-way fixed effect models with interaction terms, main analysis sample.

Description	Numeracy score		Reading score	
	Public non-targeted	All students	Public non-targeted	All students
Base FE model	0.012	0.027	0.020	0.029*
	(0.020)	(0.017)	(0.018)	(0.015)
ESL status				
ESL		0.062**		0.049**
		(0.024)		(0.023)
Non-ESL		0.017		0.023
		(0.018)		(0.016)
(ESL) – (non-ESL)		0.045**		0.026
		(0.023)		(0.022)
Home language				
English		0.018		0.020
		(0.018)		(0.016)
Chinese		0.041		0.082**
		(0.038)		(0.035)
Punjabi		0.086**		0.066**
		(0.035)		(0.033)
Other language		0.043		0.040
		(0.031)		(0.027)
(non-English) – (English)		0.038*		0.038*
		(0.021)		(0.020)
Aboriginal identity				
Aboriginal (ever)		-0.027		-0.039
		(0.025)		(0.029)
Non-Aboriginal		0.035*		0.039**
		(0.018)		(0.016)
(Aboriginal) – (non-Aboriginal)		-0.062***		-0.078***
		(0.023)		(0.028)
Gender				
Male	0.030	0.046**	0.023	0.031*
	(0.022)	(0.019)	(0.021)	(0.017)
Female	-0.006	0.008	0.018	0.028
	(0.023)	(0.019)	(0.021)	(0.017)
(Male) – (Female)	0.036*	0.038**	0.004	0.003
	(0.019)	(0.015)	(0.020)	(0.016)
Age (in months)				
Main effect	0.013	0.027	0.020	0.029*
	(0.020)	(0.017)	(0.018)	(0.015)
Interaction term	0.003	0.002	-0.001	-0.001
	(0.003)	(0.002)	(0.003)	(0.002)

Student-level control variables include gender, Aboriginal identity (ever), disability (2 categories), ESL status, home language (4 categories), and age in months. Fixed effects by school and year.
Standard errors (in parentheses) clustered by school location. Significance levels * = 0.10, ** = 0.05, *** = 0.01

The next set of results compares students by Aboriginal identity, gender, and age. We find that Aboriginal students benefit significantly less from universal FDK than non-Aboriginal students, and the overall effect on their test scores may even be negative. As in previous studies of FDK (e.g. Cannon et al. 2006; DeCicca 2007; Gibbs 2014), we see no evidence that the effect of universal FDK on reading varies by gender. However, unlike Cannon et al. (2006) and DeCicca (2007), we find a statistically significant difference between the effect of universal FDK on the numeracy scores of boys (0.05 SD in the "all students" sample, statistically significant) and girls (0.01 SD, insignificant). Finally, we see little evidence of heterogeneity by age in either numeracy or reading.

3.4 Robustness checks, main model

Table 7 reports the results from several robustness checks that evaluate alternatives to various decisions made on sample definition and choice of control variables within the context of our benchmark two-way FE model. Section 3.5 reports and discusses results for alternatives to the two-way FE model.

The first set of robustness checks involve dropping observations based on characteristics of the student's (Kindergarten) school. As the table shows, our point estimates are not particularly sensitive to excluding Francophone schools, or those few private schools that had universal FDK before 2010. The second set of robustness checks involve dropping observations for students with a non-standard trajectory between Kindergarten and Grade 4, including students who change schools and students who repeat grades. Again, the point estimates do not change dramatically though a small decrease in point estimates and increase in standard errors leads to a loss of statistical significance in some cases. The final set of checks use alternative definitions for key variables. The first of these alternative specifications use the Grade 4 school rather than

the Kindergarten school in constructing both the treatment and the fixed effect. The second alternative specification uses a 99% cutoff rather than a 50% cutoff for classifying schools as having "universal" FDK. Neither of these alternative specifications has a substantial impact on results.

To summarize, our results are robust in the sense that the point estimates do not change much in response to variations in the details of our analysis. At the same time, these estimates are all within the two-way FE framework which relies on modeling and identifying assumptions that may not hold. It is also important to consider robustness to plausible alternative models as will be investigated in Section 3.5.

Table 7: Robustness checks, two-way fixed effects models.

Description	Numeracy score		Reading score	
	Public non-targeted	All Students	Public non-targeted	All Students
Base FE model	0.012	0.027	0.021	0.029*
	(0.020)	(0.017)	(0.018)	(0.015)
Sample restrictions				
Exclude Francophone schools[5]	0.015	0.031*	0.021	0.030**
	(0.020)	(0.017)	(0.018)	(0.015)
Exclude private schools that had universal FDK before 2010		0.031*		0.032**
		(0.018)		(0.016)
Exclude students who change schools	0.018	0.030	0.019	0.029
	(0.026)	(0.023)	(0.023)	(0.020)
Exclude students who repeat grades	0.011	0.026	0.021	0.030**
	(0.020)	(0.017)	(0.018)	(0.015)
Alternative variable definitions				
School effect and treatment based on Grade 4 school	0.018	0.032	0.019	0.031*
	(0.024)	(0.021)	(0.020)	(0.018)
Schools classified as universal FDK if non-targeted FDK rate ≥ 99%	0.016	0.026	0.024	0.029*
	(0.020)	(0.017)	(0.018)	(0.016)

Student-level control variables include gender, Aboriginal identity (ever), disability (2 categories), ESL status, home language (4 categories), and age in months. Fixed effects by school and year. Standard errors (in parentheses) clustered by school location.
Significance levels * = 0.10, ** = 0.05, *** = 0.01

[5] As mentioned earlier, we have numeracy results but not reading results for most Francophone students. As a result, excluding Francophone schools has almost no effect on the results for the reading exam score.

3.5 Robustness checks, alternative models

As discussed in Section 2.3, the two-way fixed effects model is convenient but imposes several strong restrictions, including constant effects and a specific functional form. This section considers several alternative methods for measuring average treatment-on-treated (ATT) effects allowing for effect heterogeneity and imposing weaker functional form assumptions.

Our primary alternative method is Abadie's (2005) semiparametric estimator, which consistently estimates the ATT under the assumption of parallel paths conditional on a set of pre-treatment variables. To implement this estimator, we aggregate our repeated cross sections of students into a panel of schools and use 2009/10 as the pre-treatment period. We then use the `absdid` Stata package (Houngbedji 2016) to calculate inverse probability weighted estimates of the ATT effect, with propensity scores calculated using a second-order series logit estimator. In our preferred specification, the conditioning variables in the propensity score include both school-level averages of our benchmark student-level controls, as well as school characteristics – public/private, total Kindergarten enrollment, and a French Immersion indicator – that are absorbed into the school fixed effect in our previous models. Due to software limitations, standard errors are calculated assuming independence across schools rather than clustering of schools within the same location. Allowing for such clustering would presumably increase the standard errors slightly.

This semiparametric DD estimator relaxes the functional form restrictions (linearity, constant effect) of the two-way FE model but does assume correct specification of the propensity score model. We also report estimates for Sant'Anna and Zhao's (2020) doubly-robust estimator that combines flexible parametric models for both the outcome and the propensity score. This estimator is consistent if either the outcome model or propensity score model is correctly

specified. We use Sant'Anna and Zhao's "improved" doubly-robust estimator with a first-order logit model for the propensity score and a linear outcome model with interactions.

The first panel of Table 8 reports results for our benchmark two-way FE model, both in its original form (student-level, all years) and in the aggregated form we use for estimating the alternative models (school-level, 2009 and 2010 only). This allows us to distinguish between changes in results driven by the alternative model itself and changes driven by the data restrictions needed to estimate the model. As the table shows, the data restrictions reduce the estimated effect for the "all students" sample but increase it for the restricted sample. Additional investigation (not reported here) suggests that the primary source of difference is the year restrictions and not the aggregation.

The second panel reports results from several variations on the Abadie (2005) semiparametric DD model. The first row shows our preferred specification with both school-level and average student-level conditioning variables, while the second row shows a specification without the school-level characteristics. The third row shows our preferred specification with a trimmed sample that excludes schools with a propensity score over 90%. Taken together, these ATT estimates are typically somewhat higher than our two-way FE results based on the same aggregated data, with point estimates ranging from 0.02 SD to 0.09 SD. Standard errors are

substantially higher and would likely increase further with clustering on school location.

Table 8: Robustness checks, alternative estimation methods

Description	Numeracy score		Reading score	
	Public non-targeted	All students	Public non-targeted	All students
Base FE model				
Student-level, all students, all years	0.012	0.027	0.021	0.029*
	(0.020)	(0.017)	(0.018)	(0.015)
School-level, 2009/10-2010/11 only	0.017	0.014	0.036	0.023
	(0.028)	(0.023)	(0.027)	(0.023)
Semiparametric DD (Abadie 2005)				
Preferred specification: school controls	0.042	0.026	0.058	0.066
+ avg student-level controls	(0.043)	(0.034)	(0.043)	(0.052)
Avg student-level controls only	0.051	0.025	0.048	0.049
	(0.051)	(0.031)	(0.047)	(0.040)
Preferred specification, trimmed with	0.064	0.020	0.087**	0.023
propensity score <= 0.9	(0.045)	(0.035)	(0.041)	(0.034)
Doubly-robust DD (Sant'Anna/Zhao 2020)				
Preferred specification: school-level	0.026	0.023	0.090**	0.048
controls + avg student-level controls	(0.038)	(0.031)	(0.036)	(0.036)
Avg student-level controls only	0.019	0.022	0.086**	0.042
	(0.036)	(0.029)	(0.035)	(0.034)

Student-level control variables include gender, Aboriginal identity, disability (2 categories), ESL status, home language (4 categories), and age in months. School-level variables include public/private, French Immersion, and total Kindergarten enrollment. Standard errors (in parentheses) clustered by school location for 2-way FE models. Significance levels * = 0.10, ** = 0.05, *** = 0.01

The third and final panel reports results with the Sant'Anna and Zhao (2020) doubly-robust estimator. Although the standard errors are somewhat larger than for the two-way FE model, the point estimates again range from 0.02 to 0.09 SD, which is not dramatically different from our main results. One possible exception to this broad characterization is the estimated effect on reading scores in the restricted sample: 0.09 SD would clearly qualify as a "medium" effect size under the classification scheme of Kraft (2020) and would be comparable in magnitude to prior estimates on the short-term effect of FDK.

4 Conclusion

This paper provides new and credible estimates of universal FDK on test scores in a large-scale setting. Our results are generally consistent with a small positive average effect of universal FDK

on both numeracy and reading skills in Grade 4. Although the results are marginal in terms of statistical significance, the study has enough power to rule out both large positive effects and large or moderate negative effects. These results are broadly consistent with prior findings in this literature. Moving beyond average effects, we find some evidence of larger effects for specific groups, particularly students who speak English as a second language. These results are consistent with prior research on targeted FDK in B.C. but are distinct from Cannon et al.'s (2011) finding of no effect of FDK beyond the Kindergarten year on English Language Learners in Los Angeles.

Although many North American jurisdictions have already implemented universal FDK and are unlikely to return to the previous policy, some jurisdictions continue to operate half-day Kindergarten. The strong evidence of benefits for ESL students in the B.C. context suggests that these jurisdictions should consider offering targeted FDK for these students. These results further suggest that ESL students may benefit from additional classroom time more generally, either in pre-Kindergarten or in early elementary school.

The average effects that we estimate provide little support for offering FDK as a universal program. Given the relative consistency of the empirical literature on average effects of FDK, jurisdictions that implement universal programs should not expect to see substantial improvements in test scores in the medium term. The potential benefits of universal FDK on long-term outcomes via improving non-cognitive skills remains an open and important question for future research

.

5 References

Abadie, Alberto (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72(1): 1-19.

Almond, Douglas and Janet Currie (2010). Human capital development before age five. In *Handbook of Labor Economics* 4b: 1315-1486.

Bitler, Marianne P., Hilary W. Hoynes and Domina Thurston (2014). Experimental evidence on distributional effects of Head Start. NBER Working Paper 20434, National Bureau of Economic Research.

British Columbia Ministry of Education (2010). *Full Day Kindergarten Program Guide*. Crown Publications, Product Number 7530879353. Accessed June 1, 2020 at https://www.crownpub.bc.ca/Product/Details/7530879353_S.

British Columbia Ministry of Education [creator] (2020a). *Ministry of Education Data*. Population Data B.C. [publisher]. Data Extract, Ministry of Education (2020). https://www.popdata.bc.ca/data/childhood/MED.

British Columbia Ministry of Education [creator] (2020b). *Schools (K-12) with Francophone Indicators*. DataBC [publisher]. Data Extract, Ministry of Education (2020). https://catalogue.data.gov.bc.ca/dataset/schools-k-12-with-francophone-indicators.

Cannon, Jill S., Alison Jacknowitz and Gary Painter (2006). Is full better than half? Examining the longitudinal effects of full day Kindergarten attendance. *Journal of Policy Analysis and Management* 25(2): 299–321.

Cannon, Jill S., Alison Jacknowitz and Gary Painter (2011). The effect of attending full day Kindergarten on English Learner students. *Journal of Policy Analysis and Management* 30(2): 287–309.

Cascio, Elizabeth U. (2009). Do investments in universal early education pay off? Long-term effects of introducing Kindergartens into public schools. NBER Working Paper 14951, National Bureau of Economic Research.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Swchanzenbach and Danny Yagan (2011). How does your Kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, 126(4): 1593-1660.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9):2593-2632.

Child Trends (2013). *Full Day Kindergarten.* Accessed March 24, 2014 at http://www.childtrends.org/?indicators-full-daykindergarten.

Cohen, Jacob (1988). Statistical Power Analysis for the Behavioral Sciences, second edition. Hillsdale, N.J.: Lawrence Erlbaum.

Currie, Janet and Duncan Thomas (1995). Does Head Start make a difference? *American Economic Review* 85(3): 341–364.

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

DeCicca, Philip (2007). Does full day Kindergarten matter? Evidence from the first two years of schooling. *Economics of Education Review* 26(1): 67-82.

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

Dhuey, Elizabeth (2011). Who benefits from Kindergarten? Evidence from the introduction of state subsidization. *Educational Evaluation and Policy Analysis* 3(1): 3-22.

Duncan, Greg J. and Katherine Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives* 27(2): 109-132.

Dhuey, Elizabeth, Jean Eid and Christine Neill (2020). Parental employment effects of switching from half-day to full-day Kindergarten: Evidence from Ontario's French schools. *Canadian Public Policy* 46(1): 145-174.

Elicker, James, and Sangeeta Mathur (1997). What do they do all day? Comprehensive evaluation of a full-school-day Kindergarten. *Early Childhood Research Quarterly* 12(4): 459–480.

Friesen, Jane, Brian Krauth and Mohammed Reza Sattari (2017). The effect of universal full-day Kindergarten on children's behavior. Working paper, Simon Fraser University.

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

Gibbs, Chloe R. (2012). Experimental and quasi-experimental evidence on the impact of full day Kindergarten. Unpublished PhD Dissertation, University of Chicago.

Gibbs, Chloe R. (2014). Experimental evidence on early intervention: The impact of full day kindergarten. Working paper, University of Virginia.

Gibbs, Chloe R., Jens Ludwig and Douglas L. Miller (2013). Does Head Start do any lasting good? Martha J. Bailey and Sheldon Danziger, eds. *Legacies of the War on Poverty.* New York: Russell Sage Foundation: 39-65.

Houngbedji, Kenneth (2016). *ABSDID: Stata module to estimate treatment effect with Abadie semiparametric DID estimator*, Statistical Software Components S458134, Boston College Department of Economics.

Government of British Columbia (2010). *B.C. Reg. 259/2010: Amendments to the Independent School Regulation*. Victoria B.C.: Queen's Printer. http://www.bclaws.ca/civix/document/id/lc/bcgaz2/v53n18_259-2010, accessed April 20, 2020.

Kraft, Matthew A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher* 49 (4): 241–253.

Lee, Valerie E., David T. Burkam, Douglas D. Ready, Joann Honigman and Samuel J. Meisels (2006). Full-day versus half-day kindergarten: In which program do children learn more? *American Journal of Education* 112(2): 163-208.

Ludwig, Jens and Douglas L. Miller (2007). Does Head Start improve children's life chances? Evidence from a regression-discontinuity design. *Quarterly Journal of Economics* 122(1): 159-208.

McCuaig, Kerry (2014), *Policy Update: Full Day Kindergarten in Canada.* Atkinson Centre for Society and Child Development, University of Toronto. Accessed April 20, 2020 at https://www.oise.utoronto.ca/atkinson/UserFiles/File/Policy%20Commentaries/PolicyUpdate-FDKinCanada.pdf.

O'Grady, Kathryn, Marie-Anne Deussing, Tanya Scerbina, Yitian Tao, Karen Fung, Vanja Elez, and Jeremy Monk (2019). *Measuring up: Canadian Results of the OECD PISA 2018 Study*. Toronto: Council of Ministers of Education Canada. Accessed February 23, 2020 at https://www.cmec.ca/Publications/Lists/Publications/Attachments/396/PISA2018_PublicReport_EN.pdf.

Sant'Anna, Pedro H.C. and Jun B. Zhao (2020). Doubly robust difference-in-differences estimators. Journal of Econometrics 219(1): 101-122.

Votruba-Drzal, Elizabeth, Christine P. Li-Grining, and Carolina Maldonado-Carreño (2008). A developmental perspective on full- versus part-day kindergarten and children's academic trajectories through fifth grade. *Child Development* 79(4): 957-978.

Warburton, William P., Rebecca N. Warburton, and Clyde Hertzman (2012). Does full day Kindergarten help kids? *Canadian Public Policy* 38(4): 591-603.