Essays in applied econometrics

by

Reza Sattari

M.A., Simon Fraser University, 2010
B.A., Shahid Beheshti University, 2005

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2016

Approval

Name:	**Reza Sattari**
Degree:	**Doctor of Philosophy (Economics)**
Title:	***Essays in applied econometrics***
Examining Committee:	**Chair:** Nicolas Schmitt Professor

Jane Friesen
Senior Supervisor
Professor

Brian Krauth
Supervisor
Associate Professor

Chris Muris
Supervisor
Assistant Professor

Simon Woodcock
Internal Examiner
Associate Professor
Department of Economics

Abigail Payne
External Examiner
Professor
Department of Economics
McMaster University

Date Defended: 12 August 2016

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Amid growing evidence of the importance of non-cognitive skills for both cognitive skill development and long-term outcomes, understanding the effect of education policies on non-cognitive skill formation is of increasing interest. The first two chapters of this thesis studies the effect of two school interventions on student behavior.

The first paper of this Thesis (Chapter 1) provides the first evidence of the effect of multigrade classes on non-cognitive skills. I exploit strictly enforced class size caps accompanied by centralized funding rules to generate IV estimates of this effect using custom survey data administered to over 15000 parents of Kindergarten and Grade 1 students linked to publicly available administrative data on multigrade classes. I find that placing children in multigrade classes causes significantly more peer relationship problems and hyperactivity compared to single grade classrooms.

The second paper of this Thesis (Chapter 2) my coauthors and I exploit the staggered rollout of universal full-day Kindergarten (FDK) to estimate its effects on children's behavior. Our research design identifies these effects by comparing across-cohort changes in outcomes among early versus late adopting schools. We find little effect of FDK on child behavior or parents' mental health, and an increase in hours worked by parents who are employed part-time. These results hold across a range of child and family characteristics, with one exception. In families who do not speak English at home, FDK reduces child hyperactivity and peer relationship problems, improves parents' mental health and increases employment and hours.

The last paper of this Thesis (Chapter 3) was triggered by a heated debate in the Iranian parliament over the effectiveness of the "1993 Population Control Law". There has been a long debate among economists and policy makers over the effectiveness of population planning programs. The estimated program effects in the literature vary substantially. One such program is the Iranian 1993 Population Control Law that withdrew paid maternity leave and social welfare subsidies in the case of children of fourth and higher parities. My coauthor and I use data from publicly available sample 2006 census data in Iran and the annual Household Expenditure and Income Surveys (HEIS: 1988-2005) to estimate the effect of this policy on fertility outcomes. Our difference in difference method compares the change in probability of having birth in families with fewer than three children prior to the legislation to the change in probability of having birth of families with three or more children. We find that the legislation had a modest effect of 8 to 13 percent on decreasing the probability of a fourth or higher birth. The law has the highest impact after four years of implementation and after that effect size gradually goes away.

Keywords: Multigrade classes; non-cognitive skills; behavior problems; hyperactivity; program evaluation; population control policy

Dedication

To my wife Nazanin,
to our son Ryan,
& to my parents.

Acknowledgements

My deep and sincere gratitude goes first to Professor Jane Friesen who has helped and encouraged me at all stages of my thesis work with great patience and immense care. She has set an example of excellence as a researcher, mentor, instructor, and role model. I am particularly indebted to Professor Brian Krauth and Professor Muris for their invaluable comments, support, and help throughout my PhD program.

I wish to express my gratefulness to Dr. Simon Woodcock, Graduate Chair, Department of Economics for his continuous encouragement, understanding and unlimited support to me and all other graduate students in the Department of Economics.

I would especially like to thank my amazing family for the love, support, and constant encouragement they gave me over the years. I would like to thank my parents, my brother and my sisters.

Last but not least, I would like to thank my lovely wife, Nazanin for her love, advice and energy that kept me running during these years. I would like to thank our son, Ryan the most important person in the world to me and apologize for the time I missed in his life while I was writing this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

The Effect of Multigrade Classrooms on Student Behavior

Abstract

Amid growing evidence of the importance of non-cognitive skills for both cognitive skill development and long-term outcomes, understanding the effect of education policies on non-cognitive skill formation is of increasing interest. This paper provides the first quasi-experimental evidence of the effect of multigrade classes on non-cognitive skills. I exploit strictly enforced class size caps accompanied by centralized funding rules to generate IV estimates of this effect, using custom survey data administered to over 15000 parents of Kindergarten and Grade 1 students, linked to publicly available administrative data on multigrade classes. I find that placing children in multigrade classes causes a substantially increase in peer relationship problems and hyperactivity relative to placement in single grade classrooms.

Keywords: Multigrade classes; non-cognitive skills; behavior problems; hyperactivity; program evaluation; population control policy

1.1 Introduction

Recent evidence suggests both that the effects of many school interventions on cognitive skills fade out quickly (Currie and Thomas, 2000; Ludwig and Phillips, 2007; Duncan and Magnuson, 2013; Zhai et al., 2012; Bitler et al., 2014), and that these same interventions have long-term effects (Chetty et al., 2011; Ludwig and Miller, 2007; Garces et al., 1992; Deming, 2009). A growing literature indicates that the underlying mechanism that links early childhood education programs to long-run outcomes may operate through the development of non-cognitive skills such as persistence, self-control and social aptitude (e.g. Blau and Currie, 2006; Borghans et al., 2008; Cunha et al., 2006; Gibbs et al., 2011; Heckman et al., 2006). Learning how school interventions affect non-cognitive skill development therefore is critical to understanding their longer-term impacts.

This paper studies one such policy: the practice of combining students from different grades into a multigrade class with one teacher. This type of classroom is common in both rural and urban areas in a variety of jurisdictions, including 42 percent of Norwegian primary schools (Mulryan-Kyne, 2005), 25 percent of all British students (Little, 2005), 15 percent of Grade 2 students and 12 percent of Grade 3 students in California (Sims, 2008), and around 12 percent of public schools in United States offering Grade 1 (Thomas, 2012). By combining students with different school tenure and ages, multigrade classes may affect children's skill development directly via peer effects, or indirectly if teachers modify their teaching methods in response to classroom composition (Bedard and Dhuey, 2006; Sims, 2008; Leuven and Rønning, 2014).

Previous research on the effects of multigrade classes on non-cognitive skills focuses on measures of student self-concept and attitude towards school. The results suggest a small positive effect (see Veenman (1995) for a survey of this evidence). Similarly, early studies that look at the effect on cognitive measures conclude that students in multigrade classrooms perform no worse and sometimes better in terms of test scores. Both groups of studies, however, fail to address the non-random selection of students and teachers into multigrade classes.

Several more recent studies use quasi-experimental methods to address selection into multigrade classes. Using an instrumental variable strategy based on variation in student enrollment levels around the class size cap imposed by the California Class Size Reduction Program, Sims (2008) shows that placement in multigrade classes reduces test scores in Grades 2 and 3. Thomas (2012) employs a school fixed effects method with the ECLS-K data and finds that Grade 1 students who are assigned to multigrade classes do no worse relative to those placed in single grade classes. Leuven and Rønning (2014) exploit discontinuous grade mixing rules in Norwegian junior high schools in an instrumental variables setup. They conclude that the presence of lower grade peers decreases achievement, while the presence of peers from higher grades increases achievement.

None of these quasi-experimental papers considers the effect of multigrade classes on children's behavior. This paper addresses this important gap in the literature by providing the first quasi-experimental estimates of the causal effect of multigrade classes on non-cognitive skills. The analysis is based on custom survey data for over 15,000 parents of Kindergarten and Grade 1 students in British Columbia, Canada, linked to publicly available administrative data on school-by-grade measures of the share of students in multigrade classes. The survey includes widely used questions on "externalizing" behavior problems related to hyperactivity, aggression, willfulness and problems in peer relationships, as well as "internalizing" behavior problems related to anxiety and depression. Externalizing behavior problems in children have been linked to lower educational attainment, employment and earnings (e.g. Caspi et al., 1998; Farmer, 1993, 1995). Attention deficit/hyperactivity disorder (ADHD), a subcategory of externalizing behavior problems, has been linked to subsequent criminal behavior, social assistance receipt and negative school and labor market outcomes (Currie and Stabile, 2006; Currie et al., 2010; Currie and Almond, 2011; Fletcher and Wolfe, 2008, 2009; Fletcher, 2014). These measures of behavior problems enable us to directly investigate a key channel by which multigrade classes may affect long-run outcomes.

A number of important institutional features of British Columbia (B.C.)'s education system make it possible to use a highly credible research design to estimate this effect. Strictly enforced class size caps accompanied by centralized funding based on full-time equivalent enrollment create discontinuous variation in the percentage of students in multigrade classes at multiples of the class size cap. While previous authors have used discontinuities induced by similar class size rules in other jurisdictions as a source of identification for class size and multigrade effects on student achievement (Urquiola, 2006; Angrist and Lavy, 1999; Leuven et al., 2008; Fredriksson et al., 2013; Angrist et al., 2014; Leuven and Rønning, 2014), the data show that many schools in B.C. are able to control enrollment levels, so that variation in enrollment around the class size cap cannot be treated as exogenous.[1] Instead, I exploit the fact that the incentive to create a multigrade class does not depend on enrollment in a single grade, but varies with the number of students in two or more adjacent grades relative to multiples of the class size cap for each grade. I develop an instrumental variable strategy that uses variation in a multigrade measure of "excess" enrollment over ranges of grade-level enrollment that are unlikely to be manipulated by school principals (i.e. away from the vicinity of the class size caps). The identifying variation in this strategy comes from the interaction between class size limits and variation in enrollment in multiple grades in relation to those limits, which I demonstrate is plausibly exogenous.

I find that multigrade classrooms significantly increase children's behavioral problems. Specifically, students in multigrade classes on average have 0.98 of a standard deviation more peer relationship problems and exhibit 0.90 of a standard deviation more behaviors

[1] Urquiola and Verhoogen (2009) show that manipulation of enrollment around the class size caps in Chile invalidates the RD design and argue that these results should be interpreted with caution.

that are associated with hyperactivity. The effect of multigrade classes on peer relationship problems and hyperactivity is twice as large for boys as for girls. There is also weak evidence that multigrade classes increase internalizing behavior problems (the effect size for an average student is 0.66 of a standard deviation). Girls appear to show more internalizing behavioral problems when placed in multigrade classes. This effect size on girls is more than one standard deviation, compared to 0.39 standard deviation for boys.

1.2 Institutional Background

The provincial government establishes the amount of grant funding for public education annually and allocates these funds to each of 60 school districts (Ministry of Education, British Columbia, 2015). Operating funds are allocated according to a formula based primarily on full-time equivalent enrollment, with supplements for English as a Second Language, Aboriginal, and low-incidence special needs students, as well as institutional factors such as rapid enrollment changes, etc. The provincial Ministry of Education administers the funding formula and sets curriculum and other standards, while local school districts are responsible for implementation and resource allocation based on local spending priorities.

Each school has a catchment area consisting of a geographic area around the school. Students who do not reside in the catchment area of a school can apply by a designated date to attend that school. Principals determine (based on space and other resources available at the school) if and how many out-of-catchment students they will admit. In general the priority for enrollment is given in the following descending order: a catchment area child who, in the previous school year, attended the school at which the educational program is made available; a catchment area child; a non-catchment area child; and a non-school district child (Ministry of Education, British Columbia, 2014). Within these categories, principals have discretion over which students to admit.

Throughout the relevant period, all classes that include Kindergarten students are strictly capped at 22 students, and classes that include Grade 1 students (and no Kindergarten students) are capped at 24 students. All classes are taught by B.C. certified teachers. While children are eligible and expected to enroll in Kindergarten in September of the calendar year in which they turn five, schooling is not compulsory until September of the calendar year in which they turn six. In practice, nearly all children in B.C. attend Kindergarten.

1.3 Data

The primary data source for this analysis is the B.C. School Arrangements Survey (BCSAS), a custom survey administered to over 15,000 parents of Kindergarten and Grade 1 children in 2011, 2012 and 2013.[2] Twenty-one public school districts agreed to participate in the survey, including the two largest districts of Surrey and Vancouver. All public schools with Kindergarten enrollment in participating districts received surveys to distribute to families. The 2010/11 and 2011/12 surveys covered both Kindergarten and Grade 1 students, while the 2012/13 survey covered Grade 1 students only.

The key outcome variables measured by this survey are derived from responses to a set of 26 questions about children's behavior and emotional health. These questions are drawn from two sources: 23 items from the Behavior Problems Index (BPI) and 8 items from the Canadian National Longitudinal Survey of Children and Youth (NLSCY). These items are used to construct counts of the number of relevant items to which the parent responds with “sometimes true" or “often true" rather than “not true". Each item refers to some problematic behavior (e.g. “is disobedient at home"), so a higher score implies more behavior problems. Eighteen items are used to construct an overall BPI externalizing behavior scale and four BPI externalizing behavior subscales that measure hyperactivity, antisocial behavior, headstrong behavior and peer relationship problems. An internalizing behavior (anxiety/depression) score is constructed using the eight NLSCY items. To facilitate comparison of effect sizes among the regression results, all scales are standardized to have zero mean and unit variance within each grade. BCSAS contains information regarding whether an individual student is in a multigrade class when they were in Kindergarten.

My secondary data comes from the administrative records of the B.C. Ministry of Education. These records include publicly available school-level data for all public and private schools in the province. Except where specified otherwise, I restrict my attention to students enrolled in public schools administered by a school district that participated in the BCSAS survey. The school/grade-level data includes school name, type, location, and enrollment headcounts by classroom, which I use to construct average class size and number of students in multigrade classes, by grade and year.

The main BCSAS sample includes all returned surveys that could be matched with a school. The overall response rate was approximately 16.6%. Table 1.1 presents summary statistics for the main BCSAS sample. As reported in Friesen et al. (2015), the BCSAS sample does a reasonable job of capturing the diversity of the underlying population, matching well on gender, student age, and the characteristics of the classroom. While the admin data show that 11.3% of Kindergarten and 32.6% of Grade 1 students were in multigrade classes, 12.3% of Kindergarten students among BCSAS respondents are reported to be in multigrade classrooms. The average class size for Kindergarten and Grade 1 students in schools covered in the BCSAS is 19 and 20.9, respectively. Children in full-day Kindergarten (FDK) are slightly underrepresented among BCSAS respondents (68.4% versus 71.0%). It also appears that Aboriginal students are underrepresented (5.3% versus 8.5%), as are mi-

[2] This survey was developed and administered by Jane Friesen and Brian Krauth at Simon Fraser University.

nority home language students (18.2% versus 26.4%). However, responses to questions about Aboriginal identity and home language may be context specific and are known to vary from year to year for the same student within the administrative data, so that differences in response rates according to these characteristics should be interpreted with some degree of caution. The share of parents with a high school education or less is also low relative to their share in Census population data. Overall, this pattern is consistent with lower response rates among relatively disadvantaged parents and among parents who may face language or cultural barriers to survey participation.

1.4 Empirical Strategy

I am interested in estimating the effect of being in a multigrade classroom on students' emotional health and behavioral outcomes. I start with the following specification:

$$Y_i = \beta_0 + \beta_1 MG_i + X'_{i,s(i)}\Gamma + \mu_{t(i)} + \epsilon_{i,s(i),t(i)} \tag{1.1}$$

where Y_i is an outcome measure of student i, and the treatment variable MG_i, is an indicator that student i is in a multigrade classroom, and $X_{i,s(i)}$ controls for student background characteristics including gender, age, parent's education, lone parent, Aboriginal identity and ESL, and for school characteristics, including percentage of students with Aboriginal identity, percentage with a disability, percentage from English speaking families, percentage in full-day Kindergarten and enrollment in grades K-3 to control for school size. Finally, to control for unobserved time invariant characteristics I include a year effect denoted by $\mu_{t(i)}$; $\epsilon_{i,s(i),t(i)}$ is the error term.

In my primary data source (BCSAS), I observe individual treatment status (MG_i) for Kindergarten students, so I can estimate equation (1.1). However, the BCSAS data do not include a student level indicator for individual treatment status for Grade 1 students. I therefore aggregate all variables at the grade/school/year level and estimate the following equation:

$$Y_{sgt} = \beta_0 + \beta_1 MG_{sgt} + X'_{sgt}\Gamma + \mu_t + \epsilon_{sgt} \tag{1.2}$$

where Y_{sgt} and X_{sgt} are school/grade/year level averages[3] from the BCSAS data and MG_{sgt} is the percentage of students in multigrade classes at the grade/school/year level.[4]

[3]Equation (1.2) is derived from equation (1.1) and the school/grade/year level variables are calculated as:

$$W_{sgt} = \frac{\sum_i \mathbf{1}_{\{s(i)=s,g(i)=g,t(i)=t\}} \times W_i}{\sum_i \mathbf{1}_{\{s(i)=s,g(i)=g,t(i)=t\}}}$$

[4]Since MG_i is reported by parents, it could potentially be subject to measurement error. However, if parents do not systematically over or under report MG_i, then we expect the following:

$$\mathbf{E}[MG_i|Z_{st}, X_{sgt}] = \mathbf{E}[MG_{sgt}|Z_{st}, X_{sgt}]$$

The key challenge is to find plausibly exogenous variation in treatment status that can be used to identify the causal effect of being in a multigrade classroom on the outcomes of interest. There are several potential sources of non-random selection into multigrade classrooms that must be addressed. Schools may be more likely to adopt multigrade classes if they have had a successful experience in the past, and/or schools with less vulnerable students may be more likely to offer multigrade classes. At the student level, students may be assigned to multigrade classrooms based on their unobserved characteristics. For example, schools may tend to assign students who would benefit from remedial work to combination classes with students from a lower grade.

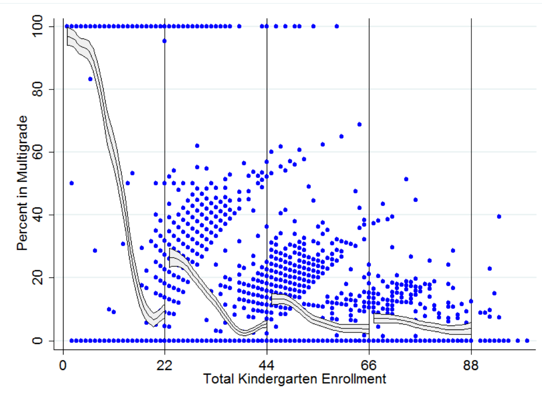
To address these issues, I exploit institutional features of the B. C. education system that influence the incentive to offer multigrade classes. Strictly enforced class size caps accompanied by enrollment based funding give schools strong incentives to form multigrade classes when doing so can reduce the number of classrooms required to accommodate all enrolled students. One of the consequences of this incentive is illustrated in Figure 1.1, which shows the percentage of students in multigrade classes in schools with different Kindergarten enrollment levels. We see discontinuous jumps in the percentage of students in multigrade classes in a small neighborhood around multiples of the class size cap. In principle, one could exploit these discontinuities to generate causal estimates of the effects of interest, using a Fuzzy Regression Discontinuity Design. The key identifying assumption in any RD design is that the conditional expectations of the potential outcomes are continuous in assignment variables at the threshold. This requires that schools are not able to precisely manipulate enrollment (McCrary, 2008).

Figure 1.2 presents a histogram of Kindergarten enrollment. We see that a substantial number of schools report enrollment levels that are exactly equal to multiples of the class size caps. This pattern suggests that popular schools may accept out-of-catchment students as long as they have space available, but not when doing so would cause them to exceed the class size cap. Figure 1.3 shows that parents of students at schools with enrollments equal to multiples of the class size cap have higher annual income and education levels on average than parents of students at schools that are just to the right of the cutoff points. These patterns strongly invalidate the RD approach in this context.

Instead, I develop an instrumental variable strategy based on plausibly exogenous variation in multigrade classes that is driven by variation in enrollment in different parts of the enrollment distribution, away from the multiples of class size cap. The possibility that schools can avoid running an additional and costly class by forming a multigrade class depends not only on the number of students enrolled in a given grade, but also on the number of students enrolled in adjacent grades, relative to the class size cap.

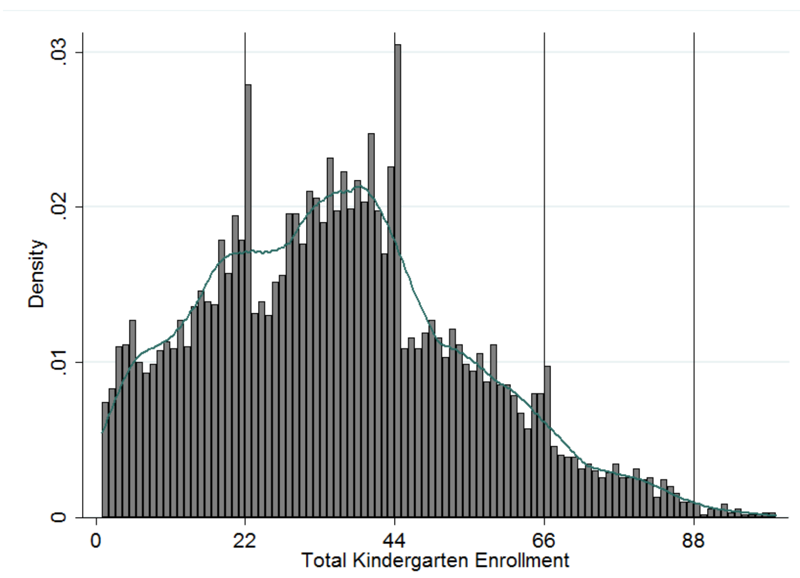
To test this, I regress MG_i on MG_{sgt} and the null hypothesis of $(H_0 : \alpha = 1)$ cannot be rejected. This suggests there is no systematic pattern in the values of MG_i reported by parents.

Figure 1.1: Scatterplot of the percentage of Kindergarteners in multigrade classes by total Kindergarten enrollment; 2007/08 - 2013/14.



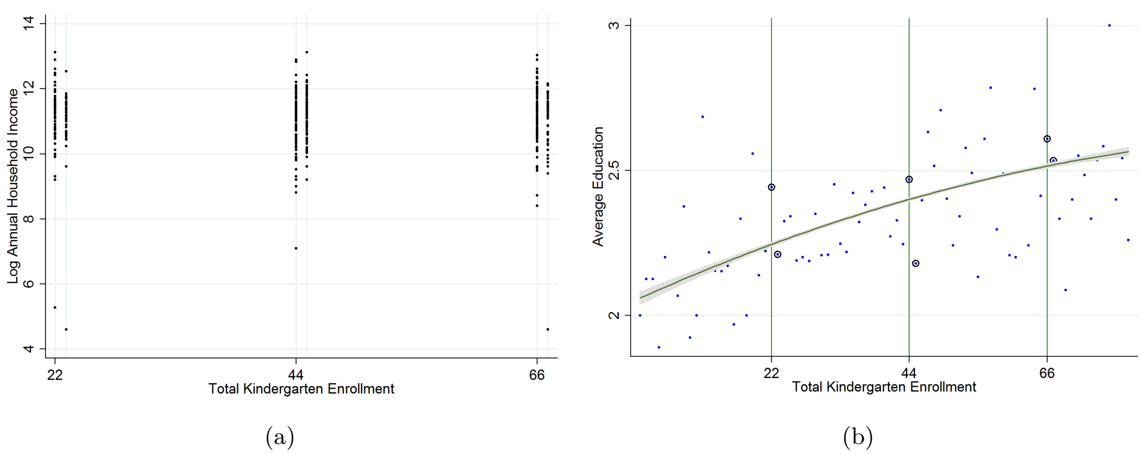
Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.
Notes: each point represents a school/year - local polynomial fits with 95% bandwidths show the relationship between variables.

Figure 1.2: Histogram of Kindergarten enrollment in British Columbia public schools; 2007/08 - 2013/14.



Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.

Figure 1.3: Comparison of parental income and education level of Kindergarten students.



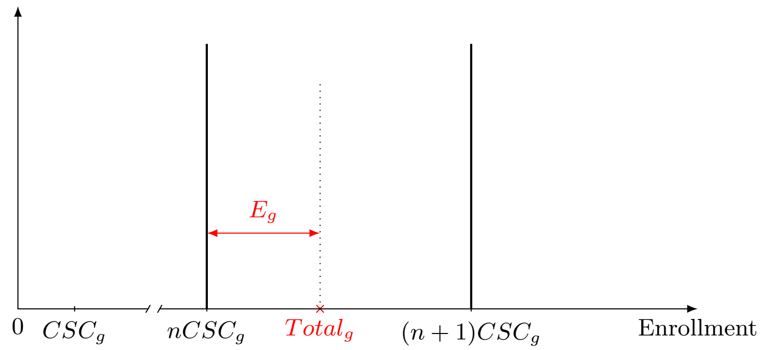


Figure 1.4: Excess Students in a given grade.

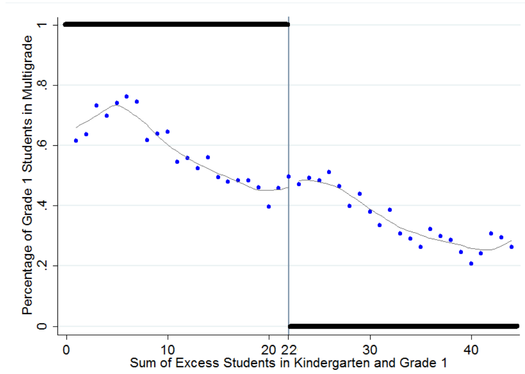
Define E_{sgt} as the number of "excess" students in school s in grade g at time t:

$$E_{sgt} = Total_{sgt} - CSC_g \times int\Big(\frac{Total_{sgt}}{CSC_g}\Big)$$

where $Total_{sgt}$ is total student enrollment and CSC_g is the relevant class size cap for that grade (22 for Kindergarten and 24 for Grades 1 and 2). Figure 1.4 provides a visual representation of this variable. If the sum of excess students in two adjacent grades is less than or equal to the class size cap ($E_{sgt} + E_{sg+1t} \leq CSC_g$), schools can save a costly class by combining the excess students into a single multigrade class. For example, if there are 33 Kindergarteners and 34 Grade 1 students in a school, instead of offering two Kindergarten and two Grade 1 classes, the school can offer one Kindergarten class and one Grade 1 class (with 22 and 24 students, respectively) and one class with 11 Kindergarteners and 10 Grade 1 students. However, if there were 36 Grade 1 students, the school cannot combine the excess students into a single classroom, since no Kindergarten student can be in a class with more than 22 students in it.

Again, these rules suggest that there may be a discontinuous relationship between the number of excess students in adjacent grades at the class size caps. Sims (2008) uses a binary multigrade predictor for a given grade that would take the value 1 if the number of classes required to serve students in that grade and students in the grade below in single grade classes is greater than the total number of classes required to serve students in both grades together, and zero otherwise. Figure 1.5 illustrates this predictor for Grade 1, using the number of excess students in Kindergarten and Grade 1. In the B.C. administrative data, the relationship between percentage of Grade 1 students in multigrade classes and the number of excess students in Kindergarten and Grade 1 does not show a discontinuous change as Sims (2008)'s multigrade indicator would predict. However, there is a clear relationship between this sum and the percentage of students in multigrade classes. If we exclude cases where grade level enrollment is exactly equal to a multiple of class size cap, we can treat this variation as exogenous.

Figure 1.5: Scatterplot of the average percentage of Grade 1 students in multigrade classes by the sum of excess students in Kindergarten and Grade 1; 2007/08 - 2013/14.



Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.

The probability that Kindergarten and Grade 1 students are placed in a multigrade class also varies with Grade 2 enrollment. For example, if $E_{sKt} + E_{s1t} > 22$ and $E_{s1t} + E_{s2t} > 24$, but $E_{sKt} + E_{s1t} + E_{s2t} \leq 46$, a school can offer a Kindergarten-Grade 1 combined class and a Grade 1-Grade 2 combined class rather than three single grade classes.[5]

I therefore use the sum of excess students in grades K-2 as an instrument to capture exogenous variation in multigrade classes:

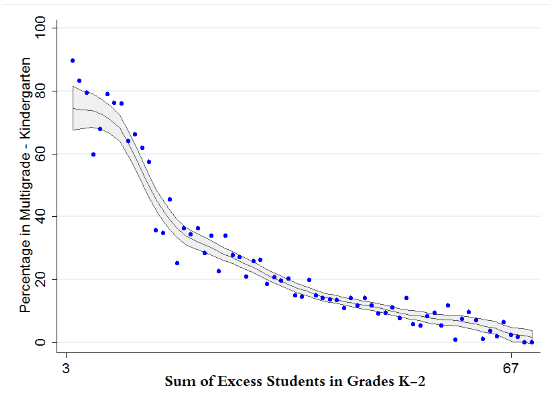
$$Z_{st} = E_{sKt} + E_{s1t} + E_{s2t}$$

Holding enrollment in Kindergarten constant, a small change in Grade 1 and Grade 2 enrollment affects the probability that Kindergarten students are in multigrade classrooms. In other words, the variation in this instrumental variable strategy comes from the interaction between class size limits and enrollment in all three grades in relation to those limits. Panels (a) and (b) in Figure 1.6 illustrate how the average percentage of students in Kindergarten and Grade 1 that are in multigrade classes changes in relation to the changes in the instrument. In schools where there a small number of excess students, multigrade classes are very appealing. As the number of excess students increases, multigrade classes are less common.

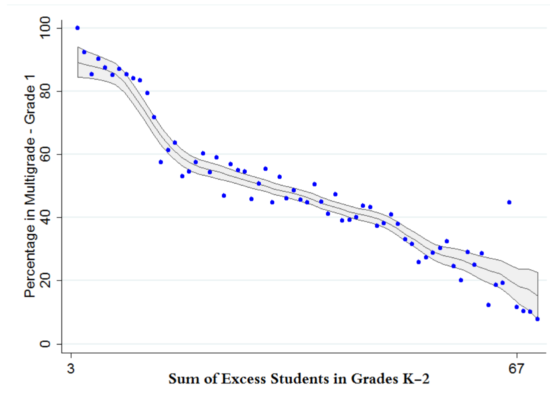
Validity of instrumental variable estimation also requires that any effect of the instrument on the outcome variables must be through an effect on the endogenous variable. The fundamental problem with testing the exclusion restriction is that it involves the structural error which is never observable. However, if unobserved student and school characteristics are correlated with the instrument, one might also expect to find a relationship between

[5]It is also true in principle that excess student from higher grades matter. In the data, however, including excess students in Grade 3 slightly weakens the instrument.

Figure 1.6: Scatterplot of the average percentage of Kindergarten (a) and Grade 1 (b) students in multigrade classes by the sum of excess students in Grades K-2 (the instrument); 2007/08 - 2013/14.



(a)



(b)

Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.

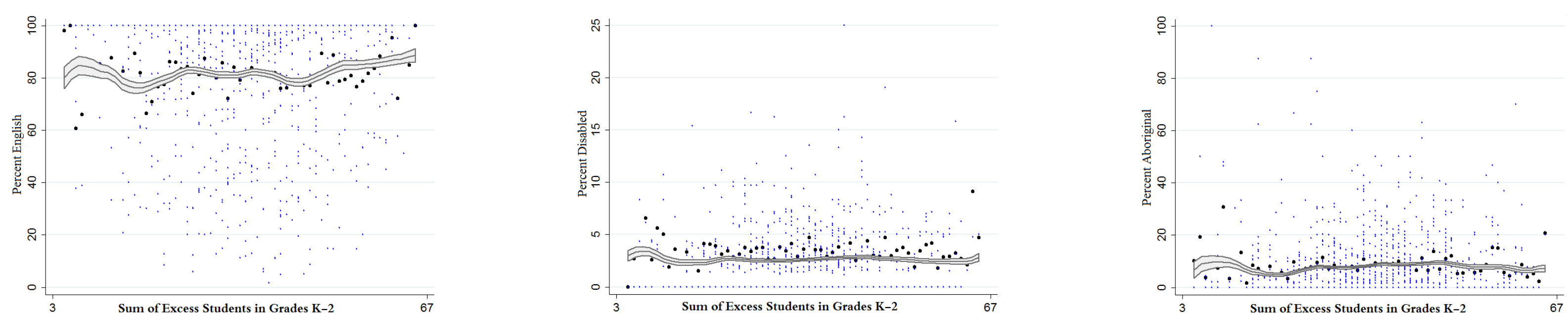
Note: Local polynomial fits with 95% bandwidths show the relationship between variables.

the instrument and observable student and school characteristics. Figure 1.7 illustrates the relationship between the proposed instrument and parents' age, education level and log of annual household income, respectively. Figure 1.8 illustrates the relationship between the number of excess students in grades K-2 and school characteristics, including percentage of students with Aboriginal identity, percentage with disability and percentage from English speaking families, respectively. We do not see a clear pattern in these graphs, suggesting that the instrument is not affecting the outcome variables through any observable characteristics other than multigrade classes.

Figure 1.7: Scatterplot of the socioeconomic status of parents by the sum of excess students in Grades K-2 (the instrument).

Note: each point represents a student in BCSAS data; local polynomial fits with 95% bandwidths show the relationship between variables.

Figure 1.8: Scatterplot of School Characteristics by the Sum of Excess Students in Grades K-2 (the Instrument).



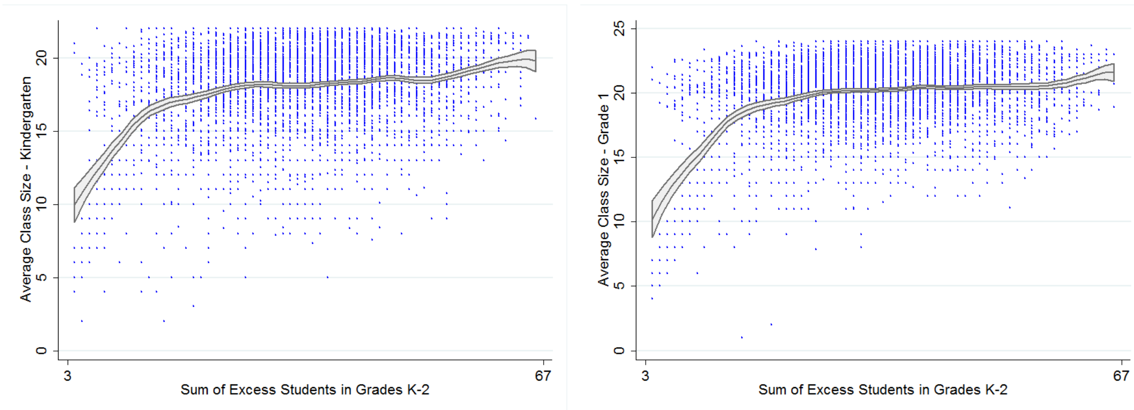
Note: each point represents a school/year; local polynomial fits with 95% bandwidths show the relationship between variables.

It is important to note that my main specification does not control for class size. Forming multigrade classrooms will likely result in changes in average class size, despite holding total grade enrollment constant. In practice, average class size for a given grade, g, will increase by the actual number of students from grade g' that are combined with grade g, divided by total number of classes serving students in grade g. There is evidence that smaller classes slightly improve student behavior (Dee and West, 2011; Finn et al., 1989; Finn, 1998; Finn et al., 2001, 2003; Fredriksson et al., 2013). To the extent that multigrade classes are correlated with average class size, the estimated coefficients capture both the direct effect of multigrade classes and their indirect effect via change in class size.[6]

As long as class size is not correlated with the sum of excess students in Grades K-2, I should be able to consistently estimate the effect of multigrade classes on student behavior. Panels (a) and (b) in Figure 1.9 illustrate the relationship between the instrument and the average class size in Kindergarten and Grade 1, respectively. There seems to be a strong positive correlation between the number of excess students in Grades K-2 and the average class size, especially when the sum of excess students is small. This positive correlation could be driven by small schools, where there are low levels of grade enrollment. In Figure 1.10 I mark the schools, where grade enrollment is less than 5 by a different color. The positive correlation between the instrument and average class size disappears when I exclude the small schools from my analysis. The reason for this pattern could be small schools, mostly located in distant rural areas, combine students from all three grades. As a result, class size will change linearly with the number of excess students in those schools.

[6] Leuven and Rønning (2014) are able to separately identify the effects of multigrade classes and class size due to the strict structure of the grade mixing rules in Norway that affect both grade composition and class size. Moreover, they limit their sample to school districts that are located in rural areas, where variation in grade enrollment is very likely to be exogenous, allowing them to find a valid instrument for class size.

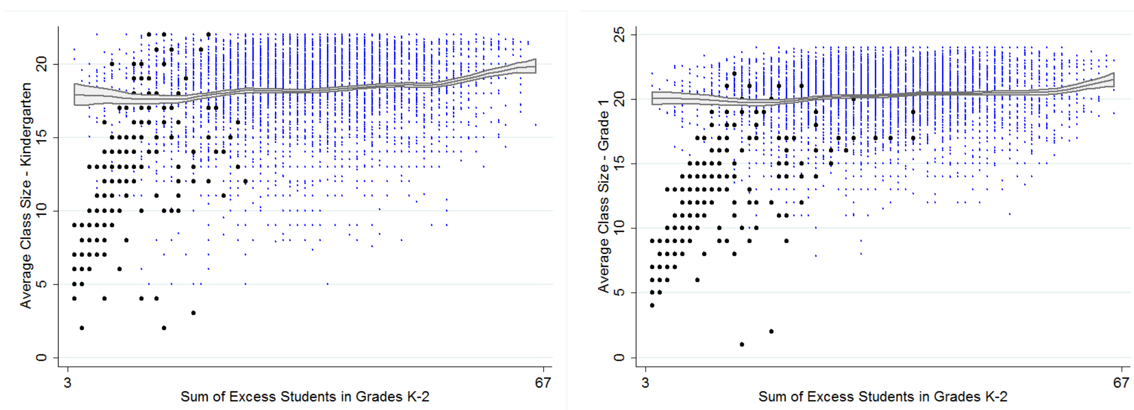
Figure 1.9: Scatterplot of the average class size by the sum of excess students; 2007/08 - 2013/14.



(a) (b)

Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.
Notes: each point represents a school/year - local polynomial fits with 95% bandwidths show the relationship between variables.

Figure 1.10: Scatterplot of the average class size by the sum of excess students, when excluding schools with less than 5 grade enrollments; 2007/8 - 2013/14.



(a) (b)

Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.
Notes: each point represents a school/year - schools with less than 5 enrollment in each grade are represented by larger black marks - local polynomial fits with 95% bandwidths show the relationship between variables.

1.5 Results

1.5.1 First Stage Estimates

Table 1.2 reports the results from the first stage regressions corresponding to my main results. The dependent variable in the first column is MG_i, an indicator that student i is in a multigrade classroom and in the second column the dependent variable is MG_{sgt}, the percentage of students in a given school, grade and year in multigrade classes. The explanatory variables include the instrument (the sum of excess students in Grades K-2) as well as the control variables and cohort/year fixed effects. The F-statistic on the excluded instrument for both specifications is above 38. This result indicates that the proposed instrument is not weak (Stock et al., 2002).

To further investigate the exclusion restriction, I estimate a set of placebo tests using the sum of excess students in grades K-2 as an instrument for observable student and school characteristics. The F-statistics for all the student and school characteristics (reported in the bottom two panels of Table 1.3) are less than 2, indicating that the variation in the instrument has no power in explaining the variation in those characteristics.

1.5.2 Main Results

My main results for Kindergarten behavior are reported in Table 1.4. All regressions include cohort/year fixed effects, and heteroskedasticity-robust standard errors are clustered at the school level.

The relationships between externalizing behavior problems and background characteristics are similar in sign and magnitude to comparable results in the ECLS-K data (see Duncan and Magnuson, 2011). Girls exhibit fewer externalizing behavior problems than boys (by 0.29 of a standard deviation), children in minority language families exhibit fewer than those speaking English at home (by 0.12 of a standard deviation), and children of lone parents exhibit more than children in multi-parent families (by 0.24 of a standard deviation). The difference between Aboriginal and non-Aboriginal children is 0.34 of a standard deviation, similar to the black-white difference in the ECLS-K. Conditional on these characteristics, parent's education is not an important predictor of externalizing behavior problems.

The results in the first column of Table 1.4 show that the effect of being in a multigrade classroom in Kindergarten on externalizing behavior is substantial (0.44 of a standard deviation) but statistically insignificant. The 90% confidence interval (-0.20, 1.09) allows us to reject the null that multigrade causes any substantial improvement in externalizing behavior problems for the average child in the sample.

The next four columns of Table 1.4 report results for the four BPI externalizing subscales. The peer relationship problems subscale refers to difficulties engaging and getting along with other children; the hyperactive subscale includes attention, impulsivity and restlessness; the headstrong subscale includes nervousness, disobedience at home, stubbornness, and tendency to be argumentative; and the antisocial subscale includes cheating, lying, bullying and not getting along with teachers. Kindergarten students who are in multigrade classrooms have more peer relationship problems and are more hyperactive than those in single grade classes. The effect size for the peer relationship problems subscale is 0.99 of a standard deviation. The hyperactivity subscale effect is also large, 0.91 of a standard deviation. Both of these effects are statistically significant at the 5% significance level.

Table A1 describes the frequency distribution of externalizing behavior as well as its peer relationship problems and hyperactivity subscales for Kindergarten students in BCSAS data calculated by Friesen et al. (2015). The maximum number of problems reported by parents are 3 and 5 for peer relationship problems and hyperactivity, respectively. The effect size for peer relationship problems is equivalent to parents reporting on average one more problem in the questionnaire. The effect size for hyperactivity is equivalent to parents reporting 2-3 more problems. The effects on the headstrong and antisocial subscales range from small negative to moderately positive, but are not statistically significant at conventional significance levels.

The last column presents results for the anxiety/depression scale. The items in this scale refer to aspects of so-called internalizing behavior, including the extent to which the child is fearful, anxious and sad. Differences by gender, Aboriginal identity and lone parenthood are much less pronounced than for externalizing behavior. Multigrade classrooms are associated with a 0.67 of a standard deviation increase in emotional problems, and this effect is significant at the 10% significance level.

Figure 1.12 plots the point estimates and 90% confidence intervals for both OLS and 2SLS results for all six outcome measures. The OLS results show small (all positive and less than 0.05 of a standard deviation) and insignificant effects of multigrade classrooms on the average student's behavior. Comparison of IV and OLS estimates show that schools with students who have more behavior problems are less likely to adopt multigrade classes. The small and insignificant OLS estimates are similar to those reported in Veenman (1995).

Results for the aggregated specification (equation (1.2)) are presented in Table 1.7. The standard deviation of each of the outcome variable is reported in brackets. Consistent with the first specification, multigrade is associated with a greater number of externalizing and internalizing behavior problems. The top panel in Table 1.7 shows results for the pooled sample of Kindergarten and Grade 1 students. The effect size for externalizing behavior problems is 0.36 (0.61 of a standard deviation). Although this effect is not statistically significant at conventional significance levels, the 90% confidence interval of (-0.02, 0.76) rules out a wide range of behavioral improvements (see Figure 1.13). Internalizing behavioral problems also increase by 0.39 (0.64 of a standard deviation). As with externalizing behavior, this effect is statistically insignificant, but the 90% confidence interval of (-0.01, 0.79) assures us that it is very unlikely that multigrade classes cause students to have fewer behavioral problems.

When I estimate the model for Kindergarten students only, the effect is larger and significant. The bottom panel in Table 1.7 reports separate results for Kindergarten and Grade 1 students. For Kindergarten students the effect size for externalizing behavior problem is 0.75 (1.27 standard deviations) and for internalizing behavior problems is 0.76 (1.24 standard deviations). In contrast, these effects are relatively small and insignificant on average for Grade 1 students. For Grade 1 students the effect size for externalizing behavior problem is 0.18 (0.30 of a standard deviation) and for internalizing behavior problems is 0.15 (0.24 of a standard deviation). One potential explanation for this relatively small effect could be that unlike Kindergarten, students in Grade 1 can be in multigrade classroom with younger students. In Kindergarten, the only possible way to be in a multigrade class is through being combined with older Grade 1 students. If being in a multigrade classroom with younger versus older students has opposite effects, these effects could cancel out each other.

Finally, to investigate heterogeneity in the effect of multigrade classes, I separately estimate specification (1.1) for boys and girls in Kindergarten. The results presented in Table 1.8 show that the effect of multigrade classes on externalizing behavior problems is larger for boys than for girls. The effect of multigrade classes on peer relationship problems and hyperactivity for boys is twice as large as the effect for girls. Boys who are assigned to multigrade classes show 1.38 and 1.22 standard deviation increase in their peer relationship problems and hyperactivity, respectively. There is also weak evidence that multigrade classes increase internalizing behavior problems (the effect size for an average student is 0.67 of a standard deviation). Girls appear to show more internalizing behavioral problems when placed in multigrade classes. This effect on girls is more than one standard deviation increase, compared to 0.39 standard deviations for boys.

1.6 Conclusion

Schools have an incentive to offer multigrade classes for pedagogical reasons or when doing so lets them save resources. This paper provides the first quasi-experimental evidence that combining students from different grades into a single classroom may affect their behavior. My main results show that placing students in multigrade classrooms causes more behavioral problems. Specifically, they show substantially more peer relationship problems and behavioral problems associated with hyperactivity. A growing literature demonstrates a strong relationship between children's behavior and long-term outcomes. This evidence suggests that the widespread use of multigrade classes may have important unintended consequences.

While my approach provides credible estimates of the effect of multigrade classes, my estimation strategy does not reveal the underlying mechanisms that drive these effects. If the costs of multigrade classes result from lack of teachers' attention and input, schools may be able to mitigate this unintended consequence of multigrade classes by providing a teacher's aide in those classes. Future research can focus on separately identifying the direct peer effect channel that works through relative age versus teachers' input and teaching method channel.

Tables and Figures

Table 1.1: Summary statistics, students in Kindergarten and Grade 1.

Variable	BCSAS data	
	Kindergarten (std dev)	**Grade 1 (std dev)**
Student characteristics:		
Male	52.2%	51.2%
Aboriginal identity	5.1%	5.5%
In a multigrade class	12.3%	
Language spoken at home is not English	17.4%	18.7%
Average age in months	62.9	74.6
	(4.7)	(5.0)
School/classroom characteristics:		
In a multigrade class	11.3%	32.6%
Average class size	19.04	20.95
	(2.2)	(1.9)
Enrolled in full-day Kindergarten	79.3%	55.8% *
Enrolled at out-of-catchment school	25.2%	26.5%
Enrolled in French Immersion program	14.9%	15.9%
Responding parent characteristics:		
Lone parent	8.2%	8.8%
Average age in years	38	39.1
	(5.9)	(6.1)
Highest level of education:		
HS dropout	3%	2.8%
HS graduate	12.5%	12.5%
Some post-secondary	42.1%	41.4%
Bachelor's or higher	42.3%	43.3%
Number of Observations (Students)	7030	8817
Number of Observations (School-Years)	893	1274

Note: The BCSAS data cover Kindergarten students in 2010 and 2011, and Grade 1 students 2010, 2011, and 2012, for the 21 public school districts that participated in the BCSAS in all three years.
* When in Kindergarten.

Table 1.2: First stage estimates.

	MG_i (Kindergarten)	MG_{sgt} (Pooled Grades)
Sum of Excess Students in Grades K-2 (Z_{st})	$-0{\cdot}003^{***}$	$-0{\cdot}0051^{***}$
	(0·0006)	(0·0006)
Total Grade Enrollment	$-0{\cdot}0011^{***}$	$-0{\cdot}0024^{***}$
	(0·0002)	(0·0001)
Female	0·01	−0·018
	(0·008)	(0·024)
Aboriginal	0·0063	
	(0·0215)	
Age In Months	0·0014	$0{\cdot}0136^{***}$
	(0·0009)	(0·0014)
Lone Parent	−0·0232	−0·0279
	(0·0148)	(0·0404)
Parent Education	0·0209	0·0139
	(0·015)	(0·0122)
Respondent is Not Mom	0·0237	0·0102
	(0·015)	(0·0391)
English Spoken at Home	$-0{\cdot}0456^{***}$	
	0·0167	
French Immersion	0·015	$0{\cdot}1419^{***}$
	(0·0256)	(0·0256)
% FDK by School/Grade/Year	−0·0002	$-0{\cdot}0711^{***}$
	(0·0002)	(0·0235)
% Disabled by School/Grade/Year	−0·003	0·0782
	(0·0027)	(0·1533)
% Aboriginal by School/Grade/Year	$0{\cdot}0027^{**}$	$0{\cdot}298^{***}$
	(0·0011)	(0·0698)
% English by School/Grade/Year	$-0{\cdot}0007^{*}$	$-0{\cdot}0578^{*}$
	(0·0004)	(0·0297)
F-statistic (excluded instrument)	38.02	76.84
Number of Observations	5365	1835
(Unit of Observation)	(Student)	(School/Grade/Year)

Notes: Dependent variables are individual student multigrade status and percentage in multigrade by School/Grade/Year. All regressions include year effects and heteroskedasticity robust standard errors are clustered at school level.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 1.3: Effect of the instrument on multigrade status, school and student background characteristics.

Variable	**Grade**	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,518)	Prob>F
Individual multigrade status (MG_i)	Kindergarten	0.059	0.059	0.009	35.306	0.000
% in multigrade classes (MG_{sgt})	K+G1	0.207	0.207	0.034	50.087	0.000
% Kindergarteners in multigrade classes (MG_{sKt})	Kindergarten	0.164	0.164	0.0342	41.032	0.000
% Grade 1s in multigrade classes (MG_{s1t})	Grade 1	0.238	0.238	0.042	33.356	0.000
% Disabled by school/grade/year	K+G1	0.030	0.03	0.000	0.022	0.89
% Aboriginal by school/grade/year	K+G1	0.114	0.114	0.000	0.219	0.64
% English by school/grade/year	K+G1	0.09	0.09	0.001	0.899	0.343
log Income	K+G1	0.009	0.008	0.000	1.344	0.247
Parent Education	K+G1	0.019	0.019	0.000	0.804	0.37
Parent Age	K+G1	0.008	0.008	0.000	1.04	0.308

Note: F-statistic (adjusted for 519 clusters in School) for a test of the hypothesis that the instrument has no effect.

Table 1.4: 2SLS estimates of the effect of multigrade classrooms on behavior problems in Kindergarten.

Explanatory Variables	**Externalizing**	**Externalizing Subscale**				**Anxiety / Depression**
		Peer Problems	Hyper-activity	Head-strong	Anti-social	
Multigrade	0.445	0.989**	0.908**	-0.098	0.312	0.673*
	(0.392)	(0.434)	(0.464)	(0.363)	(0.379)	(0.406)
Total Kindergarten Enrollment	0.001	0.001	0.001*	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.291***	-0.159***	-0.386***	-0.162***	-0.279***	-0.189***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)
Aboriginal	0.343***	0.319***	0.320***	0.194***	0.302***	0.291***
	(0.068)	(0.076)	(0.067)	(0.062)	(0.070)	(0.072)
Age In Months	-0.002	0.001	-0.003	-0.003	-0.000	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Lone Parent	0.239***	0.182***	0.177***	0.157***	0.225***	0.138***
	(0.053)	(0.057)	(0.054)	(0.049)	(0.058)	(0.053)
Parent Education	-0.014	-0.079*	-0.019	0.009	-0.022	-0.107**
	(0.038)	(0.044)	(0.038)	(0.037)	(0.041)	(0.042)
Respondent is not Mom	0.073	0.077	0.067	0.021	0.156***	0.078*
	(0.045)	(0.050)	(0.044)	(0.045)	(0.047)	(0.047)
English spoken at home	0.123***	-0.041	0.145***	0.078*	0.098**	0.018
	(0.046)	(0.051)	(0.048)	(0.047)	(0.042)	(0.047)
French Immersion	0.024	0.013	0.018	0.024	0.010	0.074
	(0.039)	(0.048)	(0.049)	(0.038)	(0.040)	(0.046)
% FDK by school/grade/year	0.002***	0.001*	0.002***	0.001***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% Disabled by school/grade/year	-0.002	-0.000	0.002	-0.004	0.002	-0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
% Aboriginal by school/grade/year	0.001	-0.001	0.002	0.004*	0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
% English by school/grade/year	-0.001	-0.001	-0.001	-0.000	-0.001	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5341	5340	5340	5340	5341	5341

Notes: All regressions include year effects and heteroskedasticity robust standard errors are clustered at the school level. The sum of excess students in Grades K-2 is used as an instrument for multigrade status. Students from schools with zero excess students in them are excluded.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 1.5: OLS estimates of the effect of multigrade classrooms on behavior problems in Kindergarten.

Explanatory Variables	**Externalizing**	**Externalizing Subscale**				**Anxiety / Depression**
		Peer Problems	Hyper-activity	Head-strong	Anti-social	
Multigrade	0.025	0.033	0.045	0.012	0.015	-0.017
	(0.038)	(0.038)	(0.041)	(0.040)	(0.040)	(0.038)
Total Kindergarten Enrollment	0.000	-0.001***	0.001	-0.000	0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Female	-0.289***	-0.151***	-0.377***	-0.165***	-0.280***	-0.184***
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)
Aboriginal	0.346***	0.327***	0.328***	0.192***	0.304***	0.295***
	(0.067)	(0.075)	(0.065)	(0.061)	(0.069)	(0.070)
Age In Months	-0.002	0.002	-0.003	-0.004	-0.001	-0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Lone Parent	0.234***	0.172***	0.159***	0.160***	0.219***	0.135**
	(0.051)	(0.054)	(0.050)	(0.049)	(0.056)	(0.050)
Parent Education	-0.002	-0.052	0.002	0.008	-0.016	-0.089**
	(0.036)	(0.040)	(0.035)	(0.035)	(0.039)	(0.039)
Respondent is not Mom	0.092**	0.105**	0.092**	0.026	0.175***	0.099**
	(0.046)	(0.047)	(0.043)	(0.045)	(0.049)	(0.045)
English spoken at home	0.109***	-0.082*	0.111***	0.088**	0.091**	-0.008
	(0.042)	(0.043)	(0.040)	(0.044)	(0.041)	(0.041)
French Immersion	0.027	0.028	0.033	0.019	0.010	0.084**
	(0.035)	(0.040)	(0.040)	(0.037)	(0.039)	(0.041)
% FDK by school/grade/year	0.002***	0.001*	0.002***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% Disabled by school/grade/year	-0.003	-0.003	-0.001	-0.003	0.001	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
% Aboriginal by school/grade/year	0.003	0.001	0.004**	0.004**	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
% English by school/grade/year	-0.001	-0.002**	-0.002**	-0.000	-0.001*	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5365	5364	5364	5364	5365	5365

Notes: All regressions include year effects and heteroskedasticity robust standard errors are clustered at the school level. The sum of excess students in Grades K-2 is used as an instrument for multigrade status. Students from schools with zero excess students are excluded from the estimation sample.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 1.6: OLS estimates of the effect of multigrade classrooms on behavior problems in Kindergarten and Grade 1 at the school/grade/year level.

	Externalizing	**Externalizing Subscale**				**Anxiety / Depression**
		Peer Problems	Hyper-activity	Head-strong	Anti-social	
	(0.59)	**(0.60)**	**(0.60)**	**(0.60)**	**(0.57)**	**(0.61)**
Pooled Grades						
	0.092*	0.051	0.096*	0.104**	0.026	0.048
	(0.049)	(0.055)	(0.050)	(0.051)	(0.044)	(0.052)
Kindergarten						
	0.103	0.084	0.074	0.130	0.079	-0.013
	(0.093)	(0.112)	(0.094)	(0.097)	(0.087)	(0.099)
Grade 1						
	0.067	0.033	0.076	0.078	-0.007	0.049
	(0.060)	(0.063)	(0.061)	(0.061)	(0.052)	(0.063)

All regressions include year effects and heteroskedasticity robust standard errors are clustered at the school level. Schools with zero excess students are excluded from the estimation sample. The standard deviation of the school/grade/year level outcome variables are reported in brackets.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 1.7: 2SLS estimation of the effect of multigrade classrooms on behavior problems in Kindergarten and Grade 1 at the school/grade/year level.

	Externalizing	**Externalizing Subscale**				**Anxiety / Depression**
		Peer Problems	Hyper-activity	Head-strong	Anti-social	
	(0.59)	**(0.60)**	**(0.60)**	**(0.60)**	**(0.57)**	**(0.61)**
Pooled Grades						
	0.368	0.296	0.391*	0.242	0.254	0.398
	(0.239)	(0.250)	(0.234)	(0.235)	(0.227)	(0.247)
Kindergarten						
	0.750*	0.824*	0.923**	0.552	0.439	0.763*
	(0.416)	(0.456)	(0.411)	(0.399)	(0.418)	(0.440)
Grade 1						
	0.185	-0.011	0.150	0.090	0.167	0.154
	(0.279)	(0.289)	(0.277)	(0.280)	(0.252)	(0.281)

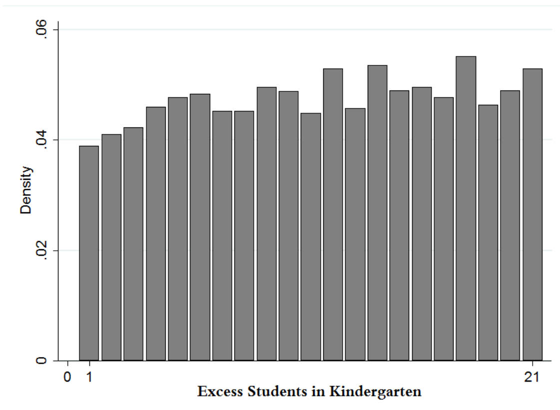
All regressions include year effects and heteroskedasticity robust standard errors are clustered at school level. The sum of excess students in Grades K-2 is used as an instrument for multigrade status. Schools with zero excess students are excluded from the estimation sample. The standard deviation of the school/grade/year level outcome variables are reported in brackets.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 1.8: 2SLS estimates of the effect of multigrade classrooms on behavior problems in Kindergarten, by gender.

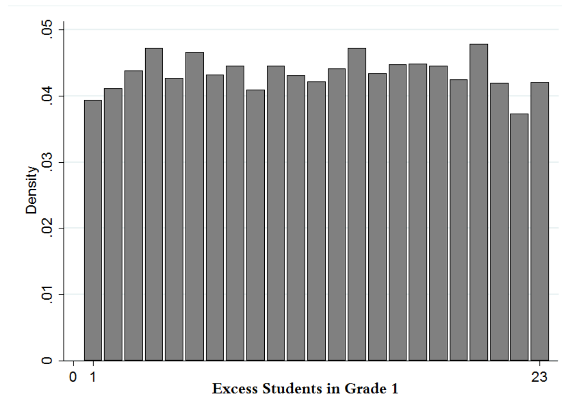
	Externalizing	**Externalizing Subscale**				**Anxiety / Depression**
		Peer Problems	Hyper-activity	Head-strong	Anti-social	
Male	0.898	1.385**	1.226*	0.313	1.103	0.388
	(0.666)	(0.701)	(0.739)	(0.585)	(0.682)	(0.624)
Female	0.027	0.638	0.608	-0.455	-0.474	1.038*
	(0.469)	(0.493)	(0.522)	(0.531)	(0.441)	(0.555)

Notes: All regressions include year effects and heteroskedasticity robust standard errors are clustered at school level. The sum of excess students in Grades K-2 is used as an instrument for multigrade status. Students from schools with zero excess students are excluded from the estimation sample. Control variables include Aboriginal identity, language spoken at home, student's age, parent's education, lone parent, respondent is not mother, French Immersion, percentage of students with Aboriginal identity, percentage with a disability, percentage from English speaking families, percentage in full-day Kindergarten and grade enrollment.
(* = significant at 10%, ** = 5%, *** = 1%)

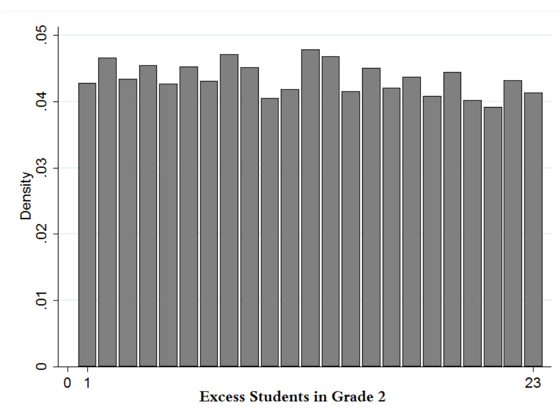
Figure 1.11: Histograms for excess students in Kindergarten (a), Grade 1 (b), Grade 2 (c) and all three grades together (d); excluding school with zero excess students; 2007/08 - 2013/14.



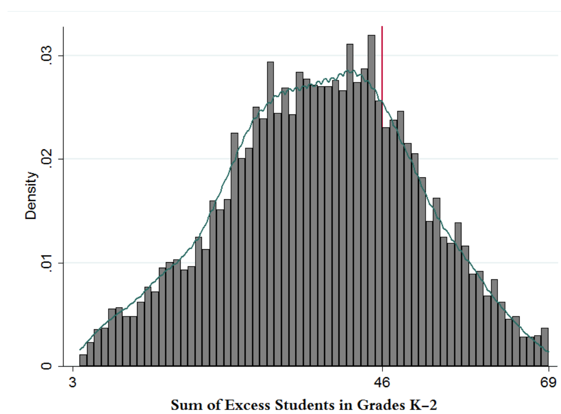
(a)



(b)



(c)



(d)

Source: Author's calculations based on publicly available school reports at www.bced.gov.bc.ca/reporting/school.php.

Figure 1.12: 90% Confidence Interval of the effect of the multigrade classes at individual level - Kindergarten.

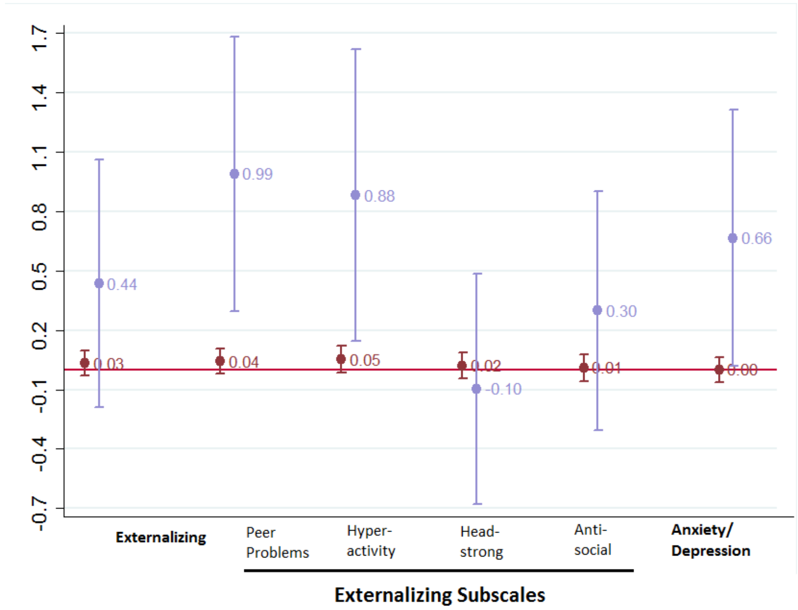


Figure 1.13: 90% Confidence Interval of the 2SLS vs OLS estimates of the effect of the multigrade classes at school/grade/year level - Kindergarten and Grade 1.

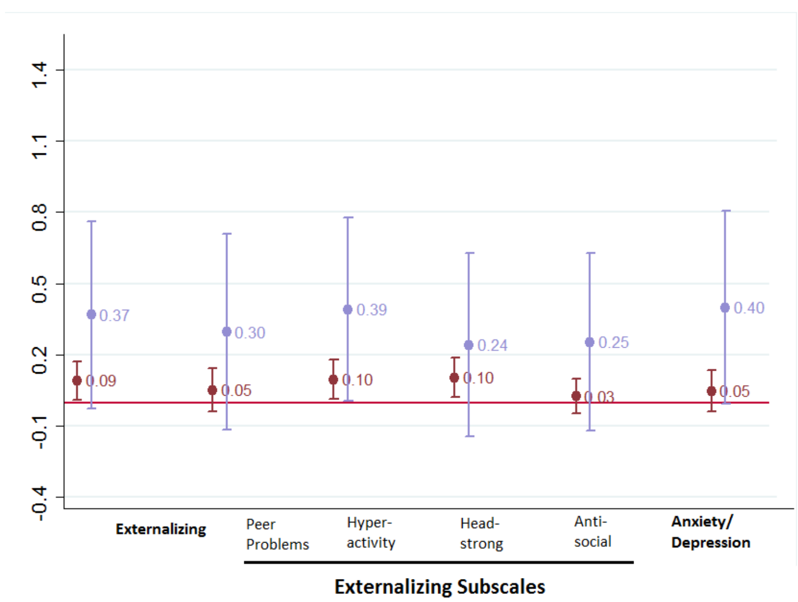


Figure 1.14: 90% Confidence Interval of the 2SLS vs OLS estimates of the effect of the multigrade classes at school/year level - Kindergarten

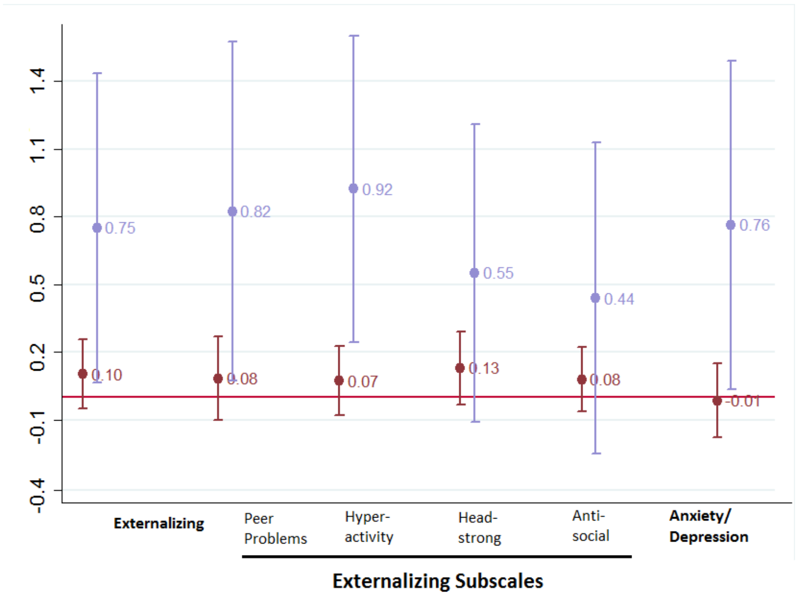
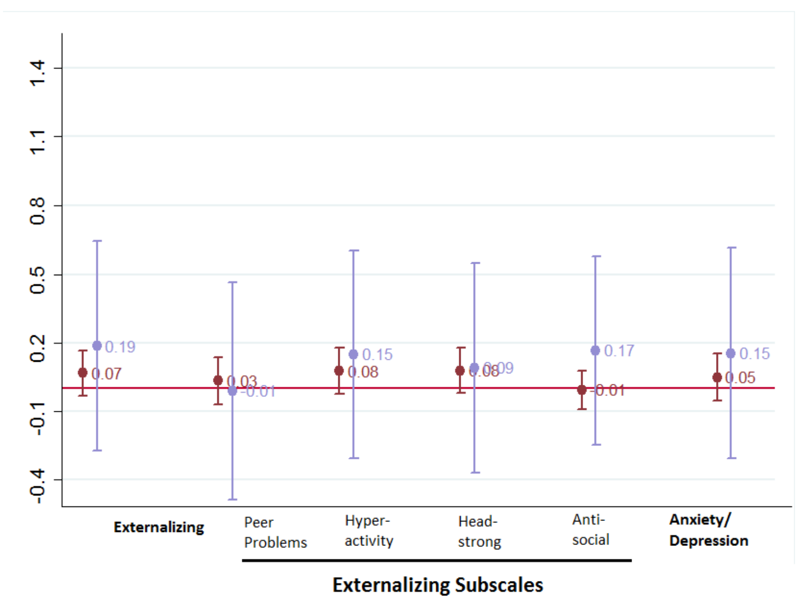


Figure 1.15: 90% Confidence Interval of the effect of the multigrade classes at school/year level - Grade 1



Chapter 2

The effect of full-day Kindergarten on children's behavior

Abstract

We exploit the staggered roll-out of universal full-day Kindergarten (FDK) to estimate its effects on children's behavior. Our research design identifies these effects by comparing across-cohort changes in outcomes in early versus late adopting schools. We find little effect of FDK on child behavior or parents' mental health, and an increase in hours worked by parents who are employed part-time. These results hold across a range of child and family characteristics, with one exception. In families who do not speak English at home, FDK reduces child hyperactivity and peer relationship problems, improves parents' mental health and increases employment and hours.

Keywords: Multigrade classes; non-cognitive skills; behavior problems; hyperactivity; program evaluation; population control policy

2.1 Introduction

Kindergarten in North America is being quietly transformed, as traditional half-day programs are replaced by programs that take up the entire school day. The share of Kindergarteners in the United States attending full-day programs rose from 28% to 76% between 1997 and 2012 (Child Trends, 2013). Six of Canada's ten provinces, including the three largest, have introduced universal full-day Kindergarten (FDK) since 1999.

The expanded Kindergarten day reflects a general trend of increasing public investment in young children. Yet there is no consensus among researchers on either the long-run benefits of such investments or their underlying mechanisms. Many early childhood programs lead to short-run achievement gains that fade out quickly (see Duncan and Magnuson, 2011, for a review), but can deliver other long-run benefits (e.g. Chetty et al., 2011; Deming, 2009; Garces et al., 2002; Ludwig and Miller, 2007). Many analysts argue that the mechanism linking early childhood programs to long-run benefits is the development of "non-cognitive" or "character" skills like persistence, self-control and social aptitude (e.g. Blau and Currie, 2006; Borghans et al., 2008; Cunha et al., 2006; Gibbs et al., 2011; Heckman et al., 2006).

Almost nothing is known about the long-run effects of FDK relative to half-day kindergarten, and relatively little is known about its short-run effects. Gibbs (2012) finds improved achievement in Kindergarten, and earlier studies find test score gains that fade out in subsequent grades (e.g. Cannon et al., 2006, 2011; DeCicca, 2007). Most research on FDK and behavior ignores non-random selection into full- and half-day programs (see the meta-analysis of Cooper et al., 2010, for details). The few studies that seriously address selection have mixed results, and rely on strong identifying assumptions (Cannon et al., 2006; Elicker and Mathur, 1997).

This paper exploits the staggered roll-out of universal FDK in British Columbia (B.C.), Canada to provide new evidence on the short-run effects of FDK on children's behavior. This evidence informs both the debate on the value of these large public expenditures and the broader discussion of investment in early childhood education. Our difference-in-differences research design identifies the effect of FDK by comparing across-cohort outcome changes in schools that adopted universal FDK in different years. This framework accounts for any time-invariant factors that vary systematically across early and late adopting schools.

Our analysis uses a custom survey administered to over 15,000 parents of Kindergarten and Grade 1 children during the roll-out period. The survey includes widely used questions on "externalizing" behavior problems related to hyperactivity, aggression, willfulness and problems in peer relationships, as well as "internalizing" behavior problems related to anxiety and depression. Externalizing behavior problems in children are linked to lower educational attainment, employment and earnings (e.g. Caspi et al., 1998; Farmer, 1993, 1995). Attention deficit/hyperactivity disorder (ADHD), a subcategory of externalizing behavior problems, is linked to subsequent criminal behavior, social assistance receipt and negative school and labor market outcomes (Currie and Stabile, 2006; Currie et al., 2010; Currie and Almond, 2011; Fletcher and Wolfe, 2008, 2009; Fletcher, 2014). Measuring these behavior problems enables us to directly investigate a key channel by which FDK may affect long-run outcomes.

In addition to effects on child behavior, we measure FDK's effect on parental labor supply and mental health. These parental outcomes may also affect the child (Frank and Meara, 2009), and have previously been found to be influenced by access to low-cost childcare (Cannon et al., 2006; Baker et al., 2008, 2015). We also evaluate heterogeneity in effects. The take-up rate of B.C.'s FDK program is nearly 100%, while take-up rates of many nominally universal early childhood programs are much lower (e.g. Gormley and Gayer, 2005; Baker et al., 2008). This feature is valuable in understanding an intervention whose effects may vary due to variation in developmental readiness, counterfactual care environments, and how families reallocate resources in response to FDK.

For the average family in our sample, we find that FDK has little impact on child behavior, and no effect on parents' mental health. It increases hours of work among parents who work part-time when their children are in Kindergarten, but has little if any effect on the likelihood that the average parent is employed or works full-time. In light of these results, the primary benefit of FDK to the average family appears to be a reduction in the private resources required to care for children during the Kindergarten year.

The estimated effects of FDK on child behavior are similar by gender, lone parent status, and parent's education. FDK affects younger children somewhat more than older children. We find very different effects in families that speak a language other than English at home. For these children, FDK causes a marginally statistically significant reduction in externalizing behavior problems in Kindergarten. This overall effect is driven primarily by large reductions in hyperactivity and peer relationship problems. The reduction in hyperactivity persists into first grade. According to Currie and Stabile (2006), improved hyperactivity scores are associated with benefits including increased achievement test scores and schooling attainment, even when scores are well below a level at which a child would be diagnosed with ADHD. Improvements in mental health could serve as a mechanism through which FDK can affect the long-term outcomes of minority language children. While the effect of FDK on peer relationship problems is also of interest, we are not aware of any evidence linking this effect to long-run outcomes. We also find that FDK increases employment of parents in minority language families, and reduces symptoms of parental depression. Together with the estimated improvements in child behavior, these results suggest that FDK may facilitate the social and economic integration of minority language families.

2.1.1 Related literature

While there is a sizeable literature on the effects of FDK on children's behavior and mental health, few papers credibly address potentially non-random selection into FDK (Cooper et al., 2010). Elicker and Mathur (1997) evaluate a small pilot project with random student assignment. They find that FDK improves children's general learning and social skills. Internal validity of this study is somewhat threatened by non-random assignment of teachers and because the relevant outcome measures are derived from teachers' assessments of their own students. The results are based on 179 participating students in a single middle-class community, limiting their external validity. Cannon et al. (2006) use a single year of data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), controlling for a rich set of covariates to account for selection into FDK as well as using state FDK policy as an instrument for student-level FDK attendance. They find that FDK increases the frequency of some behavior problems among Kindergarten students. These results have the advantage of being based on a wide cross-section of U.S. students, but their credibility rests on the strong identifying assumption that either individual FDK attendance or state-level policy is exogenous conditional on the control variables.

Given limited direct evidence on the effects of FDK, findings for related programs such as (half-day) Kindergarten and pre-Kindergarten can be informative. While a full review of those literatures is beyond this paper's scope, evidence on behavioral or long-run effects of large-scale programs (rather than pilot or demonstration projects) is particularly relevant. Results on the behavioral effects of large-scale pre-Kindergarten programs are mixed. Magnuson et al. (2007) use ECLS-K data to investigate the effects of pre-Kindergarten on behavior at the beginning of Kindergarten. They find that attending a pre-Kindergarten program housed in the school where the student subsequently attends Kindergarten has no effect on externalizing behavior, while attending a pre-Kindergarten program elsewhere slightly increases it. Berlinski et al. (2009) measure the effect of expanding pre-Kindergarten programs in Argentina in the 1990s on teacher-assessed behavior in third grade. They find that pre-Kindergarten improved attention, effort, class participation and discipline. Figlio and Roth (2009) use longitudinal administrative data from Florida to study the effect of pre-Kindergarten on behavior during the first few years of primary school. They find that pre-Kindergarten reduces behavior problems, especially among disadvantaged children. Prior studies on the long-run effects of Kindergarten relative to a non-Kindergarten counterfactual have found positive results (Cascio, 2009; Dhuey, 2011). We are aware of no credible evidence of the effects of Kindergarten on the behavior of five-year-olds.

2.2 Data and methods

2.2.1 Full-day Kindergarten in British Columbia

Education in B.C. is funded by the province using a formula based primarily on full-time equivalent (FTE) enrollment. The provincial Ministry of Education administers the funding formula and sets curriculum and other standards, while local school districts are responsible for implementation and resource allocation. Children are eligible and expected to enroll in Kindergarten in September of the calendar year in which they turn five, but schooling is not compulsory until the following September. In practice, nearly all B.C. children attend Kindergarten.

Figure 2.1 illustrates FDK enrollment in B.C. over time. Before the 2010/11 academic year, the Ministry provided half-FTE funding for most Kindergarteners and expected schools to provide half-day programs. Districts could only obtain full-FTE funding when providing FDK to Aboriginal students, English as a Second Language students or those with certain high-cost special needs. Provision of FDK to eligible students was optional for both schools and families. In August 2009, the provincial government announced that FDK would be universal by 2011/12. Implementation was staggered: the province offered funding in 2010/11 to convert up to half of each district's half-day spaces to full-day (and full-FTE). After this transition year, all Kindergarten classes were full-day and funded accordingly.

When deciding which schools would offer FDK in 2010/11, the Ministry asked districts to prioritize schools serving vulnerable populations and to provide FDK to all students in those schools except those in French Immersion programs. Both public statements and interviews with district personnel suggest that most districts used multiple criteria including the vulnerability of the school population and the availability of space and teachers.

Universal FDK implementation was not paired with any curriculum change. The additional class time was intended to provide more opportunities for individualized teacher attention and a more relaxed pace in implementing the existing play-based curriculum. Throughout the period under study, Kindergarten classes were strictly capped at 22 students, and Grade 1 classes at 24 students. The additional operating cost to the Ministry was approximately \$3,370 per affected student and \$100 million annually.

2.2.2 Data

Our main data source is the B.C. School Arrangements Survey (BCSAS), a voluntary survey of Kindergarten and Grade 1 parents conducted by Friesen and Krauth. Twenty-one public school districts agreed to participate in the survey, including the two largest districts of Surrey and Vancouver. All public schools with Kindergarten enrollment in participating districts received surveys to distribute to families. The 2010/11 and 2011/12 surveys covered both Kindergarten and Grade 1 students, while the 2012/13 survey covered Grade 1 students only. The BCSAS instrument includes 60 questions about child and parent outcomes, as well as relevant background variables, and requires 10-15 minutes to complete. Appendix A provides further details on the BCSAS.

Child behavior is characterized using 26 standard descriptions of problem behavior (e.g. "[This child] is disobedient at home") taken from the Behavior Problems Index (BPI) and/or the Canadian National Longitudinal Survey of Children and Youth (NLSCY). Each index is a simple count of behavior problems that the parent identifies as "sometimes true" or "often true" rather than "not true." We use 18 items to construct an overall BPI externalizing behavior scale and four BPI externalizing behavior subscales that measure hyperactivity, antisocial behavior, headstrong behavior and peer relationship problems. We also construct an anxiety/depression score using the eight NLSCY items in our survey. A higher score implies more behavior problems, and all scales are standardized within grade to have zero mean and unit variance to facilitate comparisons. Appendix Table A1 provides a list of items and the composition of scales.

While previous FDK studies measure behavior using teachers' assessments (e.g. Cannon et al., 2006; Elicker and Mathur, 1997), the BCSAS follows the practice of a number of influential surveys in the early childhood education literature and elicits assessments from parents.[1] Parental assessments are arguably more appropriate for our purposes, for two reasons. First, as demonstrated by Elder (2010) for ADHD, teachers' behavior assessments may reflect subjective comparisons to classmates. When all children in a school are treated, as in B.C., the treatment may affect the reference point teachers use to assess whether a student's behavior is problematic. Second, FDK doubles the length of time that a teacher spends with a child, which may influence whether the teacher finds a given level of misbehavior problematic. In both cases, treatment-induced changes in assessment criteria may result in biased estimates of the effects of FDK on the prevalence of behavior problems.

Background variables include gender, Aboriginal identity, language spoken at home, age in months, parent's education, household income, family structure (e.g. lone parent), postal code of residence, number of siblings in the home less than age 5, and whether enrolled in a multiple-grade classroom in Kindergarten. The responding parent's labor supply is measured by usual hours worked per week for pay or profit. Parental depression is

[1] A large literature in psychology demonstrates that teacher and parent assessments of children's behavior are weakly correlated (Achenbach et al., 1987). Most studies of the effects of Kindergarten and pre-Kindergarten on child behavior rely on teachers' assessments of their own students, including those that use data from the ECLS-K (e.g. Cannon et al., 2006; Magnuson et al., 2007) and from Project STAR (e.g. Finn et al., 1989; Chetty et al., 2011). Influential studies of the effects of other childhood interventions on child behavior rely on parents' assessments of their children, including those that use data from the Children of the NLSY79 (e.g. Cunha and Heckman, 2008; Cunha et al., 2010) and the Canadian National Longitudinal Survey of Children and Youth (e.g. Baker et al., 2008). In an appendix to their paper, Baker et al. (2008) provide an extended discussion of parental bias in response to questions about their own children's behavior. Currie and Stabile (2006) find that the relationship between symptoms of ADHD and subsequent educational outcomes is similar whether ADHD symptoms are assessed by parents or by teachers.

measured using a set of six questions from the Center for Epidemiologic Studies Depression Revised (CESD-R) scale. Appendix Table A.2 provides a list of these questions.

The main treatment variable is the parent's report of whether the child was in full-day or half-day Kindergarten. Respondents were also asked to name the child's current school and to indicate if the child was enrolled in a French Immersion program. French Immersion is a separate stream in many B.C. public schools in which the primary language of instruction is French. It is aimed at English-speaking children. In most cases, funds were not allocated to provide FDK to students in French Immersion during the transition year, even when they were made available to provide FDK to students in the same school's regular program.

We supplement the survey data with administrative records from the B.C. Ministry of Education. These records include both confidential student-level enrollment data (collected on September 30 of each year for funding purposes) and publicly available school-level data for all public and private schools in B.C. Unless specified otherwise, we restrict analysis to public school students in districts that participated in the BCSAS. The student-level data includes longitudinal records through the 2012/13 school year for all students who attended Kindergarten between 2009/10 and 2011/12. These records include birth date, gender, current grade, school and district identifiers, self-reported Aboriginal identity, enrollment in a language program (e.g. ESL, French Immersion, Francophone education), enrollment in a special needs program, and language spoken at home. The school/grade-level data includes school name, type, location, and enrollment headcount, as well as measures of average class size and number of students in multiple-grade classes. Students in the BCSAS cannot be matched directly with students in the administrative data, but both student-level data sets can be linked with school/grade-level data.

The main BCSAS sample includes all returned surveys that could be matched with a school. The overall response rate was approximately 16.6%. Table 2.1 presents summary statistics for the main BCSAS sample and the corresponding population in the administrative data. The BCSAS sample captures much of the underlying population's diversity, matching well on gender, student age, and classroom characteristics. FDK students are slightly under-represented among BCSAS respondents (68.4% versus 71.0). Aboriginal students (5.3% versus 8.5%) and minority home language students (18.2% versus 26.4%) also appear to be under-represented. These differences should be interpreted with caution, as responses to questions about Aboriginal identity and home language are often context specific and are known to vary over time in the administrative data for the same student. The share of parents with a high school education or less is also low relative to their share in Census population data. These patterns are consistent with lower response rates from disadvantaged families and those who may face language or cultural barriers to survey participation. As discussed below, this non-random survey response affects the interpretation of our regression results but not necessarily their validity.

2.2.3 Basic empirical framework

Our research design is based on the difference-in-differences framework. Indexing students by i, schools by s and cohorts by c, our basic model is:

$$y_i = a_{s(i)} + \sigma_{c(i)} + \theta FDK_i + X_i\beta + u_i \tag{2.1}$$

where y_i is the outcome, $a_{s(i)}$ and $\sigma_{c(i)}$ are school and cohort fixed effects, FDK_i is an indicator for enrollment in full-day Kindergarten, and X_i is a vector of control variables. The parameter of interest is θ, the effect of attending full-day ($FDK_i = 1$) versus half-day ($FDK_i = 0$) Kindergarten. In the basic model this effect is constant. Each student's treatment FDK_i depends on: (i) the school's choice to offer universal FDK, targeted FDK (for Aboriginal, ESL and special needs students only), or no FDK; (ii) the student's eligibility for targeted FDK, and; (iii) the family's choice to enroll in targeted FDK if available. Both (ii) and (iii) are influenced by unobserved student and family characteristics, so FDK_i cannot credibly be assumed exogenous.

To address this issue, we use whether the school provides universal FDK as an instrument for the student's FDK enrollment. School s is classified as providing universal FDK to cohort c ($UniversalFDK_{sc} = 1$) if a majority of its non-targeted Kindergarten students in the administrative data are in FDK, and as offering targeted FDK or no FDK ($UniversalFDK_{sc} = 0$) otherwise. Figure 2.2 shows the distribution across schools of the percentage of non-targeted students enrolled in FDK in 2010/11. FDK enrollment among non-targeted students is usually zero or 100%, so misclassification is unlikely. The FDK effect θ is identified if:

$$E\Big(u_i | a_{s(i)}, \{X_j\}_{s(j)=s(i)}, \{UniversalFDK_{s(i)c}\}_{\forall c}\Big) = 0 \tag{2.2}$$

i.e., the timing of universal FDK introduction is exogenous after conditioning on the fixed effects. This research design allows for systematic unobserved differences between schools that are early (2010/11) or late (2011/12) adopters of universal FDK (by including school fixed effects), systematic unobserved differences between students who attend FDK and half-day Kindergarten (HDK) in the same school and cohort (by using universal FDK as an instrument for full-day enrollment), unobserved variation over time in province-wide conditions (by including year/cohort effects), and common shocks or clustering of u_i within schools (provided those shocks are unrelated to the availability of universal FDK). Our results are biased to the extent that there are differential unobserved trends for early and late universal FDK adopters, or if school choice responds to FDK availability.

The issue of survey non-response can be modeled by treating the data as a random sample on the random variables $(obs_i, obs_i D_i)$ where D_i is the vector of data on student i and obs_i is an indicator of whether the student is observed in the data. The identifying assumption (2) then becomes:

$$E\Big(u_i|a_{s(i)}, \{X_j\}_{s(j)=s(i)}, \{UniversalFDK_{s(i)c}\}_{\forall c}, obs_i = 1\Big) = 0 \tag{2'}$$

Assumption (2') requires that survey response is unrelated to year-to-year within-school variation in unobserved characteristics, but allows survey response to vary with observed characteristics X_i and with the unobserved school effect. This assumption would be violated, for example, if survey response were influenced by the student's behavior or the parent's beliefs about the benefits of FDK.

2.2.4 Accounting for heterogeneity in effects

This basic model treats the FDK effect as a fixed parameter θ but is easily extended to allow for heterogeneity by observed variables by adding an interaction term:

$$y_i = a_{s(i)} + \sigma_{c(i)} + (\theta_0 + \theta_1 X_{i1})FDK_i + X_i\beta + u_i \tag{1'}$$

where X_{i1} is some subvector of X_i, and $X_{i1}UniversalFDK_{s(i)c(i)}$ is used as an instrument for $X_{i1}FDK_i$.

A particularly important form of observable heterogeneity is between students in schools that adopted universal FDK in 2010/11 (early adopting schools) versus students in schools that adopted universal FDK in 2011/12 (late adopting schools). Schools serving more vulnerable populations were chosen as early adopters, and their more disadvantaged students may respond differently to FDK. One limitation of the BCSAS is that the survey starts in 2010/11, the first year of the FDK roll-out, so we observe Grade 1 outcomes for three entry cohorts (2009/10, 2010/11, and 2011/12), but observe Kindergarten outcomes for only two (2010/11 and 2011/12). To see the consequences of this limitation, let $X_{i1} = 1$ for early adopting schools and $X_{i1} = 0$ for late adopting schools. For the BCSAS Kindergarten data, there is no within-school variation in the instrument $X_{i1}UniversalFDK_{s(i)c(i)}$ – it equals zero in both years for late adopting schools and one in both years for early adopting schools. As a result, the effect for late adopting schools (θ_0) can be estimated but the effect for early adopting schools ($\theta_0 + \theta_1$) and the difference (θ_1) cannot. Grade 1 BCSAS outcomes and all outcomes in the administrative data are observed for all three cohorts, so we can identify effects on those outcomes for both early adopting and late adopting schools. A similar analysis applies to survey non-response, implying that our estimates should be interpreted as measuring the relevant treatment effect within the subpopulation of respondents rather than the full population of students.

Unobserved heterogeneity can be understood using the LATE framework of Angrist and Imbens (1994). We can interpret θ as the local average treatment effect of FDK among compliers (students who attend FDK if and only if it is universal at their school) exposed to variation in the instrument (students in schools that vary universal FDK during the period covered by the data). Treatment effects are never identified for the 0.8% of students who are never-takers (those in half-day Kindergarten in schools classified as universal FDK) or the 21% of students (60% of targeted students) who are always-takers (those in targeted groups who would have enrolled in targeted FDK).

2.3 Results

2.3.1 Child behavior

Our main results for Kindergarten behavior are reported in Table 2.2. All regressions include school and cohort/year fixed effects, and standard errors are clustered at the school level[2]. The universal FDK instrument is very strong, with a robust F-statistic over 1,000 in all regressions. Results for Grade 1 behavior are similar.

Relationships between externalizing behavior problems and background characteristics are similar in sign and magnitude to comparable results in the ECLS-K data (see Duncan and Magnuson, 2011). Boys exhibit more externalizing behavior problems than girls (by 28% of a standard deviation), children in minority language families exhibit fewer than those speaking English at home (by 10% of a standard deviation), and children of lone parents exhibit higher rates than children in multi-parent families (by 25% of a standard deviation). The difference between Aboriginal and non-Aboriginal children is 30% of a standard deviation, similar to the black-white difference in the ECLS-K. Conditional on these characteristics, parent's education is not an important predictor of externalizing behavior problems.

In comparison, the effect of FDK on externalizing behavior is small (effect size = .052 or 5.2% of a standard deviation) and statistically insignificant. The positive sign of the point estimate implies that FDK is associated with increased behavior problems, and the 95% confidence interval [-.066, .171] allows us to reject the null that FDK causes any substantial behavioral improvement for the average child. The next four columns of Table 2.2 report results for four BPI subscales measuring different dimensions of externalizing behavior. The antisocial subscale includes cheating, lying, bullying and not getting along with teachers; the hyperactive subscale includes attention, impulsivity and restlessness; the headstrong subscale includes nervousness, disobedience at home, stubbornness, and tendency to be argumentative; and the peer relationship problems subscale refers to difficulties engaging and getting along with other children. As with the overall scale, the estimated effect size of FDK on each of these subscales is small and statistically insignificant.

[2] In some cases, a French Immersion program is housed within the same school as a regular English program, while in other cases it stands alone as a distinct school entity. When housed within the same school, we define a "school" as a school-program pair for the purpose of modeling fixed effects and for defining the availability of FDK to regular students. We treat the two programs as a single school entity when clustering standard errors.

The last column presents results for the NLSCY anxiety/depression scale. The items in this scale refer to aspects of so-called internalizing behavior, including the extent to which the child is fearful, anxious and sad. Differences by gender, Aboriginal identity and lone parenthood are much less pronounced than for externalizing behavor. FDK is associated with an increase in emotional problems of about 8.2% of a standard deviation. As with externalizing behavior, this effect is statistically insignificant and the 95% confidence interval of [-.044, .207] allows us to reject any substantial improvement for the average child.

The top panel of Table 2.3 re-displays the primary results from Table 2.2. As discussed in Section 2.2.4, these estimates measure the FDK effect only for students in late-adopting schools, and so may be missing benefits that are realized only in the more disadvantaged early-adopting schools. While it is not possible to directly investigate this heterogeneity in Kindergarten, it is possible in Grade 1. The second panel in Table 2.3 reports separate estimates for Grade 1 students in early and late adopting schools. There is little evidence that FDK has a different effect on Grade 1 behavior in early-adopting and late-adopting schools, or a substantial effect in either. While we cannot rule out a large effect on Kindergarten behavior in early adopting schools, any such effect does not persist into Grade 1.

The results in the lower panel of Table 2.3 provide little evidence that the effect of FDK varies by gender, lone parent status or parent's education. They provide weak evidence that the effect on behavior problems differs between children born in the first and second half of the calendar year. The point estimates suggest that FDK increases both externalizing behavior problems and anxiety/depression among younger children, and these effects are statistically significant at the 10% level. The final set of estimates shows a substantial and statistically significant difference by language spoken at home. The point estimates imply that FDK has little effect on children from English speaking families, but reduces externalizing behavior problems of children from minority language families by 20% of a standard deviation. This behavior improvement is statistically significant at the 10% level.

Table 2.4 reports additional detail on heterogeneity by home language. The top panel shows that the reduction in externalizing behavior problems among minority language Kindergarten children (effect size = -0.20*) is primarily through reduced hyperactivity (-0.21*) and peer relationship problems (-0.32***). The middle panel shows that the improvements in hyperactivity persist into Grade 1 (-0.25*), but other effects do not. There is also weak evidence that FDK reduces hyperactivity among English-speaking students in Grade 1 (-0.13*). To place these results in context, the lower panel of Table 2.4 provides average values of the behavior scales for a baseline population of untreated students in late adopting schools. Minority language students in this population have more behavior problems than English language students, although these differences are small and statistically insignificant. This difference is especially pronounced and highly statistically significant in the case of peer relationship problems. The estimated effect of FDK on peer relationship problems nearly closes this baseline gap between minority and English language students.

The baseline gap in hyperactivity is small, so the estimated effect of FDK reverses the direction of the gap in favor of minority language students.

2.3.2 Parental outcomes

Table 2.5 reports estimates of the effects of FDK on parents' labor supply and mental health[3]. Like the introduction of public half-day Kindergarten (Gelbach, 2002), the extension of the Kindergarten day can be thought of as offering a 100-percent marginal price subsidy for childcare among families who are not already purchasing childcare for the equivalent number of hours. For families already purchasing this childcare, the subsidy is infra-marginal and therefore provides both price and income subsidies. FDK therefore may increase labor force participation, increase or decrease hours among part-time workers, and reduce the hours of some full-time workers. This change in work hours, income, and time with children may also affect the parents' mental health.

The first row in Table 2.5 presents results for Kindergarten parents and the second row shows results in Grade 1. FDK increases overall hours in Kindergarten by more than 2 hours per week, and this effect is statistically significant. The increase occurs primarily on the intensive margin; the estimated effect on employment is small and statistically insignificant. Most of the increase in labor supply occurs between 15 and 35 hours per week; the proportion of parents who report working at least 15 hours per week increases by 8.6 percentage points, but the proportion of parents working at least 35 hours per week increases by just 1.2 percentage points. The effect of FDK on labor supply fades out by Grade 1. The effect of FDK on parental depression varies in sign across grades, but is always small (effect size < 0.03) and statistically insignificant.

The next panel of Table 2.5 shows that the effect of FDK on English and minority language parents differs substantially. Among English-speaking parents, FDK increases the proportion of those who work more than 15 hours per week by 7.1 percentage points, but has no effect on the proportion who are employed or who work full-time. In contrast, FDK increases the proportion of minority language parents who are employed by 19.6 percentage points, the proportion who work full-time by 11.6 percentage points, and the proportion who work more than 15 hours per week by 22.1 percentage points. This effect does not persist to Grade 1. FDK has almost no effect on depression among English-speaking parents, but substantially reduces depression among minority language parents (effect size = 0.32^{***}). To place these results in context, the lower panel of Table 2.5 reports average parental outcomes for the baseline population of untreated students in late adopting schools. The employment rate among minority language parents in this population is 10 percentage points lower than for English language parents. This difference is primarily due to lower rates of part-time

[3] The results reported in Table 2.5 refer to the labor supply of the parent who responded to the survey. Most but not all of these respondents were mothers. We report results for a sub-sample of mothers only in Appendix Table C.3. The results are very similar.

employment among minority language parents; full-time employment rates are similar across the two groups. The estimated effect of FDK reverses this gap in employment. Finally, minority language parents report moderately more symptoms of depression in the baseline sample (0.181**). According to the estimates in Table 2.5, FDK completely eliminates this differential.

2.4 Robustness checks

Our causal interpretation of these estimates relies on the absence of differential unobserved trends between early-adopting and late-adopting schools. This assumption could be violated if: (i) adoption of universal FDK coincided with other changes to the school environment; (ii) adoption of universal FDK influenced families' school choice decisions, or; (iii) adoption of universal FDK was influenced by unobserved trends in neighborhood composition. As discussed in Section 2.2.3, we also require the absence of differential unobserved trends in survey response patterns between early and late adopting schools.

We assess each of these assumptions in more detail below, and explore the potential consequences of specific threats to the validity of our results by calculating the range of true effects that could be masked by different types and degrees of selection. These robustness checks make extensive use of the administrative data.

2.4.1 School environments

If universal FDK is introduced simultaneously with changes to the school environment, our estimates may confound the true effect of universal FDK with the effect of these changes. While there was no change to provincial curriculum or class size rules, individual schools may respond to the need for additional teachers by increasing class sizes or the frequency of multiple-grade classes. The results in the upper panel of Table 2.6 provide no evidence that FDK affected the frequency of multiple-grade classes in Kindergarten. However, they indicate that the average Kindergarten class size increased by approximately 1 to 2 students, Grade 1 class size decreased by approximately 0.5 students, and the frequency of multiple-grade classes in Grade 1 went down by 6.3 percentage points. While statistically significant, these effect sizes are small; a small literature that estimates the effect of class size on child behavior finds very small effects from much larger class size reductions (Chetty et al., 2011; Dee and West, 2011; Finn et al., 1989). We conclude that it is very unlikely that our results are driven by changes in class size or grade composition.

2.4.2 School choice

The availability of universal FDK in some schools but not others may have influenced school choice decisions in 2010/11. We investigate this possibility using an extended version of our administrative data set that includes both public school students and the approximately 12% of students who attend private schools. We estimate the effect of the availability of universal FDK in the student's own catchment school on their decision to enroll in that school, in a public French Immersion program, in another public school or in a private school. These regressions include catchment fixed effects since early-adopting schools may be more or less attractive than late-adopting schools for other reasons. The lower panel of Table 2.6 shows that the share of Kindergarten students who enrolled in their catchment school grew on average by about 4.2 percentage points when the school introduced universal FDK. This increase corresponds to about 1.9 students at an average-sized school that enrolls 44 students per grade. The majority of this in-catchment enrollment growth came at the expense of enrollment in other public schools; the availability of universal FDK at the catchment school had virtually no effect on the share of catchment area residents who enrolled in French Immersion programs or in private schools.

While statistically significant, this enrollment shift poses only a small threat to our identification strategy. For example, consider the effect of FDK on externalizing behavior. Assuming that the availability of FDK in the catchment school affects enrollment only via the decisions of in-catchment children who would otherwise go out of catchment, we can bound the true value of the treatment effect that is consistent with our estimates. The smallest possible number of behavior problems is zero, which corresponds to a standardized score of -1.10. Since the additional in-catchment students who enroll in FDK schools form a small minority of the total number of students enrolled, the average scores of the remaining students must be close to the overall average. If all of the additional enrollees are as positively selected as possible, our point estimate that FDK increases externalizing behavior by 0.052 standard deviations (see Table 2.2) is consistent with a true effect of 0.10 standard deviations (see B for details of calculations). At the other extreme, if the additional enrollees were negatively selected such that their average score was at the 95th percentile of the range observed in the sample (1.87), the point estimate is consistent with a true effect of -0.03 standard deviations. We conclude that selection bias from this source cannot be masking a non-trivial improvement in externalizing behavior. We cannot rule out the possibility that FDK causes a small increase in externalizing behavior problems that is masked by selection bias; however, we conclude that this scenario is fairly unlikely.

2.4.3 Student characteristics

Districts may choose early-adopting schools on the basis of criteria that are related to unobserved trends in school composition. For example, if communities are rapidly polarizing by socioeconomic status, early-adopting schools will be serving increasingly vulnerable populations over time. While it is not possible to directly investigate differential unobserved trends in the student population, differential trends in observed characteristics can be found using reduced form placebo regressions of the form:

$$X_i = a_{s(i)} + \lambda_{c(i)} + \gamma UniversalFDK_{s(i)c(i)} + v_i \tag{2.3}$$

where X_i is a pre-treatment background characteristic such as gender. Since we know that the true effect of FDK on pre-treatment characteristics is zero, our research design has the testable implication that $\gamma = 0$.

The first two columns of Table 2.7 present placebo effect estimates from the administrative data, for all three cohorts and for the two cohorts that correspond to the BCSAS Kindergarten sample. The results show that schools enrolled more minority language students when adopting universal FDK, and may also have enrolled slightly older students. These results suggest that the potential bias from induced enrollment in FDK schools may be greater among minority language students than in the general student population. Again, we can bound the true effects under various scenarios. The first step is to calculate the number of additional minority language students who enroll in FDK schools. From previous calculations, total enrollment increases by 1.9 students when a school adopts universal FDK. The share of minority language students in the school increases by 4.3 percentage points (Table 2.7), on a baseline of 26.4% (Table 2.1). Assuming that universal FDK affects enrollment only via the decisions of in-catchment children who would otherwise go out of catchment, these results imply that schools gain 2.5 minority language students when adopting universal FDK (see calculations in B). These students account for 18% of minority language enrollment in universal FDK schools.

Tables 2.4 and 2.5 report statistically significant and negative estimates of FDK on hyperactivity, peer relationship problems and parental depression among minority language students. In the case of these measures, a zero raw score corresponds to a standardized score of -1.05, -0.46 and -1.01 standard deviations respectively. If all of the additional enrollees are as positively selected as possible, the estimated effect of FDK on hyperactivity (-0.209; see Table 2.4) is consistent with a true treatment effect of -0.02 standard deviations, the estimate for peer relationship problems (-0.317) is consistent with a true treatment effect of -0.23 standard deviations, and the point estimate for parental depression (-0.321; see Table 2.5) is consistent with a true treatment effect of -0.14 standard deviations (see B for details of calculations). We conclude that this form of selection bias is not masking a true increase in hyperactivity, peer relationship problems, and parental depression caused by FDK, or a null effect. We cannot rule out the possibility that the true reduction in hyperactivity for this group is small. On the other hand, the true benefits of FDK may be substantially larger than our point estimates if the additional enrollees are negatively selected. We conclude that FDK caused a significant and substantial reduction in peer relationship problems among minority language children and depression among their parents, and that these results are robust to potential selection bias from this source.

We can also assess the threat from additional minority language student enrollment for our estimates of parental employment. Our point estimates indicate that FDK raises parental employment by 19.6 percentage points (0.196, Table 2.5), weekly employment of 15 hours or more by 22.1 percentage points (0.221), and weekly employment of 35 hours or above by 11.6 percentage points (0.116). If the parents of all of the additional students were employed full-time (thus biasing our point estimates upward), these point estimates would be consistent with a true effect of 15 percentage points, 16 percentage points, and 1 percentage point respectively. In other words, our results for employment and working at least 15 hours per week are robust to selection bias, but our results for working at least 35 hours per week may not be.

2.4.4 Survey response rates

The third column of Table 2.7 presents results from placebo regressions estimated from the BCSAS survey data. None of the placebo estimates is statistically significant and the point estimates are nearly zero. While this result would normally be interpreted as supportive of our identifying assumptions, we know from the administrative data that the proportion of minority language students increased with universal FDK. Since the administrative data measure the characteristics of the population, this difference implies that the survey response rate among minority language students declined under universal FDK relative to the response rate among English language students.

Specifically, the net gain of 2.5 minority language students in schools with universal FDK did not produce any increase in the number of survey respondents. Section 2.4.3 discusses the potential threat from non-random selection of 2.5 additional students. In this case, we consider the potential threat from non-random selection of 2.5 missing survey respondents. These two threats are identical in magnitude, but opposite in sign. If non-respondents are highly negatively selected such that their average score was at the 95th percentile of the range observed in the sample, the point estimates are consistent with a true effect of 0.08 standard deviations for hyperactivity, -0.13 standard deviations for peer relationship problems, and -0.02 standard deviations for parental depression. Alternatively we can calculate what average score among non-respondents would be needed in order for the point estimates to be consistent with a true effect of zero. This score would be 1.19 for hyperactivity (between the 85.3 and 95.8 percentile of the hyperactivity distribution), 1.81 for peer relationship problems (between the 92.8 and 98.0 percentile), and 1.83 for parental depression (between the 93.5 percentile and the maximum). We conclude from these results that our results for peer relationship problems and parental depression are quite robust to this form of non-random selection, and that our results for hyperactivity are moderately robust.

2.5 Conclusion

Some early childhood interventions have been found to improve long term life outcomes even in the absence of persistent effects on academic achievement. In the face of this evidence, improvements in "non-cognitive" skills are commonly hypothesized as the main mechanism driving these long-run benefits. This hypothesis is supported by a literature linking childhood behavior and mental health to long term outcomes. Externalizing behavior, particularly behavior associated with hyperactivity, emerges as an important factor in this literature.

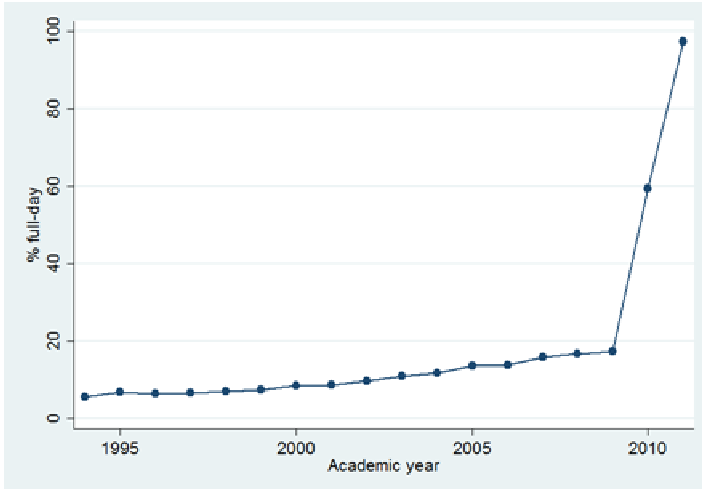
This paper advances this research agenda by providing new and credible estimates of FDK's effect on child behavior and on related parental outcomes. When interpreting these results, it is important to note that they pertain to a universal program. Our study therefore makes only a limited contribution to the literature on the effects of early childhood education programs that are targeted towards extremely disadvantaged children. It makes a larger contribution to the literature on universal programs, which are among the costliest new investments in early childhood education.

We find that FDK has little if any effect on the externalizing behavior and emotional health of the average child. This result is similar across children with a broad range of characteristics that are associated with more or less disadvantage, with the key exception of children from families who do not speak English at home. We find that FDK increases average hours worked by parents, and this increase is due to increased hours among part-time workers rather than increased employment or full-time work. At the same time, we find no evidence that these changes in parents' schedules affect their mental health. We conclude that the primary benefit of FDK for the average family is the childcare subsidy implied by the longer school day.

In contrast, FDK appears to substantially benefit minority language families: children's behavior improves and they enjoy better peer relationships, while their parents work more and show fewer signs of depression. Our research design cannot distinguish the extent to which the estimated improvements in child behavior are the direct result of attending a FDK program instead of an alternative form of care, or the indirect result of increased family income and/or improvements in parents' mental health. While this distinction is not critical for policy evaluation, in the context of the broader literature better understanding of these potential underlying mechanisms is a goal for future research.

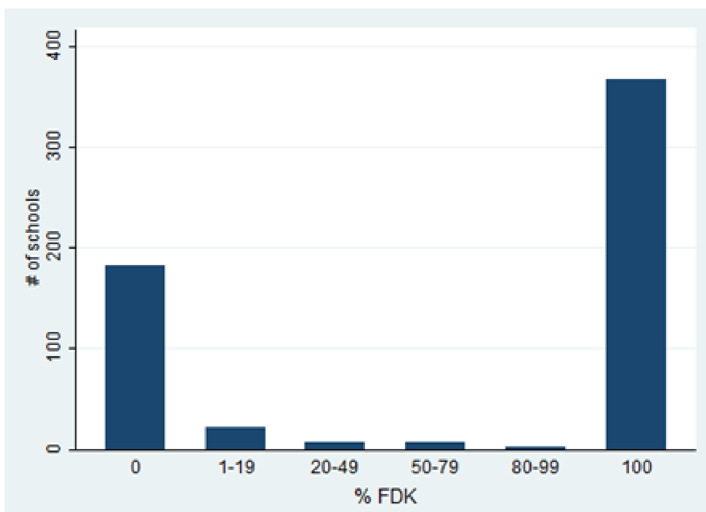
Tables and figures

Figure 2.1: Full-day Kindergarten enrollment rates in British Columbia, 1994/95-2011/12.



Source: Authors' calculations from public enrollment data; see "BC Schools - School Headcount By Grade Historical" downloaded from http://www.data.gov.bc.ca, Feb 27, 2013.

Figure 2.2: Frequency distribution: percentage of non-targeted Kindergarten students enrolled in FDK, by school, 2010/11.



Source: Authors' calculations from confidential Ministry of Education enrollment data.

Table 2.1: Summary statistics, students in Kindergarten and Grade 1.

Variable		**BCSAS data (std dev)**	**Administrative data (std dev)**
Number of observations		15,808	95,123
% of target population		16.6%	100.%
Student characteristics:			
Male		51.7%	51.6%
Aboriginal identity		5.3%	8.5%
Language spoken at home is not English		18.2%	26.4%
Average age in months		69 .4	69.6
		(7.6)	(6.9)
School/classroom characteristics:			
Enrolled in English program at catchment school		58.3%	63.2%
Enrolled in French Immersion program		15.5%	11.0%
Enrolled in English at other public school		26.2%	26.8%
Average class size		20.1	20
		(2.3)	(2.2)
School/classroom characteristics (in Kindergarten):			
Enrolled in full-day Kindergarten		68.4%	71.0%
School had universal FDK		63.4%	66.3%
In a multiple-grade class		12.6%	10.4%
Responding parent characteristics:			
Mother		87.5%	
Father		10.1%	
Lone parent		8.6%	
Average age in years		38.6	
		(6.0)	
Highest level of education:	HS dropout	2.9%	
	HS graduate	12.5%	
	Some post-secondary	41.9%	
	Bachelor's or higher	42.7%	
Labor supply:	Average weekly hours	22.8	
		(16.9)	
	Hours > 0	76.1%	
	Hours>=15	65.6%	
	Hours>=35	37.6%	

The BCSAS data cover Kindergarten students in 2010 and 2011, and Grade 1 students 2010, 2011, and 2012, for the 21 public school districts that participated in the BCSAS in all three years. Summary statistics for the administrative data are reported for this same population to maximize comparability.

Table 2.2: Effect of full-day Kindergarten on behaviour problems in Kindergarten (BCSAS data).

Explanatory variable	**Externalizing**	**Externalizing subscales**				**Anxiety/ Depression**
		Anti-social	Hyper-active	Head-strong	Peer problems	
Full-day Kindergarten	0.052	0.045	0.012	0.052	-0.107	0.082
	(0.060)	(0.058)	(0.061)	(0.062)	(0.064)	(0.064)
Male	0.282***	0.270***	0.377***	0.161***	0.148***	0.074***
	(0.026)	(0.025)	(0.026)	(0.027)	(0.026)	(0.028)
Aboriginal identity	0.300***	0.231***	0.272***	0.173***	0.306***	0.233***
	(0.066)	(0.067)	(0.062)	(0.059)	(0.072)	(0.072)
Language spoken at	-0.101***	-0.070	-0.074*	-0.092**	0.049	-0.117***
home is not English	(0.043)	(0.043)	(0.039)	(0.044)	(0.042)	(0.039)
Student's age in months	-0.003	-0.001	-0.003	-0.003	-0.001	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Parent's education (base category HS):						
HS dropout	0.063	0.088	0.100	0.016	0.057	0.084
	(0.087)	(0.094)	(0.081)	(0.084)	(0.097)	(0.083)
Some Postsecondary	0.074**	0.035	0.088**	0.081**	-0.052	0.064
	(0.039)	(0.041)	(0.038)	(0.040)	(0.042)	(0.039)
Bachelor or higher	-0.040	-0.060	-0.050	-0.041	-0.043	0.045
	(0.041)	(0.045)	(0.042)	(0.040)	(0.043)	(0.039)
Lone parent	0.249***	0.229***	0.165***	0.173***	0.174***	0.164***
	(0.051)	(0.055)	(0.049)	(0.048)	(0.054)	(0.053)
Respondent is not mother	0.091**	0.148***	0.083**	0.033	0.073	0.033
	(0.044)	(0.047)	(0.042)	(0.044)	(0.046)	(0.043)
Parent's age in years	-0.005*	0.001	-0.004	-0.007***	0.001	-0.006**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Robust 1st stage F-statistic	1664	1664	1662	1662	1662	1664
Number of observations	6403	6403	6402	6402	6402	6403
Number of schools	485	485	485	485	485	485

All regressions include year and school fixed effects (treating English and French programs in the same location as separate schools), and standard errors are clustered by school location (allowing for correlation between English and French programs in the same location). Universal FDK availability is used as an instrument for FDK enrolment.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 2.3: Effect of full-day Kindergarten on behaviour problems in Kindergarten and Grade 1, heterogeneous effects (BCSAS data).

Effect of interest		**Externalizing**	**Anxiety/ Depression**
Effect of FDK on Kindergarten behaviour			
	FDK	0.052	0.082
		(0.060)	(0.064)
Effect of FDK on Grade 1 behaviour:			
By early vs. late adopting schools	FDK x (early adopter)	-0.058	0.058
		(0.092)	(0.095)
	FDK x (late adopter)	-0.020	0.043
		(0.074)	(0.076)
	Difference (late – early)	0.038	-0.015
		(0.075)	(0.078)
Effect of FDK on Kindergarten behaviour:			
By gender:	FDK x (male)	0.063	0.070
		(0.070)	(0.074)
	FDK x (female)	0.041	0.094
		(0.070)	(0.071)
	Difference (female - male)	-0.023	0.024
		(0.072)	(0.068)
By child's age:	FDK x (older)	-0.002	0.052
		(0.064)	(0.068)
	FDK x (younger)	0.104*	0.110*
		(0.063)	(0.066)
	Difference (younger - older)	0.106***	0.057
		(0.036)	(0.040)
By lone parent:	FDK x (not lone parent)	0.045	0.084
		(0.062)	(0.065)
	FDK x (lone parent)	0.133	0.060
		(0.166)	(0.161)
	Difference (lone - non-lone)	0.087	-0.024
		(0.167)	(0.158)
By post-sec. education:	FDK x (some post-secondary)	0.042	0.080
		(0.060)	(0.063)
	FDK x (no post-secondary)	0.150	0.094
		(0.060)	(0.127)
	Difference (no post – some-post)	0.108	0.013
		(0.113)	(0.111)
By home language:	FDK x (English)	0.082	0.109*
		(0.061)	(0.065)
	FDK x (not English)	-0.202*	-0.150
		(0.115)	(0.104)
	Difference (not English – English)	-0.284**	-0.259***
		(0.109)	(0.098)

All regressions include year and school fixed effects (treating English and French programs in the same location as separate schools), and standard errors are clustered by school location (allowing for correlation between English and French programs in the same location). Universal FDK availability is used as an instrument for FDK enrolment. Control variables include gender, Aboriginal identity, language spoken at home, student's age, parent's age, parent's education, lone parent, respondent is not mother.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 2.4: Effect of full-day Kindergarten on behaviour problems in Kindergarten and Grade 1, by home language (BCSAS data).

Effect of interest		**Externalizing**	**Externalizing subscales**				**Anxiety/ Depression**
			Anti-social	Hyper-active	Head-strong	Peer problems	
Effect of FDK on Kindergarten behaviour:							
	FDK x (English)	0.082	0.066	0.038	0.081	-0.082	0.109*
		(0.061)	(0.060)	(0.062)	(0.063)	(0.063)	(0.065)
	FDK x (not English)	-0.202*	-0.134	-0.209*	-0.189	-0.317***	-0.150
		(0.115)	(0.115)	(0.117)	(0.126)	(0.122)	(0.104)
	Difference	-0.284***	-0.199*	-0.247**	-0.270**	-0.235**	-0.259***
	(not English–English)	(0.109)	(0.113)	(0.108)	(0.124)	(0.109)	(0.098)
Effect of FDK on Grade 1 behaviour:							
	FDK x (English)	-0.039	-0.034	-0.127*	0.037	-0.033	0.041
		(0.072)	(0.071)	(0.075)	(0.074)	(0.074)	(0.075)
	FDK x (not English)	-0.036	-0.106	-0.250*	0.122	-0.007	0.127
		(0.146)	(0.147)	(0.148)	(0.151)	(0.149)	(0.139)
	Difference	0.003	-0.072	-0.123	0.085	0.025	0.086
	(not English–English)	(0.121)	(0.123)	(0.117)	(0.121)	(0.120)	(0.115)
Mean values at baseline, Kindergarten children:							
	English	-0.173	-0.149	-0.139	-0.130	-0.121	-0.126
	not English	-0.057	0.013	-0.054	-0.058	0.133	-0.053
	Difference	0.116	0.161*	0.085	0.072	0.254***	0.073
	(not English–English)	(0.079)	(0.086)	(0.081)	(0.092)	(0.087)	(0.076)

All regressions include year and school fixed effects (treating English and French programs in the same location as separate schools), and standard errors are clustered by school location (allowing for correlation between English and French programs in the same location). Universal FDK availability is used as an instrument for FDK enrolment. Control variables include gender, Aboriginal identity, language spoken at home, student's age, parent's age, parent's education, lone parent, respondent is not mother.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 2.5: Effect of full-day Kindergarten on parental outcomes in Kindergarten and Grade 1, overall and by home language (BCSAS data).

Effect of interest		**Labor supply**				**Depression**
		Weekly hours	Hours > 0	Hours >= 15	Hours >=35	
Effect of FDK on Kindergarten parents:						
	FDK	2.100**	0.025	0.086***	0.012	-0.010
		(1.040)	(0.029)	(0.030)	(0.030)	(0.060)
Effect of FDK on Grade 1 parents:						
	FDK	-0.816	-0.019	-0.017	-0.048	0.028
		(1.234)	(0.033)	(0.035)	(0.036)	(0.082)
Effect of FDK on Kindergarten parents, by home language:						
	FDK x (English)	1.525	0.006	0.071**	0.001	0.026
		(1.059)	(0.029)	(0.030)	(0.031)	(0.061)
	FDK x (not English)	7.352***	0.196***	0.221***	0.116*	-0.321***
		(2.430)	(0.067)	(0.071)	(0.062)	(0.127)
	Difference	5.827**	0.190***	0.150**	0.115*	-0.347***
	(not English–English)	(2.370)	(0.064)	(0.068)	(0.060)	(0.124)
Effect of FDK on Grade 1 parents, by home language:						
	FDK x (English)	-0.810	-0.030	-0.019	-0.040	0.026
		(1.177)	(0.031)	(0.033)	(0.035)	(0.078)
	FDK x (not English)	-0.870	0.068	0.006	-0.112	0.042
		(2.533)	(0.064)	(0.070)	(0.071)	(0.158)
	Difference	-0.061	0.098**	0.025	-0.073	0.016
	(not English–English)	(2.046)	(0.050)	(0.056)	(0.059)	(0.125)
Mean values at baseline, Kindergarten parents:						
	English	21.284	0.738	0.616	0.334	-0.077
	not English	20.093	0.637	0.556	0.347	0.104
	Difference	-1.191	-0.101**	-0.060	0.013	0.181**
	(not English–English)	(1.951)	(0.051)	(0.053)	(0.051)	(0.088)

All regressions include year and school fixed effects (treating English and French programs in the same location as separate schools), and standard errors are clustered by school location (allowing for correlation between English and French programs in the same location). Universal FDK availability is used as an instrument for FDK enrolment. Control variables include gender, Aboriginal identity, language spoken at home, student's age, parent's age, parent's education, lone parent, respondent is not mother.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 2.6: Effect of full-day Kindergarten on school organization and enrolment, Kindergarten and Grade 1 (administrative data)

Effect of interest		**Kindergarten**		**Grade 1**
		Three cohorts	Two cohorts	Three cohorts
Effect of FDK on classroom environment:				
	Average class size	1.852***	2.047***	-0.577***
		(0.266)	(0.258)	(0.190)
	Multiple-grade class	0.011	-0.007	-0.063**
		(0.017)	(0.018)	(0.026)
Effect of FDK in own catchment school on school choice:				
	Own catchment school	0.042***	0.043***	0.046***
		(0.010)	(0.011)	(0.010)
	French immersion program	-0.002	-0.002	-0.004
		(0.005)	(0.007)	(0.006)
	Other public school	-0.036***	-0.032***	-0.034***
		(0.007)	(0.009)	(0.007)
	Private school	-0.003	-0.009	-0.008
		(0.006)	(0.006)	(0.006)

All regressions include year and school fixed effects, and standard errors are clustered by school location. Universal FDK availability is used as an instrument for FDK enrolment. Control variables include gender, Aboriginal identity, language spoken at home, and student's age.
(* = significant at 10%, ** = 5%, *** = 1%)

Table 2.7: Placebo regressions, "effect" of universal FDK program availability on pre-treatment variables for Kindergarten students (administrative and BCSAS data).

Effect of interest		**Administrative data**		**BCSAS**
		Three cohorts	Two cohorts	Two cohorts
Placebo effect of universal FDK program availability on:				
	Male	0.006	0.002	-0.033
		(0.019)	(0.011)	(0.028)
	Aboriginal identity	0.001	0.005	-0.002
		(0.005)	(0.005)	(0.012)
	Speaks a language other than English at home	0.035***	0.043***	0.008
		(0.008)	(0.010)	(0.019)
	Student's age in months	0.113*	0.141*	0.073
		(0.065)	(0.074)	(0.247)
	Parent is HS dropout			-0.011
				(0.008)
	Parent has some PS			0.010
				(0.029)
	Parent has university degree or higher			-0.003
				(0.025)
	Lone parent			0.005
				(0.014)
	Responding parent is not mother			0.000
				(0.018)
	Parent's age in years			-0.089
				(0.311)

All regressions include year and school fixed effects, and standard errors are clustered by school location. No additional control variables.
(* = significant at 10%, ** = 5%, *** = 1%)

Chapter 3

The Role of Financial Incentives in the Decline of Population Growth Rate

Abstract

There has been a long debate among economists and policy makers over the effectiveness of population planning programs. The estimated program effects in the literature vary substantially. One such program is the Iranian 1993 Population Control Law that withdrew paid maternity leave and social welfare subsidies in the case of children of fourth and higher parities. We use data from publicly available sample 2006 census data in Iran and the annual Household Expenditure and Income Surveys (HEIS: 1988-2005) to estimate the effect of this policy on fertility outcomes. Our difference in difference method compares the change in probability of having birth in families with fewer than three children prior to the legislation to the change in probability of having birth of families with three or more children. We find that the legislation had a modest effect of 8 to 13 percent on decreasing the probability of a fourth or higher birth. The law has the highest impact after four years of implementation and after that effect size gradually goes away.

Keywords: Multigrade classes; non-cognitive skills; behavior problems; hyperactivity; program evaluation; population control policy

3.1 Introduction

High population growth rates (or as a closely related parameter, high total fertility rates) are generally deemed undesirable by economists and politicians alike. Becker (1960) and Becker and Lewis (1974) argue that as number of children in a household increases, “quality" of the children decreases, due to the limited resources which can be spent on them. High fertility rates are often associated with a country being underdeveloped as investment in human capital is considered to be a key factor for development (Romer et al., 1990; Rebelo, 1991). Family planning is also associated with improved maternal health, increase in mothers' years of schooling and access to better job opportunities (Miller, 2010; Banerjee and Duflo, 2012). Last, but not least, is the important effect family planning programs can have on the environment. Many economists, perhaps most sonorously voiced by Sachs (2008) in recent years, have talked about irreversible damages that high population has inflicted upon the earth. For all these reasons, implementing the right population control policy at the right time is of utmost importance, especially in developing and underdeveloped countries.

Many population control policies were introduced in developing countries in the 1960s Cleland et al. (2006) and some are still in place: PROFAMILIA program in Colombia and the network of rural health houses in Iran provided easy and cheap access to contraceptives. Iran also began educating families about necessity and methods of family planning through health houses, pre-marital classes and a mandatory family planning course in universities after Iran-Iraq war ended in 1989. India has been paying couples to delay child bearing and have smaller family sizes. China's infamous one child policy and Iran's removal of government subsidies for children of fourth and higher parities are examples of legislative/punitive actions taken by governments. Providing greater access to medical facilities (like expansion to health house network in Iran) and promoting delayed marriages in countries like Kenya are other examples of policies employed to slow down population growth.

There has been a long debate among economists over the effectiveness of population control policies. Demand-side economists see fertility decline mostly a result of reduction in desired number of children. More education, increase in the opportunity cost of women and industrialization are some of the reasons for decrease in the demand for children ([see Pritchett, 1994; Breierova and Duflo, 2004; Lavy and Zablotsky, 2015). On the other hand, supply-side economists believe fertility decline is driven by the easier access to/prevalence of family planning techniques (Bongaarts et al., 2012, 1990; Robey et al., 1993). Miller and Babiarz (2016) give a comprehensive review of micro-level empirical studies on the effectiveness of family planning programs in middle and low-income countries and conclude that: “Although effect sizes are heterogeneous, long-term studies imply that in practice, family planning programs may only explain a modest share of fertility decline in real-world settings (explaining 4-20% of fertility decline among studies finding significant effects).

Family planning programs may also have quantitatively modest - but practically meaningful - effects on the socio-economic welfare of individuals and families."

Among the countries with family planning programs in place to control population growth, Iran is a unique example. It is named as one of the most successful examples (if not the most) of population control. The country's fertility rate dropped from 6.52 in 1981 to 1.87 in 2006 (a 71% drop in 25 years). Its population growth rate also decreased from 3.91 to less than 1.17 in the same period. In different points in time, Iranian government put different population control policies into action. Fo r these reasons, we think it deserves special attention.

There has been a few attempts to study the effects of these different policies on Iran's fertility rate. Raftery et al. (1995) look at the family planning programs in the Shah's regime (prior to 1979) and find no evidence that the Family Planning Program or Family Protection Act of 1967 resulted in the decline in marital fertility. Salehi-Isfahani et al. (2010) use difference in difference method and the timing of establishment of rural health houses between 1986 and 1996 to estimate the change in average fertility rate in villages in which a health house was established after 1986 and before 1996 compared to villages which didn't have a health house in either year. They find that child woman ratio declined (at most) 20 percent more in their treatment groups compared to their comparison group. Hashemi and Salehi-Isfahani (2013) try to estimate the impact of the health houses in rural areas on the hazard rate of first to third child. They study three different periods: 1979 to 1988 in which health houses offered health care services to mothers and children, 1989-1994 in which health houses began to offer birth control options and 1994-2000, the period after parliament passed Population Control Law. They find that health houses decreased the hazard of second and third child after 1989. These results, although substantial, do not explain the sharp decrease in fertility entirely. We believe that apart from establishment of health houses, other factors (possibly other components of Iran's family planning program) are responsible for the country's fertility decline. One of these components is the Population Control Law (hereafter PCL or 1993 PCL) which was passed in May 1993 and came into effect in May 1994.

The Population Control Law (explained in detail in section 2) canceled paid maternity leave and some government paid subsidies for children of fourth and higher parities born after May 1994. It also stressed the need for public education and informational campaigns about the benefits of smaller families.

Iran's 1993 Population Control Law (hereafter PCL or 1993 PCL), and in particular, withdrawal of government subsidies likely affected the fertility decisions of families who would be potentially affected by it. In this paper, we want to study the effects of the 1993 Population Control Law on the probability of having a child of fourth or higher order parities. To the best of our knowledge, this is the first paper that looks at this particular question. In order to evaluate the effectiveness of this law on reducing fertility, we employ a difference in difference strategy and take families with three or more children as treated and families with less than three children as the untreated group. Our results show that this policy had a modest effect of 8% on reducing probability of having one more child for families with three or more children. It had its maximum impact in four years and the effect size gradually decreased after that. These results are in line Miller's conclusion about the moderate share of family planning programs in fertility decline around the world in general Miller and Babiarz (2016).

The rest of this paper is organized as follows: Section 3.2.1 describes the history of family planning programs in Iran, and in particular 1993 Population Control Law. We also talk about the data sets we are using in Section 3.2.2. In section 3.3, we talk about our methodology and empirical strategy in detail. Regression results are explained in section 3.4. Finally, we conclude in section 3.5.

3.2 Institutional Background and Data

3.2.1 History of Family Planning Programs in Iran

After the Islamic Revolution in 1979, government authorities canceled the existing family planning programs in Iran. International organizations were forced to leave the country and all family planning activities were shut down in the Ministry of Health. The Islamic government proclaimed family planning as a western (and political) concept designed to weaken the country.

In addition to this, government started to support early age marriage and larger families by providing cheaper housing to larger families and exemption of men with more than three children from compulsory military service. All these plus the influx of immigrants to Iran from neighboring countries, like Afghanistan, resulted in a peak population growth rate of 3.9% in 1981 Pilehvari (2008).

Being left to deal with catastrophic damages to its infrastructure after Iran-Iraq war and the demands of a rapidly growing population, Iranian Government initiated different projects to promote family planning and smaller family size. The country's health care network started supplying contraception to married couples. In rural areas an existing system of so called "Health Houses" began growing rapidly. These health houses educated families about family planning and also provided birth control means.

One of famous campaigns starting in 1989 in Iran, was the campaign to the promote smaller family size with two simple, yet effective slogans: "Stop! Two children, tops' and "Fewer Children; Better Life". These slogans were painted on the walls, displayed on public transit and printed on milk bottles and grocery packaging. Being short and rhythmical, they were well remembered and frequently recited.

Population Control Law, passed in May 1993 and fully in effect after May 1994, canceled paid maternity leave, child care and health care subsidies for the children of fourth and higher parities who were born after May 1994. Food subsidies (then given in the form of coupons) were also withdrawn for those children. The law required Ministry of Education to effectively include topics related to population control and mother and child health, in the school curriculum. Ministry of Science, Research and Technology and Ministry of Health were asked to introduce "Population and Family Planning" as a mandatory course for all university majors. Ministry of Culture was asked to cooperate with journalists, film makers and other artists to use different mediums to encourage smaller family sizes. Finally, Iranian Broadcasting Organization was required to produce different educational programs to (directly or indirectly) educate public about mother and child health.

Iran 1979 Labor Law allowed three month paid maternity leave for every child. It also permitted the female workers to take a paid half an hour leave per three hours of work ("Child Feeding Time") and required the employers to provide child care facilities for female workers. These were all to be withdrawn for the children of fourth and higher parities from May 1994. The employee's share of monthly health care premiums (1.65% of monthly base wage for the whole family until then) did not cover children of fourth and higher parities after 1993 PCL was in effect. For each additional child born after May 1994, families had to pay approximately 1.5% of monthly base wage as health insurance premium; so a fourth child would have effectively doubled the health insurance cost of the family.

Lastly, total subsidies paid to consumers, producers and service providers was 3,686.3 billion Rials in 1994, 85% of which was consumption subsidies (63,194 Rials per person)[1] Ahmadvand and Eslami (2005). A significant part of these consumption subsidies were paid in the form of coupons (or rations) and were withdrawn for children of fourth and higher parities after 1994.

3.2.2 Data

Our main data comes from the publicly available 2% random sample of the 2006 Iranian census data[2]. The census used to take place every ten years prior to 2006 and every five years after that. It gathers information on household and individual characteristics including, family composition, month and year of birth, marital status, education, employment status, occupation, religion, citizenship and residency. Summary statistics for the key variables are reported in Table 3.1 for families with one or two children and families with three or four children (comparison and treatment groups in our difference in difference strategy).

Using census data, we construct birth records which in turn are used to define treatment status and outcome variable. Our treatment group consists of married couples with three or more children and the comparison group consists of married couples with one or two children. We limit our sample to families with the mother present in the household and between 15 to 50 years old.

[1]For comparison, note that minimum monthly wage was 116,820 Rials in the same year.

[2]Publicly available at www.amar.org.ir.

Table 3.1: Summary Statistics - 2006 Iranian census

	Families with 1 or 2 children	**Families with 3 or 4 children**
% Female Children:	46.01	46.65
Mother's Age:	35.76	41.21
Mother's Education:		
Illiterate	15.55	26.41
Primary	24.23	34.13
Secondary	44.94	29.11
Post Secondary	12.94	7.38
Unofficial Education	2.34	2.98
Father's Education:		
Illiterate	21.34	38.80
Primary	26.44	36.37
Secondary	42.09	20.84
Post Secondary	8.52	2.32
Unofficial Education	1.61	1.67
% Employed or have income:		
Mother:	12.69	9.41
Father:	92.53	92.01

Table 2.1 suggests that women with more children are on average older, probably because fertility is not completed for younger women in our sample. Also, mothers with fewer children are more likely to have higher education and to be employed. This may reflect the higher opportunity cost of having children for more educated and working women.

One important limitation of using the birth records constructed from census data for our purposes is that it will not allow us to control for any further covariates other than the very basic policy variables we need plus parents' age.

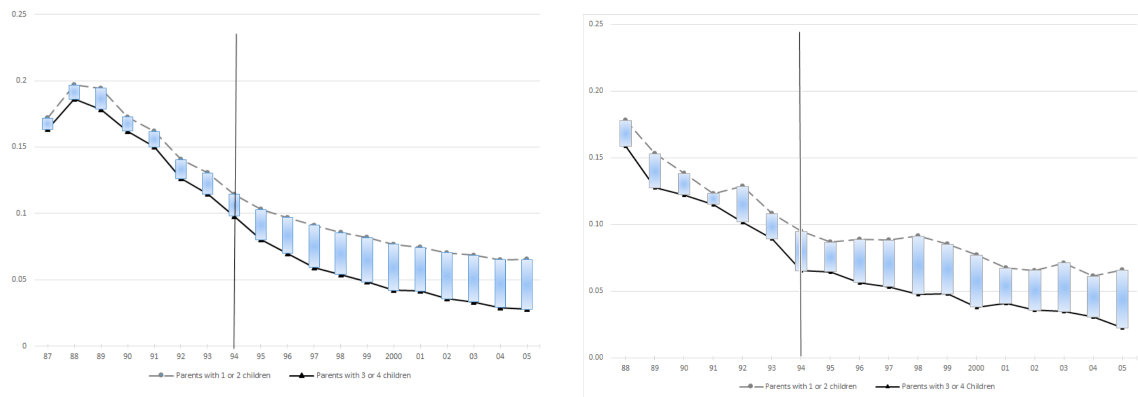
Our empirical strategy is taking care of any time-variant unobserved factors affecting our comparison and treatment groups in the same way (like access to contraceptives through health-houses) by using year fixed effects. However, there may be other factors changing systematically differently for our comparison and treatment groups over time.

Mothers with one or two children are on average younger with less household responsibility and more free time. As a result, they are more likely to participate in the labor force or continue their education, when having the opportunity to do so. One such opportunity arose with the rapid expansion of the number of universities in Iran. Mothers with fewer children were more likely to attend a university. If this increase in education further effects the outcome variable, we may over or underestimate the true effect of the Population Control Policy.

To address this issue, we use a secondary data source that allows us to control for mothers' education and employment status over the years. This data set is the annual Household Expenditure and Income Surveys (HEIS: 1988-2005) conducted by the Statistical Center of Iran. The survey gathers information on demographic characteristics for each member of the surveyed family, such as age, sex, education, employment status, marital status, and the relationship to the head of the household . It also collects detailed information on household income and expenditure. The HEIS is nationally representative and the coverage rate has increased from more than 11000 households in 1989 to more than 38000 households in 2010. Once again, we limit our sample to married couples where the mother is present in the family and is in the 15-50 age range.

The most important shortcoming of the HEIS, for our purpose, is that it does not report the year and month of birth. As the survey is conducted throughout the year, some of our important variables, including our outcome variable (having given birth in a particular year) is subject to measurement error. Consider the following example: The 1993 cycle of HEIS begins in March 1993 (March marks the beginning of the new year in Iran) and ends in April 1994. A family who gave birth to their third child in April 1992 will have a record of new born (age=0), if surveyed in March 1993. Another family with their third child born in April 1994 and surveyed in the same month will also have a record of a child with age zero in 1993 HEIS cycle. Both of these families will be considered as treated in our analysis, while the latter actually should be in our comparison group, as they had two children in May 1993. The outcome variable, whether the family had a newborn child in a given year,

Figure 3.1: Conditional Probability of giving birth to another child



(a) (b)

Source: Authors' calculation based on 2% sample of 2006 Iranian Census (a) and HEIS 1988-2005 (b), publicly available at www.amar.org.ir
Notes: Horizontal axis shows the years. Vertical axis shows the average percentage of mothers who have given birth in each year.

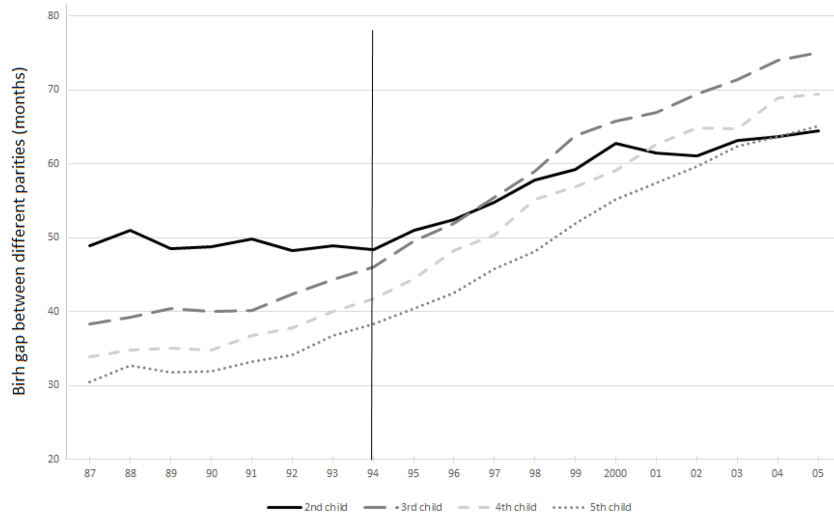
will be equal to one for both of these families, whereas neither family actually gave birth to a new child in 1993.

As the sample census data contains information on the month and year of birth, it will dramatically reduce the measurement error of the outcome variable. For this reason, we are using census as our main data source, despite its shortcomings.

Panels (a) and (b) in Figure 3.1, illustrate the percentage of families who gave birth to another child over the years for our comparison and treatment groups. Panel (a) is using the birth records constructed from 2006 census data and panel (b) is using HEIS data. We observe similar trends in both graphs. On average, the probability of having another child decreases as time goes by for both our treatment and comparison groups, and irrespective of which data set is used . However one can easily see that until 1994, when the PCL came into effect, the difference between the two groups of is small and remains constant. After 1994, this gap becomes wider and remains so over the time. In other words, families who are affected by 1993 Population Control Law become less likely to have another child.

In our opinion, PCL could have affected fertility decisions of the families through two possible channels: It might have reduced the desired number of children (total fertility rate) and/or it might have increased the gap between different parities. We are not able to separate the effects of these two mechanisms in our model, however we believe the policy has worked through both channels. Figure 3.2 illustrates the average birth spacing between different parities over the years. The solid line in Figure 3.2 shows the average wait time between first and second child in months.

Figure 3.2: Birth spacing for different parities



Source: Authors' calculation based on publicly available 2% sample of 2006 Iranian Census at www.amar.org.ir.
Notes: The horizontal axis shows the years and the vertical axis shows the average waiting time for each parity in months.

These patterns are consistent with the results of Hashemi and Salehi-Isfahani (2013). After 1989, when the Iran-Iraq war ended, and the reversal of pro-natalist policies of Iranian government, waiting time for the 3rd and higher order births gradually increased over time. However it appears that the average wait time for second child was not affected by those policies and remained just above 48 months up until 1994. It is only after this year that the average wait time for the second child starts to increase and within 6 years, it reaches 60 months. This may well be the result of the 1993 Population Control Law and the campaign to educate public about family planning.

3.3 Methodology

We are interested in the effect of 1993 Population Control Law on the probability of having children of fourth or higher parities. As this policy penalizes families with three or more children who give birth after May 1994, the natural comparison group would be families who had less than three children. Our research design is based on the following difference in difference framework:

$$Y_{it} = \alpha + \beta D_{it} + \mu_t + \sigma PCL_{it} + \theta Age_{it} + \epsilon_{it} \tag{3.1}$$

where Y_{it}, our outcome variable, is an indicator of whether or not individual i has given birth to a new child in year t, μ_t denotes a set of year fixed effects and Age_{it} controls for mother i's age in year t. The treatment status, D_{it} is an indicator for our treatment group (families with 3 or more children before the introduction of the PCL in 1994), and PCL_{it} is the interaction between D_{it} and an indicator for years after 1994.

Defining D_{it} and PCL_{it} based on the actual treatment status in year t, will violate the random assignment of treatment that assumes treatment status should not change with the outcome variable. So, to capture the causal effect of the treatment, we base our analysis on the initial assignment of individuals to treatment and comparison groups and estimate the intention to treat (ITT) effect.

As mentioned earlier, to see how sensitive our estimates are to the inclusion of socioeconomic status, namely education and employment, we are using repeated cross-section data of the Household Expenditure and Income Surveys (HIES: 1989-2005). Since these are repeated cross-section surveys, unlike the birth records constructed from 2006 Iranian Census, we are able to unbiasedly estimate the average treatment on treated (ATT) effects.

To further extend our analysis, we are using equation (3.2) to study the evolution of the treatment effect over time.

$$y_{it} = \alpha + \beta D_{it} + \mu_t + \sum_{\tau=-m}^{q} \sigma_\tau PCL_{\tau,it} + \theta Age_{it} + \epsilon_{it} \tag{3.2}$$

where y_{it}, D_{it}, μ_t and Age_{it} are defined as in equation (3.1) and $PCL_{\tau,it}$ is composed of a set of dummy variables that take value one only if individual $i$is in the treatment group and year t is $|\tau|$ years before ($\tau < 0$) or after ($\tau > 0$) the policy came into the effect. By including q lags of the treatment effect into the model, we are able to see how the treatment effect evolves over time. Moreover, inclusion of m leads of the treatment effect into the model allows us to test whether the parallel trend assumption is satisfied. We are expecting to see $\sigma_\tau = 0, \forall \tau < 0$.

We report the results for a variety of treatment and comparison groups, including families with three vs. families with two children and families with three to five children versus families with one or two children.

3.4 Results

Our main results on the effects of 1993 Population Control Law are reported in Table 3.2. The treatment group consists of families with three or four children and the comparison group consists of families with one or two children. All regressions include year fixed effects and standard errors are clustered at household level.

The second column of Table 3.2 is showing the results of our main specification. The estimated effect of Population Control Law on probability of having one more child is -0.66 in the base line of 7.9, a difference of more than 8%. This effect is very precisely estimated (significant at 0.1% level). Comparison of specifications (1) and (2) show that our results are robust to exclusion of mother's age from our model. The negative coefficient of age can be easily explained by observing two facts: First, the chances of getting pregnant decline sharply with mother's age and second, there is a higher probability for older women to have already reached their desired number of children in our sample.

Table 3.2: Effect of Population Control Law on probability of giving birth (2006 census data)

	(1)	(2)	(3)	(4)
	Total		Urban	Rural
	Comparison: Families with 1 or 2 children Treated: Families with 3 or 4 children			
PCL	$-0{\cdot}661^{***}$	$-0{\cdot}665^{***}$	$-0{\cdot}270^{*}$	$-1{\cdot}299^{***}$
	(0·102)	(0·102)	(0·124)	(0·171)
Age		$-0{\cdot}143^{***}$	$-0{\cdot}137^{***}$	$-0{\cdot}157^{***}$
		(0·002)	(0·002)	(0·003)
Mean Dependent Variable	7·914	7·914	7·138	9·883
R-square	0·04	0·04	0·04	0·05
Observations	2666613	2666613	1593666	1072947

Notes: Dependent variable takes the value of 100 if a family gave birth to another child in a given year and zero otherwise. All regressions include year effects and standard errors are clustered at family level.
(* = significant at 5%, ** = 1%, *** = 0.1%)

Columns (3) and (4) in Table 3.2 show the results for urban and rural subsamples. The policy effect is larger in rural areas (−1.299) compared to urban areas (−0.270). This can be attributed to lower desired fertility in urban areas. As an extreme example, if all families in urban areas desire to have three children, the 1993 PCL would have absolutely no effect. Another reason for the gap between rural and urban areas may be the higher sensitivity of rural families to withdrawal of subsidies (in particular, consumption subsidies).

Table 3.3 presents the results of our main specification for various treatment and comparison groups. The top panel in Table 3.3 uses families with three children as treated and families with 2 children as comparison. The middle panel compares families with one or two children with those with three, four or five children. Finally, the bottom panel compares families with less than three children to those with three to six children. The effect sizes are very similar across all versions and range from 8% to 13% and are all statistically significant at 0.1%.

Table 3.3: Effect of Population Control Law on probability of giving birth (2006 census data) - Robustness checks

	(1)	(2)	(3)	(4)
	Total		Urban	Rural
	Comparison: Families with 2 children Treated: Families with 3 children			
PCL	−0·974***	−0·976***	−0·804***	−1·032***
	(0·130)	(0·130)	(0·155)	(0·221)
Mean Dependent Variable	6·994	6·994	6·036	9·458
R-square	0·04	0·04	0·04	0·05
Observations	1366481	1366481	823802	542679
	Comparison: Families with 1 or2 children Treated: Families with 3 to 5 children			
PCL	−0·725***	−0·724***	−0·343**	−1·300***
	(0·099)	(0·098)	(0·120)	(0·163)
Mean Dependent Variable	7·851	7·851	7·080	9·777
R-square	0·04	0·04	0·04	0·05
Observations	2769502	2769502	1648775	1120727
	Comparison: Families with 1 or 2 children Treated: Families with 3 to 6 children			
PCL	−0·750***	−0·745***	−0·366**	−1·297***
	(0·098)	(0·097)	(0·119)	(0·161)
Mean Dependent Variable	7·825	7·825	7·059	9·724
R-square	0·04	0·04	0·04	0·05
Observations	2806288	2806288	1667449	1138839
Age	No	Yes	Yes	Yes

Notes: Dependent variable takes the value of 100 if a family gave birth to another child in a given year and zero otherwise. All regressions include year effects and standard errors are clustered at family level.
(* = significant at 5%, ** = 1%, *** = 0.1%)

Table 3.4 shows the results when we use data from Iranian Household Expenditure and Income Surveys (1988-2005). Columns (1) and (2) report the results for our baseline model without any controls, and with control for mother's age. Although including age dramatically reduces the estimated effect of policy (from 16.9% to 5.6%), inclusion of mother's level of education and employment status do not seem to cause any further change. Employed mothers and those with higher educations exhibit lower probabilities of having one more child, as expected.

The results for various comparison and treatment groups using HEIS (1988-2005) are presented in Table 3.5. We do not see any significant effects at conventional significance levels for families with two children vs. families with three children. However for other specifications of comparison and treatment groups, the results are precisely estimated and show the expected sign and magnitude.

Without controls for education and employment status, average treatment on treated effects vary between 18% and 36%. When we further add these two variables, these effects become closer to what we observe with 2006 census data. One must be cautious when comparing the two coefficients, however, as one estimates the average treatment on treated effect and the other estimates intention to treat effect.

Tables 3.4 and 3.5 show that although controlling for age, substantially changes the estimate of the parameter of interest, the coefficients are robust to inclusion of education and employment status. This reassures us that there was nothing else going on around the time of policy that affected our comparison and treatment groups systematically differently.

The estimated policy effect in Tables 3.2 to 3.5 are in fact the average of policy effects over all post-policy periods . To get a sense of dynamics of the effects over time, we estimate equation (3.2). Results are presented in Table 3.6. Our identifying assumption is that in the absence of the treatment, both comparison and treatment groups follow a similar trend. This specification also allows us to test the parallel trend assumption.

All four columns in Table 3.6 show a similar pattern. The estimated coefficients for years leading to adoption date of PCL are mostly small and insignificant. This suggests that our parallel trend assumption holds. For years after the policy, the coefficients are larger and highly significant.

Panel (a) in Figure 3.3 visualizes the results presented in Table 3.6, depicting σ_τs from equation (3.2). We mark 1993, the year the policy came into effect, with a vertical dashed line. We normalize the effects such that they are equal to zero in the year of adoption ($\tau = 0$). The policy has its greatest impact in four years after its adoption and the effect starts to become smaller as time goes by. The remaining panels in Figure 3.3 show the results for various comparison and treatment groups.

It is important to notice that the health house network continued to expand in rural areas even after 1993 PCL. The rural families tend to be larger (and mostly included in our treatment group). If the health houses reduced the desired number of children, the decline

Table 3.4: Effect of Population Control Law on probability of giving birth (Household Expenditure and Income Surveys, 1999-2005)

	(1)	(2)	(3)
	Comparison: Families with 1 or 2 children Treated: Families with 3 or 4 children		
PCL	−1·342***	−0·449	−0·403
	(0·250)	(0·244)	(0·245)
Age		−0·410***	−0·481***
		(0·004)	(0·005)
Education:			
Primary			−2·538***
			(0·158)
Secondary			−5·596***
			(0·170)
University			−6·038***
			(0·245)
Unofficial Education			−0·766*
			(0·303)
Employed			−0·780*
			(0·303)
Mean Dependent Variable	7·929	7·929	7·872
R-square	0·02	0·07	0·08
Observations	213008	213008	211280

Notes: Dependent variable takes the value of 100 if a family gave birth to another child in a given year and zero otherwise. All regressions include year effects.
(* = significant at 5%, ** = 1%, *** = 0.1%)

Table 3.5: Effect of Population Control Law on probability of giving birth (Household Expenditure and Income Surveys, 1999-2005) - Robustness checks

	(1)	(2)	(3)
	Comparison: Families with 2 children Treated: Families with 3 children		
PCL	0·114	0·595	0·590
	(0·322)	(0·316)	(0·316)
Mean Dependent Variable	7·168	7·168	7·112
R-square	0·03	0·06	0·07
Observations	117509	117509	116613
	Comparison: Families with 1 or 2 children Treated: Families with 3 to 5 children		
PCL	$-1{\cdot}507^{***}$	$-0{\cdot}652^{**}$	$-0{\cdot}643^{**}$
	(0·232)	(0·227)	(0·228)
Mean Dependent Variable	7·816	7·816	7·756
R-square	0·03	0·07	0·07
Observations	248096	248096	245982
	Comparison: Families with 0 to 2 children Treated: Families with 3 to 6 children		
PCL	$-3{\cdot}070^{***}$	$-2{\cdot}193^{***}$	$-2{\cdot}122^{***}$
	(0·211)	(0·205)	(0·206)
Mean Dependent Variable	8·323	8·323	8·276
R-square	0·02	0·08	0·08
Observations	308906	308906	305998
Age	No	Yes	Yes
Education	No	No	Yes
Employment Status	No	No	Yes

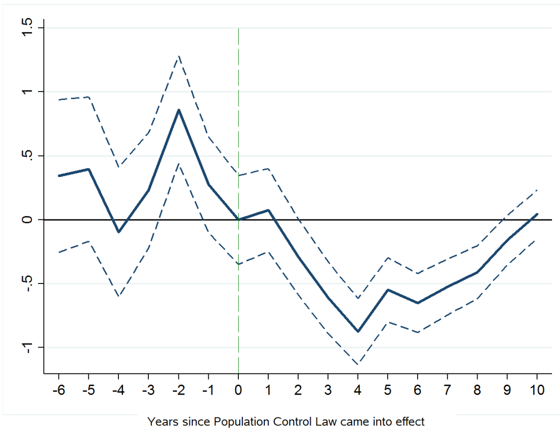
Notes: Dependent variable takes the value of 100 if a family gave birth to another child in a given year and zero otherwise. All regressions include year effects.
(* = significant at 5%, ** = 1%, *** = 0.1%)

Table 3.6: Effect of Population Control Law on probability of giving birth over time (2006 census data)

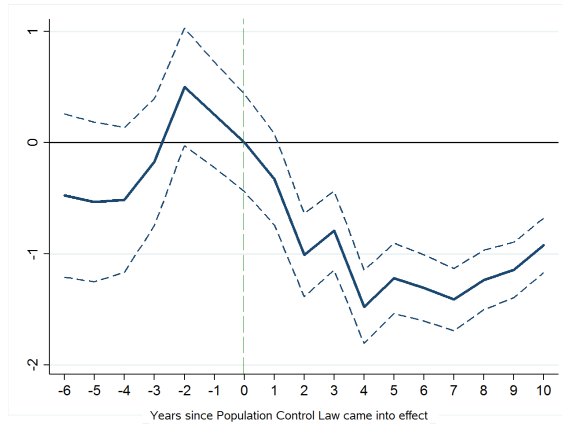
	(1)	(2)	(3)	(4)
	Total		Urban	Rural
	Comparison: Families with 1 or 2 children Treated: Families with 3 or 4 children			
PCL_{-6}	−0·032	−0·010	−0·689	1·537*
	(0·363)	(0·363)	(0·449)	(0·601)
PCL_{-5}	0·025	0·042	−0·392	0·908
	(0·343)	(0·343)	(0·424)	(0·568)
PCL_{-4}	−0·429	−0·450	−0·450	−0·695
	(0·309)	(0·309)	(0·384)	(0·505)
PCL_{-3}	−0·092	−0·124	−0·501	0·275
	(0·274)	(0·274)	(0·335)	(0·469)
PCL_{-2}	0·544*	0·506*	0·284	0·399
	(0·256)	(0·256)	(0·313)	(0·438)
PCL_{-1}	−0·044	−0·079	−0·682*	0·570
	(0·228)	(0·227)	(0·273)	(0·404)
PCL_{0}	−0·327	−0·354	−0·924***	0·153
	(0·211)	(0·211)	(0·254)	(0·377)
PCL_{+1}	−0·253	−0·279	−0·762**	0·171
	(0·198)	(0·198)	(0·237)	(0·355)
PCL_{+2}	−0·619***	−0·644***	−1·167***	0·019
	(0·181)	(0·181)	(0·215)	(0·330)
PCL_{+3}	−0·939***	−0·962***	−1·223***	−0·753*
	(0·171)	(0·171)	(0·205)	(0·307)
PCL_{+4}	−1·207***	−1·229***	−1·259***	−1·456***
	(0·158)	(0·158)	(0·190)	(0·282)
PCL_{+5}	−0·881***	−0·902***	−0·971***	−0·943***
	(0·152)	(0·152)	(0·185)	(0·268)
PCL_{+6}	−0·985***	−1·004***	−0·971***	−1·239***
	(0·140)	(0·140)	(0·168)	(0·252)
Age	No	Yes	Yes	Yes
Mean Dependent Variable	7·914	7·914	7·138	9·883
R-squared	0.04	0.04	0.04	0.05
Observations	2666613	2666613	1593666	1072947

Notes: Dependent variable takes the value of 100 if a family gave birth to another child in a given year and zero otherwise. All regressions include year effects and standard errors are clustered at family level.
(* = significant at 5%, ** = 1%, *** = 0.1%)

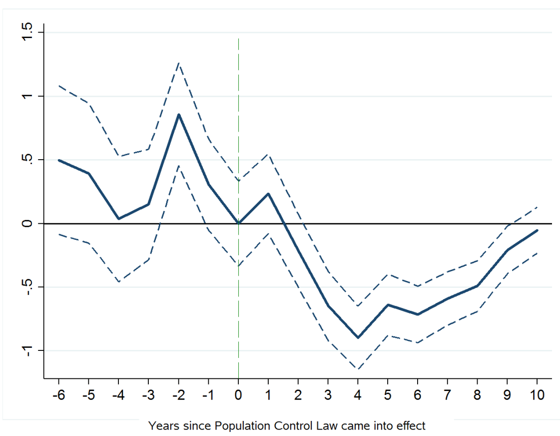
Figure 3.3: Effect of Population Control Law over time



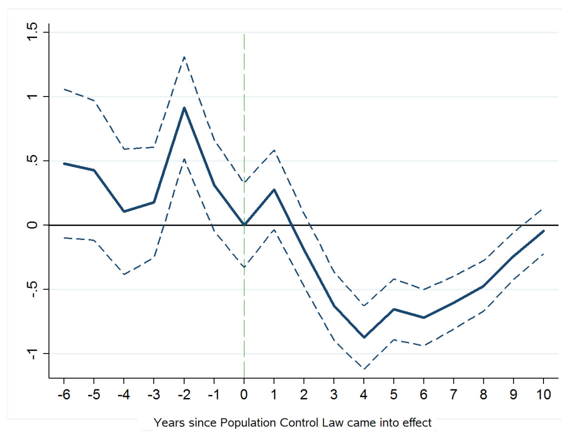
(a) Parents with 1-2 children vs. 3-4 children



(b) Parents with 2 children vs. 3 children



(c) Parents with 1-2 children vs. 3-5 children



(d) Parents with 1-2 children vs. 3-6 children

Notes: The figures provide graphical analysis of the effect of the Population Control Law over time. The year of the policy's adoption (1993) is set equal to zero and the effects are normalized so that they are equal to zero in the adoption year. The blue dashed lines show the 90 percent confidence intervals for the estimated coefficients.

in probability of having another child in rural areas might be wrongfully attributed to the policy and lead to overestimation of its effects. Consequently, the estimated coefficients must be interpreted with caution.

3.5 Concluding Remarks

There has been a long debate among economists on the effectiveness of different population control policies. In this paper, we had a closer look at one of these policies, implemented in Iran in 1993. We find that this policy had a small to moderate effect on reducing the probability of having one more child. Although we are not able to identify the exact mechanism through which the policy acted, which leaves the door open for further research, we have reasons to believe either channel would have resulted in lower population growth rates. Either the policy had changed the demand for children (reducing total fertility), or it had resulted in families spacing out their children. Even in the latter case, the fact that the female fertility declines sharply after a certain age, would mean fewer births over a woman's lifetime.

We would also like to draw readers' attention to the fact that effects of the policy are not constant over time. The policy has its biggest impact in the years immediately following its implementation date. However as time passes, these effects become less pronounced. Researchers who are trying to evaluate such policies, must pay extra attention to the time span they choose.

Bibliography

Achenbach, T. M., McConaughy, S. H. and Howell, C. T. (1987), 'Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity.', *Psychological Bulletin* **101**(2), 213.

Ahmadvand, M. and Eslami, S. (2005), 'A brief study of government subsidies paid between 1973-2003', *Journal of Commercial Surveys* **13**(3), 3–15.

Angrist, J. D., Battistin, E. and Vuri, D. (2014), In a small moment: class size and moral hazard in the mezzogiorno, Technical report, National Bureau of Economic Research.

Angrist, J. D. and Lavy, V. (1999), 'Using Maimonides' rule to estimate the effect of class size on scholastic achievement', *The Quarterly Journal of Economics* **114**(2), 533–575.

Angrist, J. and Imbens, G. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Baker, M., Gruber, J. and Milligan, K. (2008), 'Universal child care, maternal labor supply, and family well-being', *Journal of Political Economy* **116**(4), 709–745.

Baker, M., Gruber, J. and Milligan, K. (2015), Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program, Working Paper 21571, National Bureau of Economic Research.

Banerjee, A. and Duflo, E. (2012), *Poor economics: A radical rethinking of the way to fight global poverty*, PublicAffairs.

Becker, G. S. (1960), An economic analysis of fertility, *in* 'Demographic and economic change in developed countries', Columbia University Press, pp. 209–240.

Becker, G. S. and Lewis, H. G. (1974), Interaction between quantity and quality of children, *in* 'Economics of the family: Marriage, children, and human capital', University of Chicago Press, pp. 81–90.

Bedard, K. and Dhuey, E. (2006), 'The persistence of early childhood maturity: International evidence of long-run age effects', *The Quarterly Journal of Economics* pp. 1437–1472.

Berlinski, S., Galiani, S. and Gertler, P. (2009), 'The effect of pre-primary education on primary school performance', *Journal of Public Economics* **93**(1), 219–234.

Bitler, M. P., Hoynes, H. W. and Domina, T. (2014), Experimental evidence on distributional effects of Head Start, Technical report, National Bureau of Economic Research.

Blau, D. and Currie, J. (2006), 'Pre-school, day care, and after-school care: Who's minding the kids?', *Handbook of the Economics of Education* **2**, 1163–1278.

Bongaarts, J., Cleland, J., J.W., T., J.T., B. and M.D., G. (2012), 'Family planning programs for the 21st century', *New York: Population Council* .

Bongaarts, J., Mauldin, W. P. and Phillips, J. F. (1990), 'The demographic impact of family planning programs', *Studies in family planning* **21**(6), 299–310.

Borghans, L., Duckworth, A. L., Heckman, J. J. and Ter Weel, B. (2008), 'The economics and psychology of personality traits', *Journal of Human Resources* **43**(4), 972–1059.

Breierova, L. and Duflo, E. (2004), The impact of education on fertility and child mortality: Do fathers really matter less than mothers?, Technical report, National Bureau of Economic Research.

Cannon, J. S., Jacknowitz, A. and Painter, G. (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, J. S., Jacknowitz, A. and Painter, G. (2011), 'The effect of attending full-day Kindergarten on English learner students', *Journal of Policy Analysis and Management* **30**(2), 287–309.

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

Caspi, A., Wright, B. R. E., Moffitt, T. E. and Silva, P. A. (1998), 'Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood', *American Sociological Review* **63**(3), 424–451.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011), 'How does your kindergarten classroom affect your earnings? evidence from project star.', *Quarterly Journal of Economics* **126**(4).

Child Trends (2013), 'Full day Kindergarten', http://www.childtrends.org/?indicators=fullday-kindergarten. Accessed: 2014-03-24.

Cleland, J., Bernstein, S., Ezeh, A., Faundes, A., Glasier, A. and Innis, J. (2006), 'Family planning: the unfinished agenda', *The Lancet* **368**(9549), 1810–1827.

Cooper, H., Allen, A. B., Patall, E. A. and Dent, A. L. (2010), 'Effects of full-day Kindergarten on academic achievement and social development', *Review of Educational Research* **80**(1), 34–70.

Cunha, F. and Heckman, J. J. (2008), 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation', *Journal of Human Resources* **43**(4), 738–782.

Cunha, F., Heckman, J. J., Lochner, L. and Masterov, D. V. (2006), 'Interpreting the evidence on life cycle skill formation', *Handbook of the Economics of Education* **1**, 697–812.

Cunha, F., Heckman, J. J. and Schennach, S. M. (2010), 'Estimating the technology of cognitive and noncognitive skill formation', *Econometrica* **78**(3), 883–931.

Currie, J. and Almond, D. (2011), 'Human capital development before age five', *Handbook of Labor Economics* **4**, 1315–1486.

Currie, J. and Stabile, M. (2006), 'Child mental health and human capital accumulation: the case of ADHD', *Journal of Health Economics* **25**(6), 1094–1118.

Currie, J. and Stabile, M. (2009), Data appendix to "Mental health in childhood and human capital", *in* J. Gruber, ed., 'The Problems of Disadvantaged Youth: An Economic Perspective', National Bureau of Economic Research.
URL: *http://ideas.repec.org/h/nbr/nberch/11105.html*

Currie, J., Stabile, M., Manivong, P. and Roos, L. L. (2010), 'Child health and young adult outcomes', *Journal of Human Resources* **45**(3), 517–548.

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

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

Dee, T. S. and West, M. R. (2011), 'The non-cognitive returns to class size', *Educational Evaluation and Policy Analysis* **33**(1), 23–46.

Deming, D. (2009), 'Early childhood intervention and life-cycle skill development: Evidence from head start', *American Economic Journal: Applied Economics* pp. 111–134.

Dhuey, E. (2011), 'Who benefits from Kindergarten? Evidence from the introduction of state subsidization', *Educational Evaluation and Policy Analysis* **33**(1), 3–22.

Duncan, G. J. and Magnuson, K. (2011), The nature and impact of early achievement skills, attention skills, and behavior problems, *in* G. J. Duncan and R. J. Murnane, eds, 'Whither Opportunity', Russell Sage Foundation, pp. 47–70.

Duncan, G. J. and Magnuson, K. (2013), 'Investing in preschool programs', *The Journal of Economic Perspectives* **27**(2), 109.

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C. and Tien, A. (2004), Center for Epidemiologic Studies depression scale: Review and revision (CESD and CESD-R)., *in* M. E. Maruish, ed., 'The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 3rd edition', Lawrence Erlbaum Associates Publishers.

Elder, T. E. (2010), 'The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates', *Journal of Health Economics* **29**(5), 641–656.

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

Farmer, E. M. (1993), 'Externalizing behavior in the life course the transition from school to work', *Journal of Emotional and Behavioral Disorders* **1**(3), 179–188.

Farmer, E. M. (1995), 'Extremity of externalizing behavior and young adult outcomes', *Journal of Child Psychology and Psychiatry* **36**(4), 617–632.

Figlio, D. and Roth, J. (2009), The behavioral consequences of pre-Kindergarten participation for disadvantaged youth, *in* J. Gruber, ed., 'The Problems of Disadvantaged Youth: An Economic Perspective', University of Chicago Press, pp. 15–42.

Finn, J. D. (1998), 'Class size and students at risk. what is known? what is next? a commissioned paper.'.

Finn, J. D., Fulton, D., Zaharias, J. and Nye, B. A. (1989), 'Carry-over effects of small classes', *Peabody Journal of Education* pp. 75–84.

Finn, J. D., Pannozzo, G. M. and Achilles, C. M. (2003), 'The "why's" of class size: Student behavior in small classes', *Review of Educational Research* **73**(3), 321–368.

Finn, J., Gerber, S., Achilles, C. and Boyd-Zaharias, J. (2001), 'The enduring effects of small classes', *The Teachers College Record* **103**(2), 145–183.

Fletcher, J. M. (2014), 'The effects of childhood ADHD on adult labor market outcomes', *Health Economics* **23**(2), 159–181.

Fletcher, J. and Wolfe, B. (2008), 'Child mental health and human capital accumulation: the case of ADHD revisited', *Journal of Health Economics* **27**(3), 794–800.

Fletcher, J. and Wolfe, B. (2009), 'Long-term consequences of childhood ADHD on criminal activities', *The Journal of Mental Health Policy and Economics* **12**(3), 119.

Frank, R. G. and Meara, E. (2009), The effect of maternal depression and substance abuse on child human capital development, Working Paper 15314, National Bureau of Economic Research.

Fredriksson, P., Ockert, B. and Oosterbeek, H. (2013), 'Long-term effects of class size', *The Quarterly Journal of Economics* **128**(1), 249–285.

Friesen, J., Krauth, B. and Sattari, R. (2015), 'Full day Kindergarten and the development of non-cognitive skills', *Working Paper* .

Garces, E., Thomas, D. and Currie, J. (1992), 'Longer-term effects of head start', *The American Economic Review* (4).

Garces, E., Thomas, D. and Currie, J. (2002), 'Longer-term effects of Head Start', *American Economic Review* **92**(4), 999–1012.

Gelbach, J. B. (2002), 'Public schooling for young children and maternal labor supply', *American Economic Review* **92**(1), 307–322.

Gibbs, C., Ludwig, J. and Miller, D. L. (2011), Does Head Start do any lasting good?, Technical report, National Bureau of Economic Research.

Gibbs, C. R. (2012), Experimental and Quasi-experimental Evidence on the Impact of Full-day Kindergarten, PhD thesis, University of Chicago.

Gormley, W. T. and Gayer, T. (2005), 'Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program', *Journal of Human Resources* **40**(3), 533–558.

Hashemi, A. and Salehi-Isfahani, D. (2013), 'From health service delivery to family planning: the changing impact of health clinics on fertility in rural iran', *Economic Development and Cultural Change* **61**(2), 281–309.

Heckman, J. J., Stixrud, J. and Urzua, S. (2006), 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics* **24**(3), 411–482.

Lavy, V. and Zablotsky, A. (2015), 'Women's schooling and fertility under low female labor force participation: Evidence from mobility restrictions in israel', *Journal of Public Economics* **124**, 105–121.

Leuven, E., Oosterbeek, H. and Rønning, M. (2008), 'Quasi-experimental estimates of the effect of class size on achievement in Norway*', *The Scandinavian Journal of Economics* **110**(4), 663–693.

Leuven, E. and Rønning, M. (2014), 'Classroom grade composition and pupil achievement', *The Economic Journal* .

Little, A. W. (2005), 'Learning and teaching in multigrade settings', *Background paper for UNESCO (2005) EFA Global Monitoring Report* .

Ludwig, J. and Miller, D. L. (2007), 'Does head start improve children's life chances? evidence from a regression discontinuity design', *The Quarterly Journal of Economics* **122**(1), 159–208.

Ludwig, J. and Phillips, D. (2007), 'The benefits and costs of Head Start', *Society for Research in Child Development* **21**(3).

Magnuson, K. A., Ruhm, C. and Waldfogel, J. (2007), 'Does pre-Kindergarten improve school preparation and performance?', *Economics of Education Review* **26**(1), 33–51.

Manski, C. F. (1990), 'Nonparametric bounds on treatment effects', *American Economic Review* **80**(2), 319–323.

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

Miller, G. (2010), 'Contraception as development? new evidence from family planning in colombia', *The Economic Journal* **120**(545), 709–736.

Miller, G. and Babiarz, K. S. (2016), 'Family planning program effects: Evidence from microdata', *Population and Development Review* **42**(1), 7–26.

Ministry of Education, British Columbia (2014), 'School Act'.
URL: *http://www2.gov.bc.ca/gov/content/health/about-bc-s-health-care-system/office-of-the-provincial-health-officer/laws-related-to-health-in-bc/school-act*

Ministry of Education, British Columbia (2015), 'K-12 funding allocation system'.
URL: *http://www.bced.gov.bc.ca/k12funding/*

Mulryan-Kyne, C. (2005), 'The grouping practices of teachers in small two-teacher primary schools in the republic of Ireland', *Journal of Research in Rural Education* **20**(17), 1–14.

Peterson, J. L. and Zill, N. (1986), 'Marital disruption, parent-child relationships, and behavior problems in children', *Journal of Marriage and the Family* **48**(2), 295–307.

Pilehvari, S. (2008), *Family planning programs in Iran*.

Pritchett, L. H. (1994), 'Desired fertility and the impact of population policies', *Population and Development Review* pp. 1–55.

Radloff, L. S. (1977), 'The CES-D scale: A self-report depression scale for research in the general population', *Applied Psychological Measurement* **1**(3), 385–401.

Raftery, A. E., Lewis, S. M. and Aghajanian, A. (1995), 'Demand or ideation? evidence from the iranian marital fertility decline', *Demography* **32**(2), 159–182.

Rebelo, S. (1991), 'Long-run policy analysis and long-run growth', *Journal of Political Economy* **99**(3), 500–521.

Robey, B., Rutstein, S. O. and Morris, L. (1993), 'The fertility decline in developing countries.', *Scientific American* pp. 60–7.

Romer, P. M. et al. (1990), Human capital and growth: Theory and evidence, *in* 'Carnegie-Rochester Conference Series on Public Policy', Vol. 32, Elsevier, pp. 251–286.

Sachs, J. (2008), *Common wealth: Economics for a crowded planet*, Penguin.

Salehi-Isfahani, D., Abbasi-Shavazi, M. J. and Hosseini-Chavoshi, M. (2010), 'Family planning and fertility decline in rural iran: the impact of rural health clinics', *Health Economics* **19**(S1), 159–180.

Sims, D. (2008), 'A strategic response to class size reduction: Combination classes and student achievement in California', *Journal of Policy Analysis and Management* **27**(3), 457–478.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business & Economic Statistics* **20**(4), 518–529.

Thomas, J. L. (2012), 'Combination classes and educational achievement', *Economics of Education Review* **31**(6), 1058–1066.

Urquiola, M. (2006), 'Identifying class size effects in developing countries: Evidence from rural Bolivia', *Review of Economics and Statistics* **88**(1), 171–177.

Urquiola, M. and Verhoogen, E. (2009), 'Class-size caps, sorting, and the regression-discontinuity design', *The American Economic Review* **99**(1), 179–215.

Veenman, S. (1995), 'Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis', *Review of Educational Research* **65**(4), 319–381.

Zhai, F., Raver, C. C. and Jones, S. M. (2012), 'Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings', *Children and Youth Services Review* **34**(5), 946–954.

Appendix A

The B.C. School Arrangements Survey

This appendix provides additional background information on the BCSAS. Twenty-one districts participated fully in the survey. Three additional districts agreed to more limited participation and are excluded from the main analysis. Two only allowed us to write to schools asking for their participation (only one school agreed to participate), and another only allowed us to survey parents of Grade 1 students in 2011-2012 and 2012-2013.

To maximize participation, respondents were offered the option of completing the survey online or by mail. In April or early May, schools were sent a letter to be sent home with each student, inviting parents to participate in the online survey. A few weeks later, schools were sent a reminder letter to be sent home with each student, along with a mail-in version of the survey. All survey materials were made available in English, (traditional) Chinese, and Punjabi to minimize language barriers to participation. Surveys were not distributed for children enrolled in distance education programs or other alternative schooling arrangements. To further encourage participation, a cash prize drawing was offered to schools with participation rates above 30%. The online survey was open from early April to July 25 each year.

The behavior items in the BCSAS are drawn from two sources. The Behavior Problems Index (BPI) was developed by Peterson and Zill (1986) to measure the incidence and severity of child behavior problems in a survey setting. Items from this index appear in a number of prominent survey instruments, including the NLSY, the PSID Child Development Supplement and the ECLS-K. The National Longitudinal Survey of Children and Youth is a Canadian survey used to assess a wide range of child health and behavior outcomes. Table A1 lists the behavior items in the BCSAS and the corresponding BPI and NLSCY items. There is overlap between the items in the NLSCY anxiety/depression scale and the BPI items included in our survey, so the BCSAS includes a total of 26 behavioral and emotional items. Currie and Stabile (2009) provide a discussion of the differences between the BPI and NLSCY emotional items.

The adult depression items are listed in Appendix Table A.2, and were drawn from the revised version of the Center for Epidemiologic Studies Depression (CESD-R) scale, developed by Radloff (1977) and revised by Eaton et al. (2004). The CESD-R scale is widely used in national surveys, including both the NLSY and the NLSCY.

Table A.1: Correspondence between BCSAS, NLSCY and BPI child behavior and emotional health items

BCSAS item	NLSCY anxiety/depression item	BPI item	BPI[1] Scale	BPI Sub-scale
This child...				
		Has sudden changes in mood or feeling	E	Anxious/ Depressed
		Feels/complains no one loves him/her	I	Anxious/ Depressed
Is rather high strung, tense, and nervous	Is nervous, high strung or tense	Is rather high strung, tense, and nervous	E	Headstrong
Cheats or tells lies		Cheats or tells lies	E	Antisocial
Is too fearful or anxious	Is too fearful or anxious	Is too fearful or anxious	E/I	Anxious/ Depressed
Argues too much		Argues too much	E	Headstrong
Has difficulty concentrating/paying attention		Has difficulty concentrating, paying attention	I	Hyperactive
Is easily confused, seems in a fog		Is easily confused, seems in a fog	E/I	Hyperactive
Bullies or is cruel/mean to others		Bullies or is cruel/mean to others	E	Antisocial
Is disobedient at home		Is disobedient at home	E	Headstrong
Does not seem to feel sorry after misbehaving		Does not seem to feel sorry after misbehaving	**	Antisocial
Has trouble getting along with other children		Has trouble getting along with other children	E	Peer Problems
Is impulsive or acts without thinking		Is impulsive or acts without thinking	E	Hyperactive
		Feels worthless or inferior	I	Anxious/ Depressed
Is not liked by other children		Is not liked by other children	E	Peer Problems
Has trouble getting mind off certain thoughts		Has trouble getting mind off certain thoughts	E	Hyperactive
Is restless, overly active, cannot sit still		Is restless, overly active, cannot sit still	E	Hyperactive
Is stubborn, sullen, or irritable		Is stubborn, sullen, or irritable	E	Headstrong
Has strong temper and loses it easily		Has strong temper and loses it easily	E	Headstrong
Seems to be unhappy, sad or depressed	Seems to be unhappy, sad or depressed	Is unhappy, sad, or depressed	E/I	Anxious/ Depressed
Is withdrawn, does not get involved with others		Is withdrawn, does not get involved with others	I	Peer Problems
Breaks things deliberately		Breaks things deliberately	E	Antisocial
		Clings to adults	I	Dependent
Cries a lot	Cries a lot	Cries too much	I	Dependent
		Demands a lot of attention	I	Dependent
		Is too dependent on others	I	Dependent
Is worried	Is worried	Worries too much	***	***
Is disobedient at school		Is disobedient at school	E	Antisocial
Has trouble getting along with teachers		Has trouble getting along with teachers	E	Antisocial
Is not as happy as other children	Is not as happy as other children			
Appears miserable, unhappy, tearful, or distressed	Appears miserable, unhappy, tearful, or distressed			
Has trouble enjoying him/her self	Has trouble enjoying him/her self			

Sources: For the BPI: Centre for Human Resource Research (2009). Child and Young Adult Data Users Guide. Ohio State University. For the NLSCY: National Longitudinal Survey of Children and Youth, Cycle 8. Microdata User Guide. Statistics Canada.

Table A.2: BCSAS items included in adult depression (CESD-R) scale

	BCSAS item
During the past week. . .	I had trouble keeping my mind on what I was doing
	I felt depressed
	I felt that everything I did was an effort
	My sleep was restless
	I felt sad
	I could not get going

Appendix B

Robustness calculations

This appendix reports additional details on the robustness calculations described in Section 4. These calculations are based on a simple bounding exercise in the spirit of Manski (1990). We abstract from the precise details of the model in order to focus on the problem of bounding average treatment effects when survey response depends on the treatment.

Suppose that a given school features a population of potential student respondents characterized by a binary treatment $t \in 0, 1$, bounded potential outcomes $y(t) \in [y^{min}, y^{max}]$ and participation choices $z(t) \in 0, 1$. The treatment is universal FDK[1] provision at the school in one's Kindergarten year. Depending on the application, non-participation here can mean either attending a different school or non participation in the survey. The key assumption in these calculations is that participation is monotonic in the treatment. That is, if a school's enrollment increases from 44 students to 46 students in response to the treatment, we assume that those 46 students can be divided into 44 "stayers" who would have attended the school either with or without the treatment and two "movers" who would not have attended the school without the treatment. Monotonicity here requires that either $z(1) \geq z(0)$ with probability one or $z(1) \geq z(0)$ with probability one.

We observe (t, z, yz), i.e. we only observe outcomes for participants. To keep things simple suppose that t is randomly assigned[2] (i.e., independent of both $y(.)$ and $z(.)$), and that we estimate a regression of y on t conditional on participating ($z = 1$). Then the regression coefficient β can be interpreted as measuring:

$$\beta = E(y|t = 1, z = 1) - E(y|t = 0, z = 1)$$

Expressed in terms of potential outcomes:

$$\beta = E(y(1)|t = 1, z(1) = 1) - E(y(0)|t = 0, z(0) = 1)$$

Since t is randomly assigned:

$$\beta = E(y(1)|z(1) = 1) - E(y(0)|z(0) = 1)$$

[1]Here we ignore the distinction between the treatment (student enrollment in FDK) and the instrument (universal FDK); this distinction can be accommodated by conditioning on being a complier.

[2]Again, the background characteristics and fixed effects have been left out here for simplicity, but could be accommodated by conditioning on them.

Rearranging produces:

$$\beta = E(y(1)|z(1)=1) - E(y(0)|z(1)=1) + E(y(0)|z(1)=1) - E(y(0)|z(0)=1)$$
$$= \underbrace{E(y(1)-y(0)|z(1)=1)}_{\text{true effect}} + \underbrace{E(y(0)|z(1)=1) - E(y(0)|z(0)=1)}_{\text{bias}}$$

That is, we can interpret the regression coefficient as a true effect (average treatment effect on those who participate when treated) plus a bias term that reflects the gap in average baseline outcomes between those who attend the school when universal FDK is available and those who attend when it is not:

$$bias = E(y(0)|z(1)=1) - E(y(0)|z(0)=1)$$

Applying the law of total probability:

$$\begin{aligned} bias = & E(y(0)|z(0)=1, z(1)=1)Pr(z(0)=1|z(1)=1) \\ & + E(y(0)|z(0)=0, z(1)=1)Pr(z(0)=0|z(1)=1) \\ & - E(y(0)|z(0)=1, z(1)=1)Pr(z(1)=1|z(0)=1) \\ & - E(y(0)|z(0)=1, z(1)=0)Pr(z(1)=0|z(0)=1) \end{aligned}$$

When the treatment increases the likelihood of appearing in the data, the monotonicity assumption implies that $Pr(z(1)=1|z(0)=1)=1$. Substituting in and rearranging produces:

$$bias = \underbrace{\left(E(y(0)|z(0)=0, z(1)=1) - E(y(0)|z(0)=1, z(1)=1)\right)}_{\text{mover-stayer outcome gap}} \times \underbrace{Pr(z(0)=0|z(1)=1)}_{\text{proportion movers}}$$

The first part of this equation is the gap in average untreated outcomes between those who only participate when treated (movers) and those who always participate (stayers) and can be bounded since the outcomes themselves are bounded. The second part of this equation is the fraction of participants who are movers, and can be estimated directly from the data. Together these results can be used either to place bounds on the true treatment effect:

$$\text{true effect} = \beta - (\text{mover-stayer outcome gap}) * (\text{proportion movers})$$

or to calculate the mover-stayer outcome gap required to explain the measured effect given some presumed value for the true effect:

$$\text{mover-stayer outcome gap} = \frac{\beta - \text{true effect}}{\text{proportion movers}}$$

In contrast, when the treatment decreases the likelihood of appearing in the data, the monotonicity assumption implies that $Pr(z(0)=1|z(1)=1)=1$. Substituting in and rearranging produces:

$$bias = \underbrace{E(y(0)|z(0)=1, z(1)=1) - E(y(0)|z(0)=1, z(1)=0)}_{\text{stayer-mover outcome gap}} \times \underbrace{Pr(z(1)=0|z(0)=1)}_{\text{proportion movers}}$$

and the remaining calculations carry through.

Appendix Table B.1 below describes the frequency distribution of externalizing behavior as well as its hyperactivity and peer relationship problems subscales and the index of parental depression. As can be seen from the table, these distributions are highly skewed. Most parents report none or only a few problems, so the lowest possible score (no problems) is only one standard deviation below the mean while the highest possible score is two to four standard deviations above the mean. That is, it is not possible for a child to have dramatically above-average behavior but it is possible to have dramatically below-average behavior.

Table B.1: Distribution of raw and standardized behavior scores, Kindergarten students in BCSAS.

Raw score (# of problems)	**Externalizing**		**Hyperactivity**		**Peer relationships**		**Parental depression**	
	Standardized Score	% ≤ score	Standardized score	% ≤ score	Standardized score	% ≤ score	Standardized score	% ≤ score
0	-1.11	18.6	-1.05	33	-0.46	78.9	-1.01	30.9
1	-0.84	30	-0.39	54.3	1.06	92.8	-0.47	51.8
2	-0.57	40.9	0.27	71.7	2.57	98	0.07	68.9
3	-0.3	51.9	0.93	85.3	4.09	100	0.61	80.1
4	-0.03	60.8	1.59	95.8			1.15	87.7
5	0.24	68.9	2.25	100			1.69	93.5
6	0.52	76					2.23	100
9	1.33	90.8						
11	1.87	95.3						
18	3.76	100						

Appendix Table B.2 uses this information to perform the robustness calculations described in Section 2.4.

Table B.2: Data and formulas used to calculate values reported in Section 4, Robustness checks.

Quantity	Baseline parameters	Calculated Values		Source
Increase in school enrollment				
Enrollment in representative school	44			Admin data, average school Kindergarten enrollment
Estimated effect of FDK on school enrollment	0.043			Admin data, Table 6
Implied increase in number of students enrolled		1.9		=.043*44
As a proportion of students in FDK school		0.041		=1.9/(44+1.9)
Estimated effect of FDK on externalizing behavior	0.052			BCSAS data, Table 2
Minimum (no problems) behavior score	-1.11			BCSAS data, Table B1
95th percentile of behavior score	1.87			BCSAS data, Table B1
Lower bound on true effect (movers are at 95th percentile)		-0.03		=0.052-(1.87-0)*0.041
Upper bound on true effect (movers have no behavior problems)		0.10		=0.052-(-1.11-0)*0.041
Increase in school enrollment, minority language students				
Minority language share of students	0.264			Admin data, Table 1
Minority language enrollment in representative school		11.6		=44*0.264
Estimate of FDK on school minority language share	0.043			Admin data, Table 7
Implied increase in minority language enrollment		2.5		=(.264+.043)*45.9-11.6
As a proportion of minority language students		0.18		=2.5/(11.6+2.5)
Effect of FDK on hyperactivity				
Point estimate	-0.209			BCSAS data, Table 4
Minimum score (no problems)	-1.05			BCSAS data, Table B1
High score (95.8 percentile)	1.59			BCSAS data, Table B1
Lower bound on true effect		-0.49		=-0.209-(1.59-0)*0.18
Upper bound on true effect		-0.02		=-0.209-(-1.05-0)*0.18
Average score for movers if true effect is zero		-1.19		=0-0.209/0.18
Effect of FDK on peer relationship problems				
Point estimate	-0.317			BCSAS data, Table 4
Minimum score (no problems)	-0.46			BCSAS data, Table B1
High score (92.8 percentile)	1.06			BCSAS data, Table B1
Lower bound on true effect		-0.50		=-0.317-(1.06-0)*0.18
Upper bound on true effect		-0.24		=-0.317-(-0.46-0)*0.18
Average score for movers if true effect is zero		-1.81		=0-0.317/0.18
Effect of FDK on parental depression				
Point estimate	-0.321			BCSAS data, Table 4
Minimum score (no problems)	-1.01			BCSAS data, Table B1
High score (93.5 percentile)	1.69			BCSAS data, Table B1
Lower bound on true effect		-0.62		=-0.321-(1.69-0)*0.18
Upper bound on true effect		-0.14		=-0.321-(-1.01-0)*0.18
Average score for movers if true effect is zero		-1.83		=0-0.321/0.18
Effect of FDK on parent hours > 0				
Point estimate	0.196			BCSAS data, Table 5
Average (minimum is zero, maximum is one)	0.761			BCSAS data, Table 1
Lower bound on true effect		0.15		=0.196-(1-0.761)*0.18
Upper bound on true effect		0.33		=0.196-(0-0.761)*0.18
Effect of FDK on parent hours ≥ 15				
Point estimate	0.221			BCSAS data, Table 5
Average (minimum is zero, maximum is one)	0.656			BCSAS data, Table 1
Lower bound on true effect		0.16		=0.221-(1-0.656)*0.18
Upper bound on true effect		0.34		=0.221-(0-0.656)*0.18
Effect of FDK on parent hours ≥ 35				
Point estimate	0.116			BCSAS data, Table 5
Average (minimum is zero, maximum is one)	0.376			BCSAS data, Table 1
Lower bound on true effect		0.01		=0.116-(1-0.376)*0.18
Upper bound on true effect		0.18		=0.116-(0-0.376)*0.18
Survey response rates				
Overall response rate	0.166			
% minority language students in sample	0.182			
Response rate of minority language students		0.114		=0.166*0.182/0.264
Number of minority students in sample, representative school		1.33		=0.114*11.6
Predicted number of minority language students in sample, constant survey response rate, FDK		1.61		=0.114*(11.6+2.5)
Share of minority language students in FDK "missing" from sample		0.18		=(1.61-1.33)/1.61
Effect of FDK on hyperactivity				
Lower bound on true effect			-0.39	=-0.209-(0-(-1.05))*0.18
Upper bound on true effect			0.07	=-0.209-(0-1.59)*0.18
Average score for movers if true effect is zero			1.19	=0-(-0.209/0.18)
Effect of FDK on peer relationship problems				
Lower bound on true effect			-0.40	=-0.317-(0-(-0.46))*0.18
Upper bound on true effect			-0.13	=-0.317-(0-1.06)*0.18
Average score for movers if true effect is zero			1.81	=0-(-0.317/0.18)
Effect of FDK on parental depression				
Lower bound on true effect			-0.50	=-0.321-(0-(-1.01))*0.18
Upper bound on true effect			-0.02	=-0.321-(0-1.69)*0.18
Average score for movers if true effect is zero			1.83	=0-(-0.321/0.18)
Effect of FDK on parent hours > 0				
Lower bound on true effect			0.06	=0.196-(0.761-0)*0.18
Upper bound on true effect			0.24	=0.196-(0.761-1)*0.18
Effect of FDK on parent hours ≥ 15				
Lower bound on true effect			0.11	=0.221-(0.656-0)*0.18
Upper bound on true effect			0.28	=0.221-(0.656-1)*0.18
Effect of FDK on parent hours ≥ 35				
Lower bound on true effect			0.05	=0.116-(0.376-0)*0.18
Upper bound on true effect			0.23	=0.116-(0.376-1)*0.18

Appendix C

Additional results

This appendix reports additional results that have been omitted from the main body of the paper in the interest of space, but are of potential interest.

Appendix Table C.1 reports the results from the "first stage" regression corresponding to our main results, i.e., a linear regression in which the dependent variable is FDK enrollment, and the explanatory variables include the instrument (universal FDK availability) as well as the control variables and fixed effects. Each FE-IV regression reported in the main body of the paper has a slightly different first stage as observations are dropped for missing values of the dependent variable, so Table C.1 reports the first stage for the full set of observations for which it could be estimated. Only a few observations are affected (e.g., the regression for externalizing behavior in Table 2.2 is based on 6403 of the 6431 observations used in Table C.1) and we have verified that the results reported here are representative.

Appendix Table C.2 reports the estimated effect of FDK on behavior problems in Grade 1, in the same format as Table 2.2 reports the estimated effect of FDK on behavior problems in Kindergarten.

Appendix Table C.3 reports the effects of FDK on parental outcomes among respondents who are mothers, in the same format as Table 2.5 reports the effects of FDK on parental outcomes among all respondents.

Appendix Table C.4 reports the effects of FDK on child behavior problems in Kindergarten under a number of alternative specifications:

- The "baseline specification" is just the specification used in Table 2.2, reproduced here for convenience.
- The "no controls" specification drops the individual-level control variables, while keeping the fixed effects.

- The "additional controls" specifications add further control variables to the model. The additional "school context" controls include enrollment in a multiple-grade class (both an indicator for the child's own participation, and the % of same-grade schoolmates in multiple-grade classes), average class size for current school and grade level, and % Aboriginal and % disabled in current school and grade level. The additional "family context" controls include indicators for speaking a Chinese language at home, speaking Punjabi at home, having younger siblings, family income in the lowest quartile, family income in the second-lowest quartile, whether the parent works for pay, and whether the parent works full-time.

- The "catchment school substituted for actual school" specification bases the instrument, fixed effects and clustering on the student's (English-language) catchment school rather than his or her actual school. Students can be matched to a catchment school based on their postal code of residence; approximately 87% of the sample can be matched. Non-matches can occur because the parent did not report a valid postal code, because the postal code cannot be matched to a specific location within a BCSAS-participating district, or because the district does not have catchment boundaries.

- The "behavioral outcomes measured by two-factor model" specification uses dependent variables that have been constructed by estimating a factor model. That is, a two-factor model was estimated for the full set of child behavior questions using the iterative principal factor method. A two-factor model was chosen primarily for comparability; two factors is also the number that would be selected using the conventional "eigenvalue > 1" criterion. The factors were then rotated using the oblique quartimin method, and factor scores constructed by the Bartlett scoring method. To facilitate comparison with the simple indices, the factor scores are standardized to have mean zero and unit variance within the grade and signed so that higher scores can be interpreted as reflecting more behavior problems. The first factor is reported as "externalizing" and the second factor is reported as "anxiety/depression" in accordance with the correlation between these factors and the two simple indices.

The results in Appendix Table C.4 indicate that the results reported in the main text are robust to these alternative specifications.

Appendix Table C.5 presents placebo estimates of the effects of universal FDK availability in a school on the characteristics of students who reside within that school's catchment area (rather than those who attend the school). These effects are estimated from the sub-sample of students in the administrative data for whom we could identify the catchment school. The results show that the introduction of universal FDK coincided with an increase in the residential concentration of minority language students in a school's catchment area. Comparing this point estimate to the corresponding estimate in the first column of Table 2.7 demonstrates that almost two-thirds (0.023/0.035) of the overall placebo effect of universal FDK on minority language status can be accounted for by changing patterns of residential choice, or of reported residential location.

Table C.1: "First stage" regression, BCSAS.

Explanatory variable	**FDK enrollment** (FDK_i)
Universal FDK availability ($UniversalFDK_{s(i)c(i)}$)	0.881***
	(0.022)
Male	-0.003
	(0.003)
Aboriginal identity	0.005
	(0.013)
Language spoken at home is not English	0.005
	(0.008)
Student's age in months	-0.0003
	(0.0005)
Parent's education (base category HS):	
HS dropout	-0.01
	(0.015)
Some Post-secondary	-0.013
	(0.008)
Bachelor or higher	-0.011
	(0.009)
Lone parent	0.006
	(0.007)
Respondent is not mother	-0.004
	(0.007)
Parent's age in years	-0.0002
	(0.0004)
Number of observations	6431
Number of schools	506

Dependent variable is FDK enrollment. All regressions include year and school fixed effects, and standard errors are clustered by school location.
(* = significant at 10%, ** = 5%, *** = 1%)

Table C.2: Effect of full-day Kindergarten on behavior problems in Grade 1.

Explanatory variable	Externalizing	Externalizing subscales				Anxiety/ Depression
		Anti-social	Hyper-active	Head-strong	Peer problems	
Full-day Kindergarten	-0.039	-0.042	-0.141	0.047	-0.030	0.051
	(0.075)	(0.074)	(0.078)	(0.077)	(0.077)	(0.077)
Male	0.247***	0.237***	0.367***	0.136***	0.088***	0.063***
	(0.023)	(0.022)	(0.023)	(0.024	(0.023)	(0.023)
Aboriginal identity	0.232***	0.205***	0.226***	0.148***	0.232***	0.147**
	(0.057)	(0.060)	(0.0057)	(0.055)	(0.060)	(0.058)
Language spoken at home is not English	-0.039	-0.006	-0.020	-0.052	0.062	-0.052
	(0.040)	(0.039)	(0.038)	(0.042)	(0.039)	(0.037)
Student's age in months	0.001	0.002	-0.001	-0.001	0.004	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Parent's education (base category HS):						
HS dropout	0.144*	0.103	0.181**	0.135	0.001	0.066
	(0.082)	(0.089)	(0.072)	(0.082)	(0.090)	(0.088)
Some Post-secondary	-0.035	-0.126***	-0.004	-0.036	-0.002	0.059
	(0.039)	(0.042)	(0.038)	(0.040)	(0.043)	(0.037)
Bachelor or higher	-0.098**	-0.166***	-0.079*	-0.095**	-0.005	0.058
	(0.041)	(0.042)	(0.041)	(0.042)	(0.042)	(0.039)
Lone parent	0.203***	0.216***	0.115**	0.159***	0.163***	0.202***
	(0.050)	(0.051)	(0.047)	(0.046)	(0.048)	(0.051)
Respondent is not mother	0.021	0.040	0.063	-0.014	0.044	0.028
	(0.040)	(0.039)	(0.039)	(0.039)	(0.041)	(0.037)
Parent's age in years	-0.007***	-0.003	-0.008***	-0.005**	-0.005**	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Robust 1st stage F-statistic	1211	1214	1212	1210	1209	1211
Number of obs	7749	7746	7746	7749	7745	7750
Number of schools	485	485	485	485	485	485

All regressions include year and school fixed effects, and standard errors are clustered by school location. Universal FDK availability is used as an instrument for FDK enrollment.
(* = significant at 10%, ** = 5%, *** = 1%)

Table C.3: Effect of full-day Kindergarten on maternal outcomes in Kindergarten and Grade 1, overall and by home language (BCSAS data).

Effect of interest	**Labour supply**				**Depression**
	Weekly hours	Hours > 0	Hours ≥ 15	Hours ≥ 35	
Effect of FDK on Kindergarten mothers:					
FDK	1.624	0.013	0.079**	-0.002	-0.004
	(1.097)	(0.032)	(0.032)	(0.032)	(0.064)
Effect of FDK on Grade 1 mothers:					
FDK	-0.233	-0.01	-0.014	-0.026	0.01
	(1.317)	(0.035)	(0.037)	(0.039)	(0.085)
Effect of FDK on Kindergarten mothers, by home language:					
FDK x (English)	1.049	-0.007	0.062*	-0.012	0.031
	(1.107)	(0.032)	(0.032)	(0.032)	(0.064)
FDK x (not English)	7.669***	0.219***	0.258***	0.099	-0.349***
	(2.748)	(0.077)	(0.083)	(0.069)	(0.144)
Difference	6.621**	0.227***	0.196**	0.110*	-0.380***
(not English–English)	(2.644)	(0.073)	(0.078)	(0.066)	(0.141)
Effect of FDK on Grade 1 mothers, by home language:					
FDK x (English)	-0.279	-0.018	-0.018	-0.02	0.012
	(1.264)	(0.034)	(0.035)	(0.038)	(0.081)
FDK x (not English)	0.212	0.065	0.023	-0.087	-0.006
	(2.825)	(0.074)	(0.08)	(0.078)	(0.175)
Difference	0.491	0.082	0.041	-0.067	-0.017
(not English–English)	(2.337)	(0.06)	(0.066)	(0.065)	(0.135)

All regressions include year and school fixed effects, and standard errors are clustered by school location. Universal FDK availability is used as an instrument for FDK enrollment. Control variables include gender, Aboriginal identity, language spoken at home, student's age, parent's age, parent's education, lone parent, respondent is not mother.
(* = significant at 10%, ** = 5%, *** = 1%)

Table C.4: Effect of full-day Kindergarten on behavior problems in Kindergarten (BCSAS data), alternative specifications.

Specification	**Externalizing**	**Anxiety/ Depression**
Baseline specification from Table 2	0.052	0.082
	(0.06)	(0.064)
No control variables	0.039	0.074
	(0.06)	(0.064)
Additional control variables, school context	0.035	0.076
	(0.063)	(0.066)
Additional control variables, family context	0.028	0.034
	(0.066)	(0.069)
Catchment school substituted for actual school	0.044	0.164
	(0.099)	(0.105)
Behavioral measures constructed using two-factor model	0.061	0.032
	(0.062)	(0.067)

All regressions include year and school fixed effects, and standard errors are clustered by school location. Universal FDK availability is used as an instrument for FDK enrollment. Control variables in baseline specification include gender, Aboriginal identity, language spoken at home, student's age, parent's age, parent's education, lone parent, respondent is not mother.
(* = significant at 10%, ** = 5%, *** = 1%)

Table C.5: Placebo regressions, "effect" of universal FDK program availability in catchment school on pre-treatment variables for Kindergarten students (administrative and BCSAS data).

Effect of interest	**Administrative data**
	Catchment school, three cohorts
Placebo effect of universal FDK program availability on:	
Male	0.004
	(0.007)
Aboriginal identity	0.001
	(0.004)
Speaks a language other than English at home	0.023***
	(0.008)
Student's age in months	0.006
	(0.063)

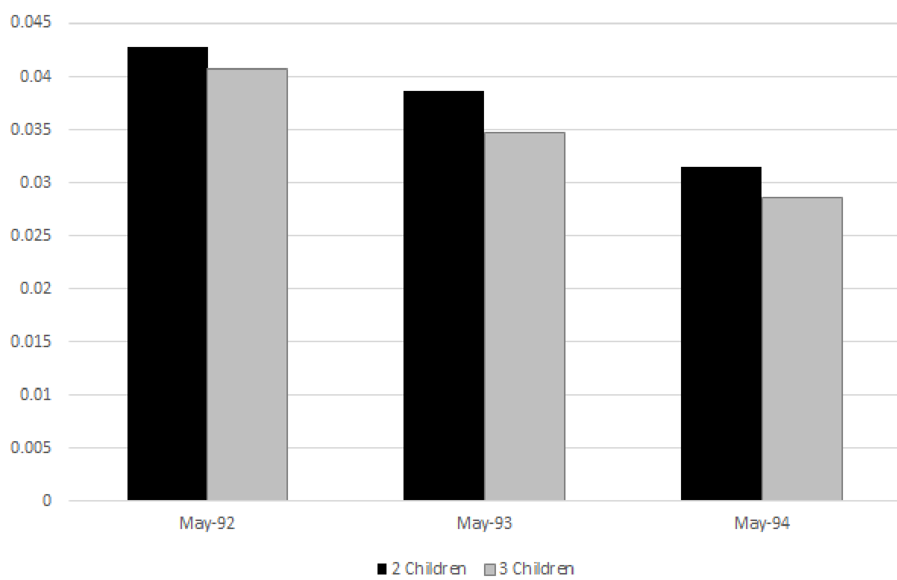
All regressions include year and school fixed effects, and standard errors are clustered by school location. No additional control variables.
(* = significant at 10%, ** = 5%, *** = 1%)

Appendix D

Heterogeneous effects

As there is a one year gap between the passage of the PCL and its implementation, one may expect an increase in the number of higher order births (4 and more) in this "Transition Period". Parents with more than three children as of May 1993, who desired even a greater number of children would have seen the Transition Period as the time to increase their family size without suffering the penalties 1993 PCL would have imposed on them otherwise. Figure D.1, shows the probabilities of having another child for people with two and three children exactly one year before passage of PCL, in May 1993 when the law was passed in parliament and in May 1994, exactly one year after. Our regression results also do not show a significant difference in the gap between the two probabilities during this period.

Figure D.1: Heterogeneous effects of 1993 PCL in transition period



Source: Authors' calculation based on publicly available 2% sample of 2006 Iranian Census at www.amar.org.ir.
Notes: Average percentage of families with two or three children in May 1992, May 1993 and May 1994 who gave birth in the following nine to twelve months.