The Long-Run Effect of Public Universal Pre-K*

Assaf Kott
Brown University

Abstract

I estimate the effect of attending public pre-K at age three on matriculating high school at age eighteen. My setting is municipalities with an Arab majority in Israel that experienced a universal public pre-K expansion in 2000-2004. Leveraging the staggered rollout of the expansion for identification, I find that attending pre-K positively affects performance in high-school matriculation exams in the Hebrew and Arabic languages. A marginal treatment effect (MTE) analysis reveals a reverse selection on gains: children who benefit the most from pre-K are the least likely to enroll.

*I thank Anna Aizer, Jesse Shapiro, Nicola Babarcich, Chien-Tzu Cheng, and Marcella Mello for their helpful comments. Special thanks goes to Israel's Ministry of Education for allowing access to the online protected research lab and to Eliad Trefler for assistance with the data. All mistakes are my own.

1 Introduction

Universal public pre-K programs have gained popularity among the general public, politicians and policy experts over the past few years. In the 2020 democratic primary, for example, all leading candidates supported federal funded universal programs,[1] with Joe Biden's plan suggesting that offering pre-K for all three- and four-year-olds "will ease the burden on our families, help close the achievement gap, (and) promote the labor participation of parents who want to work."[2] In addition, within the U.S. several states and cities have begun to implement their own plans toward universal public pre-K. Oklahoma and Georgia, for instance, have been providing free public pre-K since the late 1990s (Cascio and Schanzenbach 2013) and New York City has been rapidly expanding its own public pre-K program since 2014 to allow all four-year-old residents to enroll (Brown 2019).

While there is a broad consensus on the effectiveness of high-quality targeted early childhood education, there is less evidence on large-scale universal public programs.[3] In Head Start, the most studied and large-scale targeted preschool program in the U.S, the intervention is aimed at a poor population and includes services, such as instruction for parents and nutrition and health guidance. Public pre-K programs, on the other hand, have a much narrower scope which in most places includes five days a week half-day educational programs that aim to prepare children for kindergarten and primary school. This raises the question of how much evidence from the high-quality targeted programs can be extended to universal pre-K programs.[4]

In this study, I explore the long-term effects of a large-scale reform in Israel that significantly increased public pre-K enrollment in the Arab population. The quality of the program I study is of a lower quality compared to that of European programs: classrooms are bigger, adult to children ratio is lower, requirements for teachers are more lenient and expenditure per pupil is lower. This reform was first studied by Schlosser (2011) who found that it increased maternal labor force participation, with the increase being driven by more educated mothers.

[1] See for example an analysis by the Washington post in https://www.washingtonpost.com/graphics/politics/policy-2020/education/ (accessed 5/12/2020)

[2] https://joebiden.com/education/ (accessed 5/12/2020)

[3] For Perry preschool and the Abecedarian project studies have found that for each dollar the government invested it recouped between four to seven dollars in taxes (see Currie and Almond (2011) for a review). The returns on Head Start, which has been more extensively studied, have been less consistent with some studies finding the program does not pay off. However, Kline and Walters (2016) show that accounting for the fact that the programs draw participants from other public programs, the returns are positive even under the strictest assumptions.

[4] For further discussion on the problems of extending evidence from targeted programs to universal pre-K program see Baker (2011), who also points out that policy makers who advocate for universal pre-K often cite evidence from these targeted programs.

To conduct my analysis, I use administrative data from the Israeli education ministry. I restrict my analysis to 28 municipalities with Arab majorities that experienced a sharp increase in enrollment in public pre-K at the age of three during the school years 1999-2000 to 2003-2004. My main analysis sample covers 54,751 students who were born between 1996 and 2000.

I address endogenous selection to pre-K by employing an instrumental variable approach. The instrument, pre-K coverage rate, makes use of the fact that logistical issues have prevented a full and immediate implementation of the reform, and thus have induced a temporal and geographical variation in the roll-out of pre-K classrooms. The analysis is conducted in two parts. In the first part, I examine standard two stage least squares (2SLS) estimators which identify local average treatment effects (LATE). In the second part, I adopt a marginal treatment effect (MTE) framework. This method, introduced by by Björklund and Moffitt (1987) and generalized by Heckman and Vytlacil (1999, 2005) and Heckman et al. (2006), allows the econometrician to draw a more complete picture of heterogeneity in the treatment effect and provides estimates for the average treatment effect (ATE), treatment on the treated effect (TT), and treatment on the untreated effect (TUT). These estimates allow me to draw conclusions about the nature of selection on gains and prevalence of heterogeneous treatment effects.

I find that attending pre-K at the age of three has a positive effect on performance in high-school graduation exams in the Hebrew and Arabic languages and on the probability of graduating with a high-school diploma. I do not find an effect on the two other academic subjects I study (English and Math). The pooled sample estimates mask heterogeneity: while the estimated LATE on Hebrew test scores is similar across genders - about a fifth of a standard deviation and significant at the 1% level - the effects on graduation with a diploma and on Arabic test scores are only significant for girls (at the 10% and 5% levels, respectively) and are indistinguishable from zero for boys. There is less heterogeneity across parental education. The MTE analysis reveals a reverse selection on gains. Taken together, my findings support the case for public universal universal pre-K programs. However, these positive returns are highly dependent on reaching an almost full take-up.

This study contributes to the growing literature on public universal pre-K in two ways. First, I add to the limited literature on universal pre-K long-term outcomes. Most existing studies focus on short-term outcomes (e.g. effects on school-readiness exams) or on medium-term outcomes (e.g. effects on primary and middle school standardized test scores).[5] The few papers that study long-run

[5] Gormley and Gayer (2005) find evidence that disadvantaged preschool attendees in Tulsa, Oklahoma, scored higher than their peers on tests at the end of the preschool year. Fitzpatrick (2008) finds that the Georgia program increased test scores of disadvantaged children as late as the fourth grade, but the effects appear smaller and concen-

outcomes reach contradicting conclusions. Baker et al. (2008, 2019) investigate the consequences of an extensive reform in the 1990s in the Canadian province of Quebec in which a generous subsidy became available for public child care for children ages zero to five. They find that in the long run, this program has increased participation in crime and reduced life satisfaction. They do not find any lasting effect on cognitive skills. On the other hand, Havnes and Mogstad (2011) study a major childcare reform that took place in the 1970s in Norway and find a positive effect on long-term outcomes. They show that the reform significantly increased years of education, college attendance, and earnings at the age of 30, and reduced the probability of having children. The contradictory results of these two studies may be due to the varied quality of programs and alternatives. In the Canadian case, the reform affected mostly children from more affluent families that substitute away from private care, and the policy positively affected mothers' labor force participation (LFP). In the Norwegian case, participants in the public program substitute away from informal care, and the program did not increase mothers' LFP. In addition, the Norwegian model seems to be an outlier in its high-quality pre-K program.[6] Finally, Gray-Lobe et al. (2021) study the long-term effects of public preschool in Boston. While this program was open to all students, capacity was limited and thus assignment was done via lottery. The authors find that attending pre-K boosts college-going as well as high school graduation achievements with the effects being driven by boys. I add to this literature by exploring a universal program in which participants switched from informal care to public care, which affected a lower SES population with a low maternal LFP and which, in several dimensions, is of inferior quality.

My second contribution is to the understanding of heterogeneous treatment effects and selection on gains in the public pre-K arena. I follow Cornelissen et al. (2018) in adopting an MTE framework to estimate the ATE, TT, and TUT and to explore heterogeneous treatment effects on observables and unobservables. Cornelissen et al. (2018) studied an expansion of universal pre-K in Germany during the 1990s and used the gradual roll-out of new classrooms as an instrument in an MTE framework. Their study examines the effect on short-term outcomes that are measured at the end of kindergarten. They find a reverse selection on gains: the TUT estimate for an indicator of school readiness is large and significant, while the TT is negative. Another related study is Kline

trated in a small group of disadvantaged children living in small towns. Cascio and Schanzenbach (2013) combines Georgia's and Oklahoma's programs to study the reduce-form effect of preschool on family and children outcomes. The longest term outcomes they study are test scores in the eighth grade, and they find a positive effect on tests in reading and math.

[6] Specifically, educational requirements for teachers, adult to children ratio, and expenditure per student in Norway dominate over those in other developed countries (see Havnes and Mogstad 2011)

and Walters (2016). They find that accounting for the fact that Head Start draws many students from other public programs, makes the program cost-effective under a wide range of assumptions. They also find that Head Start generates the highest returns for children who are the least likely to attend the program. I add to this literature by exploring selection and long-term heterogeneity patterns in public universal pre-K. My findings of reverse selection on gains are consistent with those of Cornelissen et al. (2018) and Kline and Walters (2016), and should serve as a warning that high returns on these types of early education investments are contingent on full take-up.

2 Background

2.1 The Policy

Like most developed countries, Israel has a compulsory education law and it provides state funded public free schooling for kindergarten through the 12th grade. This law is colloquially termed the Free and Compulsory Education law. In 1984 the *Knesset* (the Israeli parliament) amended the law and extended it to pre-K - children ages three and four. The implementation of the amendment, however, has been suspended by a temporary legislation for many years due to budgetary constraints.

During the time the amendment was suspended, local municipalities could choose to open pre-K classrooms. These local programs were regulated by the central government, which determined tuition, requirements for teachers, and subsidies among other regulations. Subsidies were available based on household income, and were funded by both the central government and the local municipalities while the rules regarding subsidies were set at the national level. This created an incentive for municipalities with a substantial share of low income households to opt out of opening pre-K programs.

The first step for the implementation of the amendment took place in 1999 when the minister of education issued a special order to implement the law in specific neighborhoods and municipalities ("The 1999 Reform"). This was supposed to be the first phase in a national effort to fully implement the amendment within ten years, but this was ultimately the only step taken.[7] There were three criteria for places to be included in the 1999 Reform. These consisted of municipalities which were either ranked at the two bottom deciles of an official economics deprivation index,

[7]Only in 2011 after a series of large protests against the high costs of living in Israel, the government resumed implementing the amendment and in 2016 the roll out was completed.

places declared as confrontation line, and places defined as national priority areas.

The 1999 Reform induced two main changes. The first change is due to the free component of the amendment to the Free and Compulsory Education law. That is, in places where pre-K programs were already in place and enrollment was high, the policy effect was in eliminating tuition fees for parents. However, since municipalities which were affected were mostly the poorest municipalities in the nation, most parents in these places were already eligible for a significant subsidy; the highest subsidy level was 90% for the poorest families. The second change was due to the compulsory component of the law which determined that every child is entitled to public pre-K.[8] Thus, municipalities that previously opted out of providing pre-K or that only offered a limited number of pre-K programs, were now compelled to provide this service to all who required them. It is important to note that while the free component immediately applied in participating municipalities, the compulsory component had a more gradual effect. I explain more below on the circumstances which prevented Arab-majority municipalities, the population I study, from immediately providing pre-K services. But first, I explain why this reform has mostly affected municipalities with Arab majorities.

The fact that prior to the reform, coverage of pre-K classrooms was determined by the local municipality created a geographical variation in pre-K coverage in the pre-reform period. The variation was driven by differences in the economic stability of local municipalities and differences in their demand for pre-K services. These two forces created a division in pre-K coverage rate along ethnicity lines: municipalities with a Jewish non-orthodox majority with a high LFP rate and with a financially healthy budget were offering pre-K programs since the 1980s; on the other hand, municipalities with a non-Jewish majority, mostly places with an Arab majority, with a low LFP rate and with a tighter budget had limited pre-K coverage, if any. Therefore, the 1999 Reform mainly affected Israel's poorest municipalities with Arab majorities.

The data on pre-K enrollment confirm the idea of differences in enrollment by sector and age group. Figure 1 shows attendance rates for the 1996 birth cohort (three years old during school year 1999-2000) through the 2002 birth cohort (three years old during school year 2005-2006).[9] The figure shows attendance rates for pre-K at age three, pre-K at age four, and kindergarten at age five, separately for the Jewish and Arab sectors. For kindergartens, enrollment in both sectors

[8] It also compelled parents to enroll their children in pre-K, however this is rarely enforced.

[9] Conveniently, in Israel there is an almost complete overlap between a birth cohort and the age cutoff for qualifying for pre-K. For instance, children born between 12/4/1995 and 12/23/1996 were eligible to enroll age 3 pre-K in school year 1999-2000. The cutoff in Israel changes from year to year because it is defined as the first day of the Hebrew month of *Shvat*. For more information on cutoff rules in Israel see Attar and Cohen-Zada (2018).

was stable and very high (90 to 97%). This was a result of the Free and Compulsory Education law which began at age five before the amendment. For the Jewish sector, enrollment in pre-K at age three and pre-K at age four was stable and high: more than half of the students in every cohort attended public pre-K. The amendment did not seem to affect enrollment in the Jewish sector. This is not surprising since municipalities with Jewish majorities were already operating pre-K programs before the amendment. The patterns for ages three and four in the Arab sector are quite different. In this sector the reform seems to have had a substantial effect on enrollment. For pre-K at age three, enrollment had almost doubled during the sample period. A similar pattern is also evident for pre-K at age four. To summarize, the 1999 Reform significantly increased attendance of three year-olds from the Arab population in public pre-K.

2.2 Exogeneity of the roll-out of new classrooms

My estimation strategy relies on the idea that the temporal and geographical variation in implementing the 1999 Reform were driven by external forces, specifically logistical constraints, that were independent of unobserved determinants of children's educational achievements.

Knesset records from a discussion held in March 2001 reflect lawmakers' frustration with the slow opening of new classrooms. The transcripts indicate that in Arab municipalities included in the reform, the main constraints for opening new classrooms were: 1. lack of transparent communication between the central government and municipalities in regards to funding and the number of children in each cohort, 2. logistical issues in finding appropriate classrooms, and 3. the resignation of the minister of education and political instability at the national level which made it difficult to allocate funds and to make plans for the next school year. Tamar Gojansky, one of the key lawmakers who were pushing for this reform, summarized the feeling of participants in this discussion:[10]

> ... We failed to implement the 1999 decision and it has been dragging on through 2000, 2001, and 2002. We need to add probably 13,000 more children and you (the finance ministry) do not know exactly how many and exactly where they are located, and you cannot say yet where and how many classrooms will be open...we are stepping in place...

[10] https://www.nevo.co.il/law_html/law103/chinuch2001-02-30.htm (accessed on 05/13/2020). Author's translation from Hebrew.

These records suggest that: 1. policy makers had no intention to gradually open classrooms and that the order in which classrooms were opened within the 1999 Reform municipalities was not premeditated, 2. 1999 Reform municipalities were treated as one cohesive group, and 3. that the policy goal was to achieve sufficient pre-K coverage to allow enrollment of entire cohorts in Arab-majority municipalities. Indeed, an examination of enrollment by town reveals a subgroup of Arab-majority towns achieved full enrollment during school years 1999-2000 to 2003-2004.[11] This motivates my sample selection to Arab-majoriy municipalities (on average more than 90% of students) in the 1999 Reform municipalities. This sample restriction leaves an average of 10,950 students per cohort out of a total of an average of 31,200 Arab students per cohort (who consist 25% of the total student body of an average pre-K cohort in the Israeli formal education system during my sample period).

The variation in pre-K enrollment in sample municipalities is illustrated in figure 2. The figure shows the distribution of public pre-K enrollment rates in the 28 sample municipalities for birth cohorts 1996 through 2000 (school years 1999-2000 to 2003-2004). The figure reveals that while some towns were quick to reach the goal of almost full enrollment, other towns took a slower and more gradual path toward full enrollment. I will exploit this variation in timing and intensity across municipalities and cohorts to identify the causal effect of attending pre-K at age three on long-term outcomes.

2.2.1 The Instrument

The instrument I will use throughout the paper is pre-K coverage rate. I follow Cornelissen et al. (2018) and define the instrument as :

$$Coverage_{jc} = \frac{30 \cdot Classrooms_{jc}}{pop_{jc=1996}}$$

Where $Classrooms_{jc}$ is the number of classrooms at municipality j when cohort c was three-years-old, and pop_{j1996} is the size of the 1996 birth cohort at municipality j. The numerator of this term can be thought of as the number of pre-K spots if each classroom consisted of exactly 30 students. Thus, the coverage rate equals one if there are exactly enough spots to cover all children

[11] I choose 2003-2004 as the end period for several reasons. First, figure 1 indicates that by this school year most of the increase in enrollment in the Arab sector had occurred. Outside sources also support the notion that this process ended in 2003-2004 (Blass and Bleikh (2012)). Another reason to choose this end period is that the cohort that was three years old during 2003-2004 - birth cohort 2000 - is the youngest cohort for which I have graduation data.]

of birth cohort 1996. In practice, most classrooms are of mixed ages and enroll four-year-olds, and even five-year-olds in some cases, and therefore the coverage rate could be higher than one without full enrollment of three-year-olds. I keep fixed the denominator at the size of the 1996 cohort to use only variation that is induced by the reform and not by changes in cohort size.

Figure 3 shows the evolution of pre-K coverage rate over time. The figure reveals that the distribution of coverage rate shifted to the right throughout the sample period. The figure also shows that there is a variation in the cross-section of coverage rate. Both of these patterns are consistent with the patterns in enrollment rate presented in figure 2. Indeed, a regression of enrollment rate on coverage rate controlling for municipality and cohort fixed effects yields a strong first stage. Column 1 in Table 1 shows that the coefficient in this regression is a statistically significant 0.387. We expect this coefficient to be about one, if enrollment was solely driven by available spots and if classrooms only consisted of three-year-olds. Since classes also enrolled older children and since older children had precedence, we expect this number to be lower than one. I provide additional evidence on the first-stage relationship in section 5.

After establishing the first stage relationship, I now turn to exogeneity of pre-K coverage rate. For the instrument to be valid it must be orthogonal to unobserved determinants of educational outcomes and to selection into pre-K. While, by definition, I cannot empirically test this assumption, there is some suggestive evidence that supports this idea. First, as I discussed, the order in which classrooms were opened was not planned. The goal of policy makers was to achieve full enrollment in participating municipalities as soon as possible. Yet, it might be the case that there were factors which were unobserved by the policy makers and the econometrician that determined the gradual roll-out of classrooms. My analysis will be confounded if these factors are also unobserved factors in determining educational outcomes and selection into pre-K.

Next, I show that deviations from year and municipality averages of important observed determinants of educational outcomes cannot be explained with pre-K coverage rate. Column 2 (column 3) of table 1 shows results from a municipality-level regression of the mean mother (father) years of education on pre-K coverage rate and school year and municipality fixed effects. For both dependent variables I cannot rule out that the coefficient on pre-K coverage rate is zero and can rule out a value larger than 0.3 years of schooling.[12] To sum up, changes in pre-K coverage are mainly determinant by initial level of coverage and do not appear to have explanatory power in predicting

[12]In the sample, an increase of father's (mother's) education by one year of schooling is associated with an increase of 0.041 (0.044) percentage points in the probability of obtaining a high school diploma (controlling for town and cohort fixed effects).

other important covariates.

2.3 Child Care Provision in Israel and The Arab Sector

To allow comparison of my study to other policies around the world, I briefly review the main elements of public pre-K in Israel and provide some additional information on the Arab sector. In Israel, public pre-K programs are half-day and run six days a week for nine months. In the periods and places where the amendment was not implemented, an annual fee for attending public pre-K was charged. The annual tuition sticker price was set at the national level and was usually slightly lower than the average monthly wage in Israel. A subsidy schedule was set also at the national level and was based on household size and income. On average parents paid 64% of the full tuition.[13]

The Israeli education ministry mandates that each class have a teacher and an assistant teacher. The teacher is required to have a teaching certificate and in many cases also holds a bachelor degree, however this was not a requirement during the period I study. The teaching assistant has to have graduated from high-school and commit to undertaking a pre-K related coursework during their first two years on the job. There is some evidence that these requirements were not strictly enforced in the Arab sector during the period I study (Zaban 2012). The education ministry also set rules for the maximum classroom size which at the time of my study was 35 students, and the average class size in my sample was 32. This implies an adult to children ratio in the Israeli program, 1:16, which is higher than most peer pre-K programs.[14] While I do not have data on expenditure per student in the period I study, in 2015 the estimated expenditure per student was $4,300 (PPP), which is much lower than most European countries and Head Start.[15]

Next, I provide more information on the Arab sector in Israel. Arabs consist about 20% of Israeli citizens and numbered 1.3 million people at the end of 2004. Arabs in Israel belong to three major religious groups: 83% are Muslims, 9% are Christians, and 8% are Druze (CBS 2005). The majority of the Arab population, 71%, live in towns or villages with a majority (90%) of Arab residents. The rest of the Arab population resides in mixed towns and cities.[16] There are substantial

[13]For example in the year 2001 full tuition was 5,323 NIS, and average salary was 5,500 NIS. source: https://www.gov.il/BlobFolder/dynamiccollectorresultitem/ministry_of_education_2001/he/state-budget_2001_gov-ministries-budget_educ.doc (accessed 05/14/202)

[14]Norway, Germany, UK, and France have an adult to children ratio of 1:8, 12.5:1, 10:1, 25:1, respectively; The ratio in Head Start is 10:1 (Cornelissen et al. 2018 and Havnes and Mogstad 2011)

[15]In the same year the estimated expenditure per student in Norway, and Germany was 11,000$ and 7,700$, respectively. Source: https://www.oecd.org/els/soc/PF3_1_Public_spending_on_childcare_and_early_education.pdf (accessed 05/14/2020)

[16]Author's calcutions of Israel's Central Bureau of Statistics (CBS) 1999 Local Authorities Sruvey. source:

economic disparities between the Arab and Jewish populations in Israel. In 2004, income of Arab households was 36% lower than that of Jewish households (CBS 2004a). Part of this gap is driven by the lower participation of Arab women in the labor force which in 2004 was 21% for women ages 25 - 64, compared to 69% for Jewish non-orthodox women.[17] In the context of Pre-K, the low labor force participation of women contributed to the high number of children ages 0-6 who were in home-care. In a survey conducted by the Israeli Central Bureau of Statistics (CBS) in 2004, 80% of the children not enrolled into public pre-K stayed at home with the mother or another relative. The other 20% are attending some other informal care (CBS 2004b). Additionally, data available only for school year 2004-2005 indicate that only 3.1% of Arab children ages 0 to 6 attended private programs (the corresponding rate in the Jewish sector is 9.5%).[18]

2.4 Secondary Schooling In Israel

I examine the effect of attending pre-K at age three on outcomes measured in high school. In Israel, high school students are examined in a series of matriculation exams, where each exam is associated with a number of credits (1 to 5) and to be awarded a high school diploma, students must receive a passing grade in exams worth at least 20 credits. Some subjects have a minimum credit requirement and all students have to take at least the lowest level exam in these subjects. Each student and school are associated with one of four educational tracks and the curricular of matriculation subjects and minimum requirements is set at the track level. I will focus on the Arab track where, in addition to the "regular" core subjects such as Mathematics and English, students are required to be examined at at least three credits level of Arabic (for native Arabic speakers) and Hebrew (for non-native Hebrew speakers).

The main outcome I examine is graduating high school with a diploma. This is an important outcome which is a necessary requirement for enrolling in post-secondary studies and in many entry-level positions which do not require a college degree. Importantly, previous research shows that interventions in Israel that successfully improved probability of getting a high school diploma also improved labor market outcomes in adulthood (see for example Lavy et al. 2020).

Another important outcome is tests in Hebrew. A lack of sufficient proficiency in Hebrew among the Arab population is hypothesized by many to be a driver of the economic disparities

---

https://old.cbs.gov.il/webpub/pub/text_page.html?publ=58&CYear=1999&CMonth=1 (accessed 05/14/2020)

[17] source: https://www.boi.org.il/he/NewsAndPublications/PressReleases/Documents/%D7%9E%D7%A6%D7%92%D7%AA%20%
(accessed 05/14/2020)

[18] source: https://old.cbs.gov.il/www/publications/children05/pdf/intro_h.pdf (accessed 05/14/2020)

between Jews and Arabs in Israel (see for example: Yashiv and Kasir 2014). The Israeli government has been investing in programs for students and adults that aim to improve Hebrew skills.[19] Therefore, if public pre-K can help improve Hebrew skills in adulthood, it is likely to also have a positive effects on labor market outcomes.

3 Data

The data used in this study come from the Israeli Education Ministry administrative records. The files identify the student's schools and pre-K institutions and record student's demographic characteristics including gender, ethnicity, immigration status, and parental schooling. These files are available from school year 1995-1996 to 2017-2018 for first through twelfth grade students and are available from school year 1999-2000 to 2017-2018 for public pre-K and kindergarten participants. In addition, the ministry keeps record of student achievements in high-school matriculation tests.

I construct my sample by first creating a census of all children born in birth cohorts 1996 to 2000. I do this by looking at all first graders who reside in my sample municipalities. Since attending elementary school is mandatory and since there is almost no private schooling in the sample municipalities, I assume this is the near-universe of children of these birth cohorts. Next, I link students to their past selves and record whether they attended pre-K when they were three-year-olds. I also record the number of classrooms that operated in their municipality at that time to create my instrument. Finally, I observev students' high school outcomes that include graduating with a diploma, total number of credits attempted, credits attempted in English, Math, Arabic, and Hebrew, and test scores in baseline-level Arabic and Hebrew exams.[20] For test scores I compute z-scores which are calculated for each cohort separately over all test takers (regardless of residing in sample municipalities). Table 2 shows summary statistics of the analysis sample.

[19]see an overview of programs in this link (accessed 05/14/2020)

[20]In Hebrew and Arabic all student in a given cohort are examined in the same baseline-level exam. Those who choose to take higher level in these subjects are tested in additional exams. In English and Math this is not the case and there is no signal exam everyone takes. Thus, I look only on test scores in Hebrew and Arabic.

4 Empirical Framework

4.1 Two Stage Least Square

I am interested in the effect of attending pre-K at age three on long run outcomes. I posit the following relationship between a long-run outcome and pre-K enrollment:

$$y_{ijc} = X_{ijc}\gamma + Enroll3_{ijc} \cdot \tilde{X}_{ijc}\beta + \alpha_j + \theta_c + \varepsilon_{ijc} \tag{1}$$

Where y_{ijc} is the outcome of student i, of birth cohort c, who resides in municipality j, X_{ijc} is a vector which non-parametrically controls for students background characteristics,[21] $Enroll3_{ijc}$ is an indicator for public pre-K enrollment at the age of three, $\tilde{X}_{ijc}$ is a subset of variables in X_{ijc} on which I allow heterogeneous treatment effects, α_j are municipality fixed effect, θ_c are cohort fixed effects, and ε_{ijc} is the error term. I address endogeneity concerns by restricting my sample to only 1999 Reform municipalities and by instrumenting enrollment with pre-K coverage rate. The first-step of the model is:

$$Enroll3_{ijc} = X_{ijc}\gamma^0 + Coverage_{jc} \cdot \tilde{X}_{ijc}\beta^o + Coverage^2_{jc} \cdot \tilde{X}_{ijc}\gamma^0 + \alpha^0_j + \theta^0_c + \varepsilon^0_{ijc} \tag{2}$$

Where $Coverage_{jc}$ is pre-K coverage in town j when student i was three years old.

As I discuss in section 2.2.1, initial pre-K coverage rate plays an important role in explaining changes in pre-K coverage overtime. Thus, controlling for municipality fixed effects, which capture time-invariant differences across municipalities (such as initial pre-K coverage), is important for the validity of the instrument. The role of the other controls is less in the reduction of bias, since as I showed in table 1 they are not correlated with the instrument, but rather in improving precision and allowing to identify heterogeneous treatment effects. I use quadratic function of coverage rate to allow a non-linear relationship between coverage and enrollment.

I now discuss the assumptions under which β can be given a causal interpretation. First, there must exist a first-stage such that the instrument induces sufficient variation in the endogenous variable. As discussed above, column 1 in table 1 presents evidence of a first-stage relationship. The second assumption requires that the instrument should be as good as random conditional on my set of controls. I discussed the exogeneity of pre-K coverage rate within the sample municipalities in section 2.2.1.

[21] I control for gender, family size, order of birth, month of birth, and five groups of parental years of education.

Finally, in the presence of heterogeneous treatment effects, an additional assumption is required for a LATE interpretation - monotonicity. In the context of my model I assume that for all students an increase in coverage rate does not reduce the probability of enrolling in pre-K. Again, I cannot support this assumption with my data, but it is hard to imagine a scenario in which a parent decides to enroll their child into pre-K when there is a low number of classrooms in the municipality but choose not to when this number increases.[22] One possible case can be when expansions in coverage rate are coupled with a reduction in pre-K quality and thus some parents may choose not to enroll their child. This is unlikely to be the case in my sample since there were almost no private pre-K programs in the Arab sector and therefore the alternative was either of lower quality (unregulated programs) or costly (home-care).

Under these assumptions the 2SLS estimand can be interpreted as the weighted average of causal effects of attending pre-K at age three, for the sub-group of students whose enrollment was affected by the instrument (the compliers). In the presence of a continuous variable this LATE can be hard to interpret. Cornelissen et al. (2016) show that this LATE can be written as a weighted average of all pairwise LATE estimates.[23] The weights for each pairwise LATE, Cornelissen et al. (2016) showed, are proportional to the strength of the first-stage and number of observations. Therefore, in the presence of substantial heterogeneity the 2SLS estimand might have little relevance for policy. With this limitation in mind, I next turn to an MTE framework to estimate other interesting treatment parameters.

4.2 Marginal Treatment Effect

4.2.1 Model

I now present the MTE framework. This section builds on Heckman and Vytlacil(1999, 2005), and Cornelissen et al. (2018). The purpose of this section is to ease the exposition and does not show new theoretical results. I start by rewriting equation 1 in a potential outcome framework:

$$y_{ijc}^{k} = X_{ijc}\beta^{k} + \alpha_j + \theta_c + U_{ijc}^{k}\; k = 0,1 \tag{3}$$

[22]Note that this is a stronger assumption than the standard monotonicty assumption. As noted by Heckman et al. (2006) in the presence of continuous instrument a more approptiate term for the monotonicity assumption is a uniformity.

[23]For example, if we partition coverage rate to n equal size intervals, we would average over $\frac{n(n-1)}{2}$ possible LATE estimates.

Where y^0_{ijc} is the potential outcome of student i who did not attended pre-K at age three, and y^1_{ijc} is the potential outcome of a student who did. β^0 is the vector of coefficients in the untreated state, and β^1 is the vector of coefficients in the treated state. The ATE will be captured by the difference in the coefficient of the constant in β^1 to its coefficient in β^0. Similarly, heterogeneous treatment effects on observables will be captured by the different components of the difference $\beta^1 - \beta^0$. U^k_{ijc} is the unobserved component in each state and its expectation conditional on the controls is normalized to zero. The treatment effect can be thus written as:

$$y^1_{ijc} - y^0_{ijc} = X_{ijc}(\beta^1 - \beta^0) + U^1_{ijc} - U^0_{ijc}$$

This formulation makes clear that under the linear separability assumption implicit in 3, the treatment effect is the sum of the effect on observables and the effect on unobservables.

Next, I turn to modeling selection into treatment. Let the choice enrolling in pre-K be given by $PreK3 = 1[Z_{ijc}\delta - V_i \geq 0]$. Z_{ijc} is a vector that includes in addition to the controls and fixed effects in 3, the excluded instrument(s). V_i is a random variable with cumulative distribution Φ and can be thought of as the unobsreved resistance to treatment. Note that student for whom $\Phi(Z_{ijc}\delta) = \Phi(V_i)$, are indifferent between treatment and no treatment. The left hand side of this equation is the propensity score, which I denote as $P(Z)$. The right hand side represents the quantiles of the unobserved resistance, which I will note as $U_D = \Phi(V_i)$. The MTE is defined by:

$$MTE(x, u_D) = \mathbb{E}[y^1_{ijc} - y^0_{ijc} | X = x, U_D = u_D] \tag{4}$$

Assumptions required for the MTE estimation are closely related to the assumption in 4.1. First we maintain the first stage assumption. Second, we impose a conditional independence of the instrument assumption: $Coverage \perp (U^1, U^0, V) | X, \theta, \alpha$. We also keep the monotenicity assumption. Finally, we assume separability between the observed and the unobserved treatment effect:[24]

$$MTE(x, u_D) = x(\beta^1 - \beta^0) + \mathbb{E}[U^1_{ijc} - U^0_{ijc} | U_D = u_D] \tag{5}$$

Thus, holding x constant, changes in u_D only affect the unobserved component and they identify the slope of the MTE with respect to u_D.

[24] As noted by Bhuller et al. (2020) the separability between observed and unobserved heterogeneity in treatment effects is a weaker assumption than the additive separability assumption usually employed in IV models (for example in equation 1).

4.2.2 Estimation Method

I follow Heckman et al. (2006) and Cornelissen et al. (2016) who showed that the model presented in the previous section can be written in the following outcome equation as a function of the propensity score:

$$\mathbb{E}(Y_{ijc}|X_{ijc}, \alpha_j, \theta_c, P(Z_{ijc}) = p_{ijc}) = X_{ijc}\beta^0 + \alpha_j + \theta_c + X_{ijc}(\beta^1 - \beta^0)p_{ijc} + K(p_{ijc}) \tag{6}$$

with p being the propensity score and $K(p)$ an unspecified function of the propensity score. As these authors showed the MTE can thus be expressed as the derivative of 6 with respect to the propensity score:

$$MTE(x, u_D) = \frac{\partial \mathbb{E}(Y_{ijc}|X_{ijc} = x, P(Z_{ijc}) = p)}{\partial p} = x(\beta^1 - \beta^0) + \frac{\partial K(p)}{\partial p} \tag{7}$$

I employ a two-stage estimation procedure to estimate the MTE curve. In my main specification, X includes the same controls as in section 4.1 and cohort and municipality fixed effects, and Z includes, in addition to X, coverage rate and its square interacted with a dummy for boys and a dummy for higher-educated mother. I allow for observable heterogeneity only across these two dimensions. In the first stage I estimate the propensity score, specifying Φ as a logistic distribution. In the second stage I estimate 6 by a OLS regression and, as is common in recent applications (see for example Cornelissen et al. (2018) and Bhuller et al. (2020)), model $K(p)$ as a global polynomial. In my main analysis I show the results for a third order polynomial, but results are stable over other orders. To compute standard errors I bootstrap the two-stage procedure and cluster at the cohort-municipality level. Finally, to compute the ATE, TT, and TUT, I aggregate over the estimated MTE curve using weights introduced by Heckman et al. (2006) and Carneiro et al. (2011).

5 Results

5.1 Two Stage Least Square

5.1.1 Main Results

I start by showing the results on the 2SLS estimates for the effect on graduating with a diploma. Table 3 shows results of the 2SLS procedures where I do not allow for treatment heterogeneity (that is $\tilde{X}$ in equation 1 includes only a constant). The first column in this table shows the OLS estimate of equation 1. As one might expect there is a positive correlation between attending pre-K and being awarded high-school diploma. I should note that even though in a general population we would be skeptical of attributing a causal interpretation to such a correlation, in this case some of the usual selection confounders might have a more minor role. This is because the regression is estimated on a restricted sample of municipalities, where variation in enrollment is driven mostly by an exogenous policy change and controls for a rich set of covariates. Nevertheless, I move to use variation in enrollment that only comes from variation of pre-K coverage rate.

Column 2 in table 3 shows the coefficients of coverage rate in the first-stage estimation which is described in equation 2. As was hinted by table 1, there is a strong positive and concave first-stage. Column 3 shows that the reduced form version of equation 1 yields a positive but imprecise effect of coverage rate on high-school diploma. Finally, column 4 shows the 2SLS estimate of the effect; the estimate, 4.5 percentage points, is not statistically significant and the OLS estimate in column 1 lies in its confidence interval.

I now turn to other graduation outcomes. Table 4 shows the 2SLS estimates of the effect on the total number of credits, on number of credits in Arabic, Hebrew, English, and Mathematics, and the effect on z-scores in Arabic and Hebrew. I use the same specification as in table 3. The effect on total credits is positive but is not statistically different from zero, and the effect on credits in Mathematics and English is also indistinguishable than zero but has a negative and small point estimate. On the other hand, students achievements in the Hebrew and Arabic Languages appear to be positively impacted by attending pre-K, with a stronger effect on Hebrew. The estimate for the number of credits in Hebrew is about half a credit and the effect on z-score is a about a fifth of a standard deviation (both are statistically significant). For Arabic, only the effect on the total number of credits is significant and the effect on z-score is marginally significant. In other words, students who attend pre-K are being tested in higher levels of Hebrew and preform better in the

Hebrew exam, while in Arabic these students are tested in higher levels but there is no significant effect on their performance.

5.1.2 Heterogeneous Effect and Mechanisms

The estimates presented thus far are focused on the effect on the LATE estimand. However, examining heterogeneity in the treatment effect across different populations could help understand when and where pre-K is a useful policy tool and what are the mechanisms that make it beneficial. In this section, I first explore the effect by gender. In my sample, gender has a significant role in predicting high school graduation (and other graduation outcomes); girls probability to graduate is 26.0 percentage points higher than boys. As some policy makers and other studies hypothesizing that pre-K is an equalizer,[25] examining heterogeneity across genders in my setting may illuminate this idea. The next heterogeneity dimension I study is parental education. This analysis builds on Schlosser (2011) that found that the 1999 Reform increased labor force participation among higher-educated mothers.

The top panel in table 5 shows the effect separately for boys and girls. In these regressions $\tilde{x}$ includes a dummy for boys in addition to a constant. We see that for girls, attending pre-K increases their probability of graduating with a diploma by 6.7 percentage points (p=0.070). The effect for boys, however, is only 2.8 percentage points and not significant. A similar pattern emerges on scores in Hebrew and Arabic. In contrast, the point estimates on total credits and credits in Hebrew and Arabic are higher for boys than girls, but only the effect on credits in Arabic is statistically significant in favor of boys.

The bottom panel in table 5 shows the effect separately by mother education. In these regressions $\tilde{x}$ includes a dummy for mother having more than 12 years of education in addition to a constant. Overall, there are almost no differences across the two groups. The exceptions are credits in Hebrew, where students of lower-educated mothers more positively impacted, and test scores in Arabic, where children of higher-educated mothers are more positively impacted.

[25] For example: https://www.washingtonpost.com/business/2019/01/13/castros-policy-agenda-universal-pre-k-comprehensive-immigration-reform-medicare-for-all/ (accessed 05/14/2020)

5.2 Marginal Treatment Effect

5.2.1 First stage

In this section I show results from the MTE estimation procedure described in 4.2. I start by showing results on the selection model. In Figure 4 I plot the predicted probability against the instrument using a LOWESS smoothing function. The figure shows that there exists a strong first-stage; the variation in the instrument alone induces a variation in the propensity score between 0 and 0.92. The figure also suggest that the effect of the instrument on enrollment is similar across sub-groups. Specifically, there are no differences between boys and girls (both in levels and slopes) and there are minor differences across mother-education groups, with students of higher-educated mothers being more affected by the instrument in some region of its domain.

Thus, figure 4 shows evidence for two important conditions for the MTE estimation. First, as I showed previously, a first stage relationship exists. Second, the variation in the instrument creates a substantial variation in propensity score over its full support. This condition is necessary to calculate the ATE, TT and TUT, from the MTE curve (See Cornelissen et al. 2016). Next, I show that in addition to a full support, there is also a common support. Figure 5 shows the distribution of the estimated propensity score by treatment status. Partitioning the propensity score support (the unit interval) to 100 equal size bins, there are more than 10 students in each treatment group for values of propensity score which range between 0.01 and 0.94.

5.2.2 MTE Curve and Treatment Estimates

Panel A in Figure 6 shows the estimated MTE curve for equation 7 where the outcome is graduating with a diploma and the intercept is taken at the covariates means. The figure shows the relationship between the unobserved treatment effect and unobserved resistance into treatment. Therefore, the figure allows us to analyze selection on gains due to unobserved resistance. The MTE exhibits a U-shape curve. Since higher values of u_D imply higher resistance, this U-shape suggests that the children that are most likely to enroll and least likely to enroll in pre-K, gain the most of attending pre-K. However, only for the highest levels of u_D the MTE is statistically different than zero, and thus figure 6 suggests that there exist a reverse selection gains.

Next, I aggregate over the estimated MTE curve and compute the standard treatment parameters (ATE, TT, and TUT) as demonstrated by Heckman et al. (2006) and Carneiro et al. (2011). These estimates for graduating with a diploma are plotted as horizontal lines in figure 7 (I also show the 2SLS estimate) and are listed for graduation with diploma and z-scores in table 6. Several interesting patterns are revealed from these estimates. First, the ATE point estimate for graduating with a diploma is 11.5 percentage points and is significant at the 10% level. This should be interpreted as the effect of taking a student at random from the population and assigning them to pre-K. Additionally, the TUT lies above the TT, and the ATE estimates; the TUT is large and significant while the TT is not. This is another illustration of reveres selection on gains.

The next interesting pattern is that all of these three treatment estimates lie above the 2SLS estimate. To shed more light on this pattern, I plot in panel B of figure 6 the weights on the MTE curve used to compute each of the treatment parameters and also show the implied weights for the 2SLS estimate.[26] We learn from this figure that the 2SLS estimate gives the highest weights to intermediate resistance students with the lowest marginal treatment effect. Moreover, for quantiles with the highest treatment effect the 2SLS gives a *negative* weight.[27] Thus, analysis of the 2SLS weights suggests that the 2SLS estimate, which can interpret as a weighted average of pairwise LATEs, is not close to any of the standard treatment parameters and it is hard to think of a case when the 2SLS weights represent an interesting estimand.

Finally, adding more outcomes to the MTE curve analysis in table 6, the pattern across outcomes is similar to the pattern presented in section 5.1.1. For test scores in Hebrew there is a strong and significant effect for all three treatment parameters, magnitudes are of the same order as the 2SLS estimates in table 4, and the differences across parameters are not large. For test scores in Arabic, I find no significant ATE and TT effects, and marginally significant TUT effect of 0.193 standard deviations. Importantly, all outcomes exhibit a reverse selection on gains.

6 Conclusions

In this paper I study the effect of attending public universal pre-K on educational achievements measured upon high school graduation. I analyze a reform in Israel that increased pre-K participation of children of the Israeli Arab population. I use the gradual roll out of the program to identify the causal effect of attending pre-K and I employ two empirical frameworks to that end: 2SLS and MTE.

[26]I follow the procedure suggested by Cornelissen et al. (2016) to implement Heckman et al. (2006) method to compute the 2SLS implied weights.

[27]The possibility of negative weights in 2SLS estimation where the instrument is continuous is discussed in Heckman et al. (2006)

I find that that in both approaches the estimated effect on test scores in the Hebrew language (a non-native language for children in the Arab sector) is large and statistically significant. This is an important result since many policy makers and researchers in Israel have argued that part of the persistent economic disparities between Jews and Arabs are a result of lack of proficiency in Hebrew among the Arab population. While I am not aware of intentional programs to improve Hebrew skills among pre-K students, pre-K teachers are more likely to have a better proficiency in Hebrew since most of them attended post-secondary schooling in Hebrew.[28] Unfortunately, I can not test this conjecture in the data.

I also find evidence for a positive effect on high-school graduation with a diploma: 2SLS estimates show that pre-K going has a positive effect only for girls; MTE estimates show a large and significant TUT and ATE and a non significant TT. Since there are a number of papers that study the relationship between high school diploma in Israel and labor market outcomes, I can make an attempt to do a back-of-the-envelope calculation of the returns on pre-K. A recent study by Lavy et al. (2020) on a high school intervention that had a positive effect on high school diploma of 12 percentage points (a similar magnitude to my ATE), found that in the long-run participants of that program earned annually additional $900. With an annual expenditure of $4,300 per student in pre-K and about 20% income tax, the government is expected to recoup its investment in 24 years after students enter the labor force. I believe this number is too low because: 1. I only have estimates for the cost of pre-K in 2015, and I expect the cost in 2000-2004 to be lower; 2. I expect the effect on earnings to be higher because the pre-K also affected proficiency in Hebrew. Therefore, more information on the effect of this program on labor force outcome is necessary for a more reliable cost-benefit estimate.

Taken together, my findings support the idea of a positive long-lasting effect of attending pre-K at age three with a few caveats. First, reverse selection on gains implies that an almost full take up is required to recoup the program's costs. Second, when extrapolating my results to other programs two issues should be taken into account: alternatives for treatment and quality of treatment. I study a setting in which no private childcare nor an alternative public childcare were available and children were in parental or informal care. Additionally, the program I study is of lower quality in most dimensions in comparison to other public pre-K programs in developed countries.

[28] An anecdotal illustration for this mechanism is presented in this video which recently gained traction in Israel and features an Arab Hebrew teacher who took an initiative and created her own method to teach her students Hebrew in a playful manner.

References

Attar, I. and D. Cohen-Zada (2018, September). The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. *Journal of economic behavior & organization 153*, 38–57.

Baker, M. (2011, November). Innis lecture: Universal early childhood interventions: what is the evidence base? *Canadian Journal of Economics/Revue canadienne d'économique 44*(4), 1069–1105.

Baker, M., J. Gruber, and K. Milligan (2008, August). Universal child care, maternal labor supply, and family WellâBeing. *J. Polit. Econ. 116*(4), 709–745.

Baker, M., J. Gruber, and K. Milligan (2019, August). The Long-Run impacts of a universal child care program. *American Economic Journal: Economic Policy 11*(3), 1–26.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020, April). Incarceration, recidivism, and employment. *J. Polit. Econ. 128*(4), 1269–1324.

Björklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in Self-Selection models. *Rev. Econ. Stat. 69*(1), 42–49.

Blass, N. and H. Bleikh (2012). Demographical change and kindergarten attandance. In A. Kimhi (Ed.), *Pre-Primary Education in Israel: Organizational and Demographic Perspectives*, Chapter 4, pp. 127–153. Taub Center.

Brown, J. (2019). Does public Pre-K have unintended consequences on the child care market for infants and toddlers?

Carneiro, P., J. J. Heckman, and E. Vytlacil (2011, October). Estimating marginal returns to education. *Am. Econ. Rev. 101*(6), 2754–2781.

Cascio, E. and D. W. Schanzenbach (2013). The impacts of expanding access to High-Quality preschool education. *Brookings Pap. Econ. Act. 44*(2 (Fall)), 127–192.

CBS (2004a). *Income Survey 2004*. Central Bureau of Statistics.

CBS (2004b). *Social Survey 2004*. Central Bureau of Statistics.

CBS (2005). *Statistical Abstract of Israel No. 56.* Central Bureau of Statistics.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016, August). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Econ. 41*, 47–60.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy 126*(6), 2356–2409.

Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of labor economics*, Volume 4, pp. 1315–1486. Elsevier.

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal Pre-Kindergarten on children's academic achievement.

Gormley, W. T. and T. Gayer (2005). Promoting school readiness in oklahoma: An evaluation of tulsa's Pre-K program. *J. Hum. Resour. 40*(3), 533–558.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2021, May). The Long-Term effects of universal preschool in boston.

Havnes, T. and M. Mogstad (2011, May). No child left behind: Subsidized child care and children's Long-Run outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006, August). Understanding instrumental variables in models with essential heterogeneity. *Rev. Econ. Stat. 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (2005, April). Structural equations, treatment effects and econometric policy evaluation.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proc. Natl. Acad. Sci. U. S. A. 96*(8), 4730–4734.

Kline, P. and C. R. Walters (2016, November). Evaluating public programs with close substitutes: The case of head start. *Q. J. Econ. 131*(4), 1795–1848.

Lavy, V., A. Kott, and G. Rachkovski (2020, December). Does remedial education at late childhood pay off after all? Long-Run consequences for university schooling, labor market outcomes and Inter-Generational mobility.

Schlosser, A. (2011). Public preschool and the labor supply of arab mothers: Evidence from a natural experiment. Unpublished Manuscript.

Yashiv, E. and N. Kasir (2014). The labor market of israeli arabs: Key features and policy solutions. *Unpublished working paper, Tel Aviv University*.

Zaban, H. (2012). Preschool teachers training in israel. In A. Kimhi (Ed.), *Pre-Primary Education in Israel: Organizational and Demographic Perspectives*, Chapter 4, pp. 127–153. Taub Center.

Tables

Table 1: Changes in Municipality-level Variables against Changes in Coverage Rate

	Dependent variable:			
	Pre-K3 Enrollment	Mother Education	Father Education	Family Size
	(1)	(2)	(3)	(4)
Pre-K3 Coverage Rate	0.387***	0.011	0.001	0.003
	(0.018)	(0.145)	(0.142)	(0.018)
Standard Deviation	0.301	1.062	1.186	0.962
Observations	140	140	140	140
Adjusted R^2	0.926	0.617	0.707	0.961

*p<0.1; **p<0.05; ***p<0.01

Notes: The table presents results from regressing pre-K coverage on the variable in the column header. Each regression includes municipality and school year fixed effects. The unit of observation is municipality-school-year. Robust standard errors are presented in parenthesis.

Table 2: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Enrolled in PreK-3	54,751	0.739	0.439	0	1
Father Schooling	49,623	10.222	2.963	0	22
Mother Schooling	48,678	9.979	2.870	0	21
Girl	54,751	0.487	0.500	0	1
Family Size	53,529	3.150	1.590	0	13
Birth order	52,185	2.968	1.846	1	14
Graduated w/ diploma	54,751	0.505	0.500	0	1
Total Credits	40,872	24.613	9.103	1	49
Credits in Arabic	34,108	3.252	0.674	1	5
Credits in Hebrew	36,543	3.523	1.111	1	5
Credits in Math	31,766	3.448	0.687	1	5
Credits in English	36,101	3.839	0.703	1	5
z-score Arabic	35,123	0.072	0.955	−4.556	2.283
z-score Hebrew	36,444	0.137	0.901	−4.240	2.216

Notes: The table presents summary statistics for birth cohorts 1996-2000 in the 1999 Reform municipalities. Summary statistics ignore missing values. z-score variables are computed with respect to all test takers in the country at a given school year.

Table 3: First stage, Reduced form and 2SLS estimates for Graduating with a Diploma

	Dependent variable:			
	Graduated	Enrolled PreK3	Graduated	
	OLS	First Stage	Reduced Form	2SLS
	(1)	(2)	(3)	(4)
Enrolled in Pre-K3	0.038^{***}			0.045
	(0.009)			(0.035)
Pre-K Coverage Rate		0.577^{***}	0.018	
		(0.035)	(0.034)	
Pre-K Coverage Rate Squared		-0.081^{***}	-0.0001	
		(0.014)	(0.012)	
Background controls		Yes		
Town FE		Yes		
Cohort FE		Yes		
Observations	54,751	54,751	54,751	54,751
Adjusted R^2	0.196	0.422	0.195	0.196

*p<0.1; **p<0.05; ***p<0.01

Notes: The table presents results from estimating the 2SLS model described in 4.1 where $\tilde{x}$ in equation 1 includes only a constant and the outcome is graduating with a diploma. Column 1 presents results from estimating the OLS version of equation 1, column 2 presents the first-stage estimates, column 3 present the reduced form and column 4 presents the 2SLS results. All regressions include include a dummy for gender, fixed effects for month of birth, order of birth and family size, and indicators for five groups of mother and father years of education. For each variable I add an indicator for missing values. Standard errors in parenthesis are clustered at the municipality-cohort level. The unit of observation is a student.

Table 4: 2SLS Estimates for Other Outcomes

	Dependent variable:						
		Credits				*Scores*	
	Total Credits	Arabic	Hebrew	English	Math	Arabic	Hebrew
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Enrolled in Pre-K3	0.735	0.149^{**}	0.437^{***}	-0.002	-0.011	0.103	0.215^{***}
	(0.673)	(0.064)	(0.118)	(0.052)	(0.027)	(0.065)	(0.078)
Observations	41,256	34,463	36,543	36,452	32,065	35,123	36,444
Adjusted R^2	0.179	0.108	0.138	0.185	0.125	0.185	0.183

Notes: The table presents results from estimating the 2SLS model described in 4.1 where the outcome is listed in the column header. See notes for table 4 for additional details.

Table 5: Heterogeneous 2SLS Estimates

	Dependent variable:							
			Credits				*Scores*	
	Graduation	Total Credits	Arabic	Hebrew	English	Math	Arabic	Hebrew
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Girls	0.067*	0.479	0.095	0.403***	0.009	0.003	0.164**	0.270***
	(0.036)	(0.710)	(0.073)	(0.123)	(0.055)	(0.032)	(0.073)	(0.084)
Boys	0.028	1.083	0.235***	0.487***	−0.020	−0.032	0.009	0.136*
	(0.037)	(0.706)	(0.065)	(0.121)	(0.055)	(0.036)	(0.066)	(0.078)
Mother education Low	0.050	0.934	0.144**	0.568***	−0.011	−0.057*	0.073	0.212***
	(0.037)	(0.702)	(0.067)	(0.130)	(0.059)	(0.034)	(0.069)	(0.081)
Mother education High	0.042	0.502	0.155**	0.295**	0.007	0.033	0.135**	0.219***
	(0.034)	(0.700)	(0.068)	(0.120)	(0.054)	(0.035)	(0.068)	(0.081)
Observations	54,751	41,256	34,463	36,543	36,452	32,065	35,123	36,444
Adjusted R^2	0.196	0.179	0.108	0.137	0.185	0.125	0.185	0.183

Notes: The table presents results from estimating the 2SLS model described in 4.1. Each cell shows results from a separate regression where $\tilde{x}$ in equation 1 includes in addition to a constant, a dummy for the variable listed in the first column and the outcome is listed in the column header. See notes for table 4 for additional details.

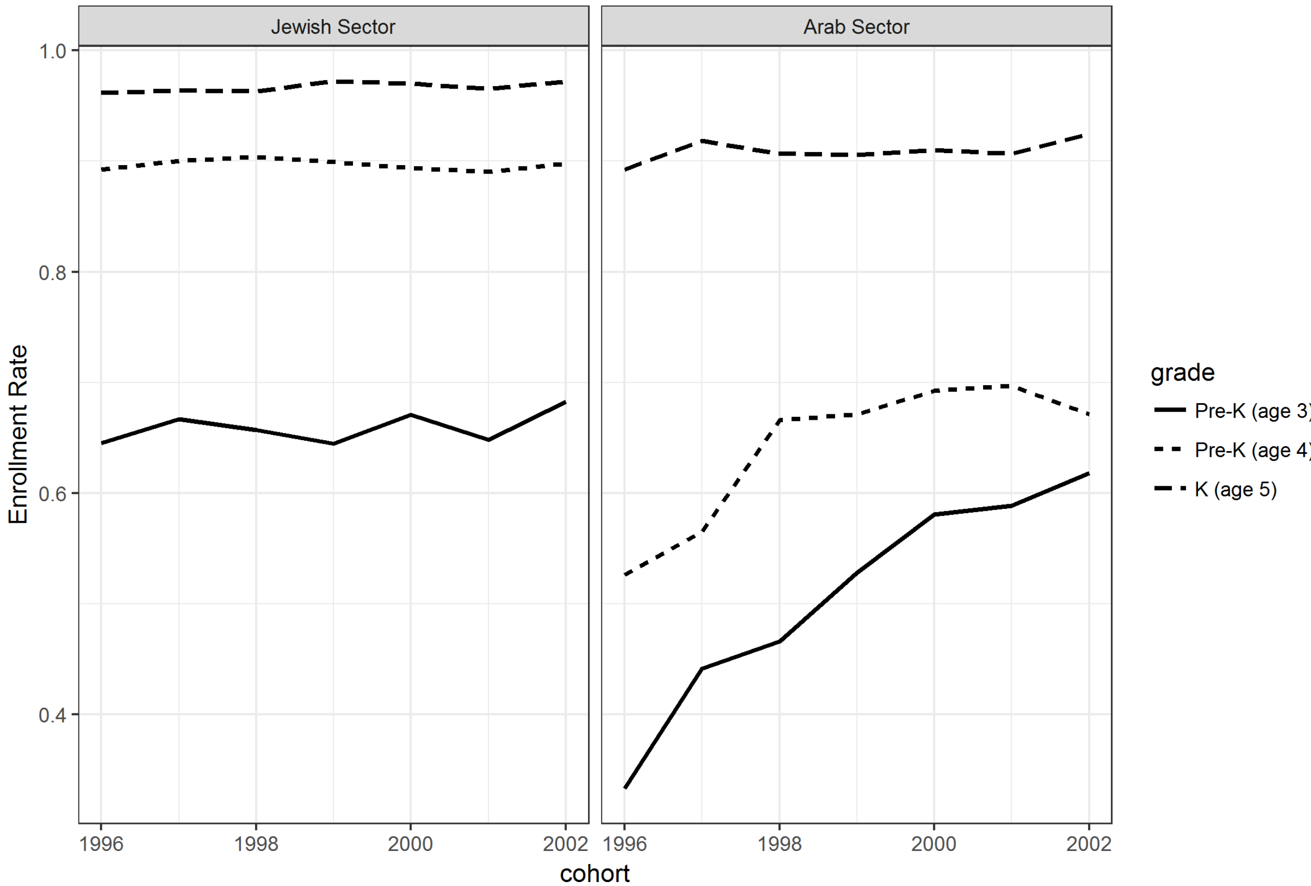
Table 6: MTE Estimates for Graduation Outcomes

	Outcome:		
	Graduation	Hebrew z-score	Arabic z-score
	(1)	(2)	(3)
ATE	0.115**	0.250***	0.112
	(0.058)	(0.076)	(0.074)
TT	0.088	0.228***	0.086
	(0.055)	(0.071)	(0.068)
TUT	0.201**	0.310***	0.193*
	(0.082)	(0.097)	(0.102)

Notes: The table presents results from estimated MTE model described in 4.2. The estimated parameters are the ATE, TT, and TUT, and the outcomes are listed in the column header. The estimates are calculated by aggregating over the MTE estimated curve (shown in figure 6) and using weights as demonstrated in table 1B in Heckman et al. (2006). Bootstrapped standard errors clustered at the municipality-school-year level are presented in parenthesis.

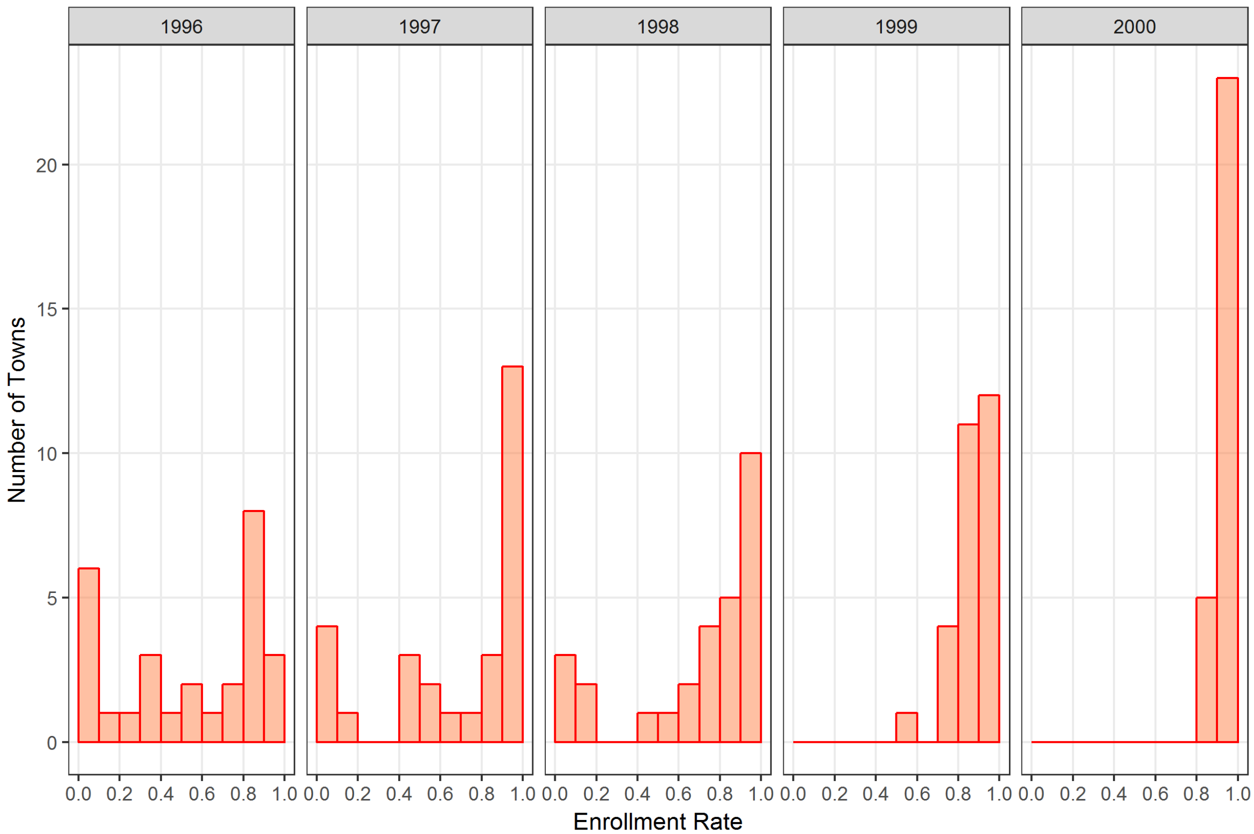
Figures

Figure 1: Trends in Pre-K and Kindergarten Attendance, by Cohort and Sector



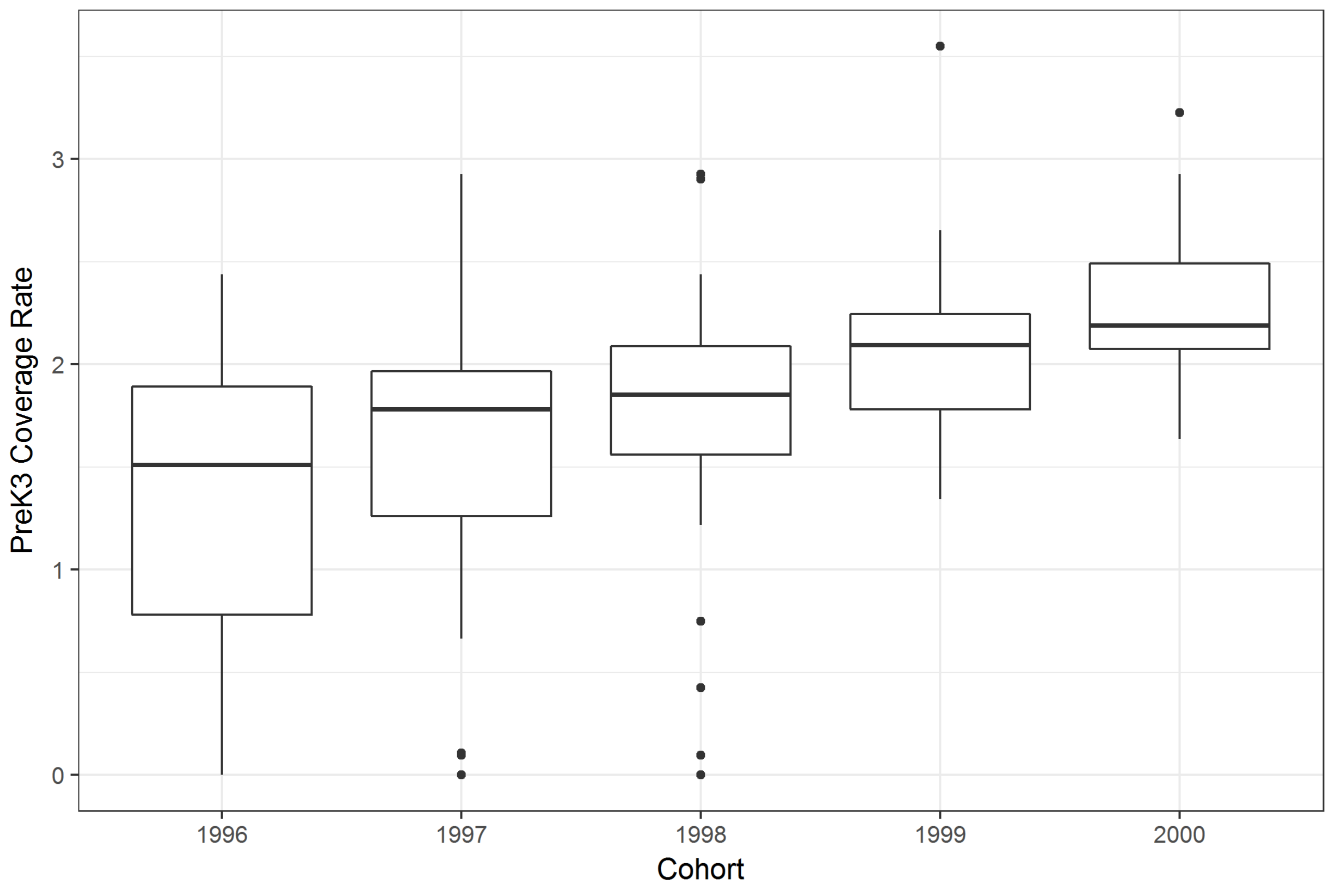
Notes: Each line indicates the share of students who attended pre-K for age 3, pre-K for age 4, and kindergarten for a given birth cohort of Arab and Jewish students.

Figure 2: The Distirbution of Enrollment rate in Sample Municipalites by birth cohorts



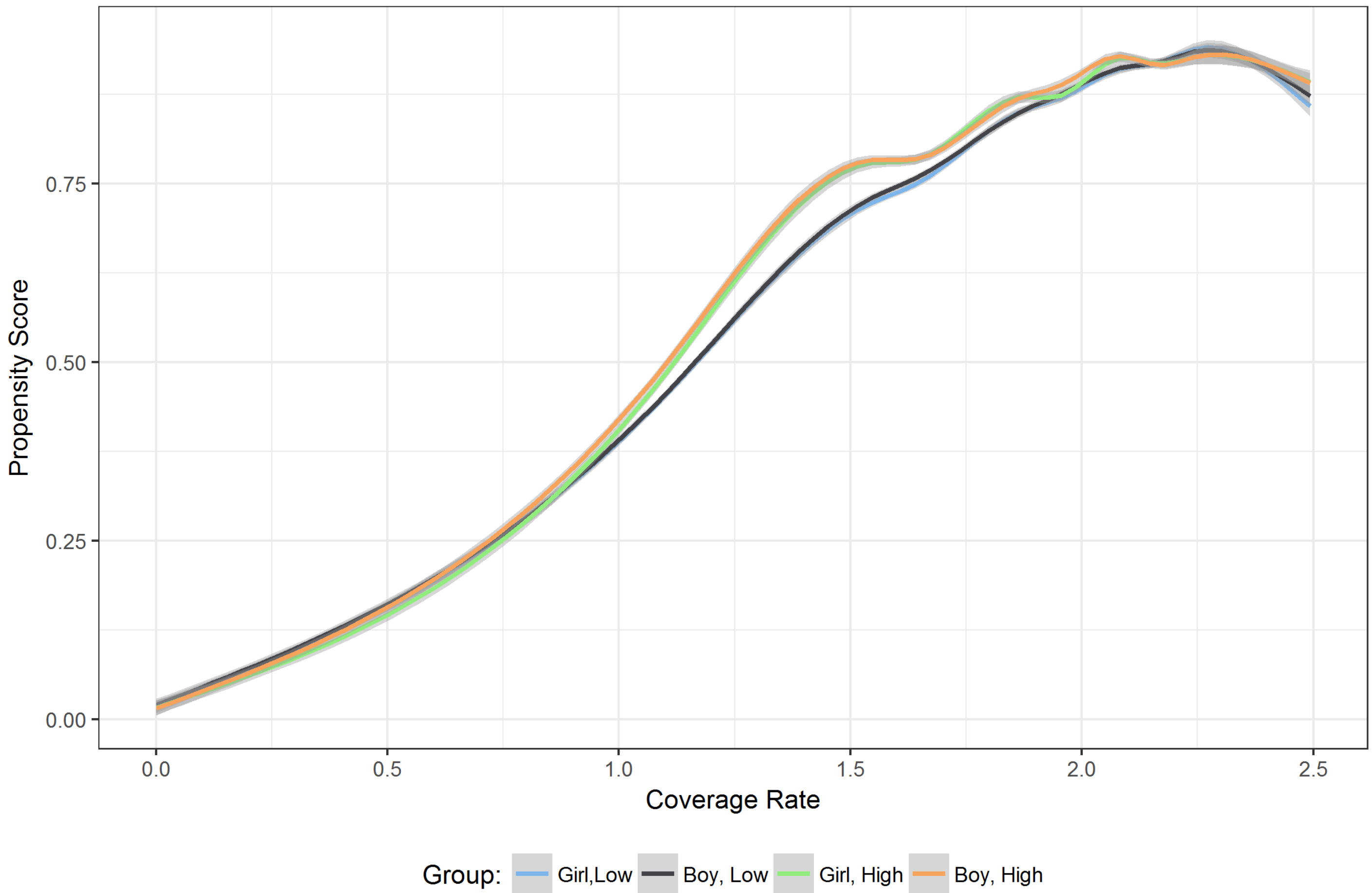
Notes: Each box presents the distribution of pre-K enrollment of three-year-olds across the 28 sample municipalities for birth cohorts 1996 to 2000. Birth cohort 1996 was eligible for three-year-olds public pre-K in school year 1999-2000, birth cohort 1997 was eligible in 2000-2001, and so on.

Figure 3: Evolution of Pre-K Coverage Rate



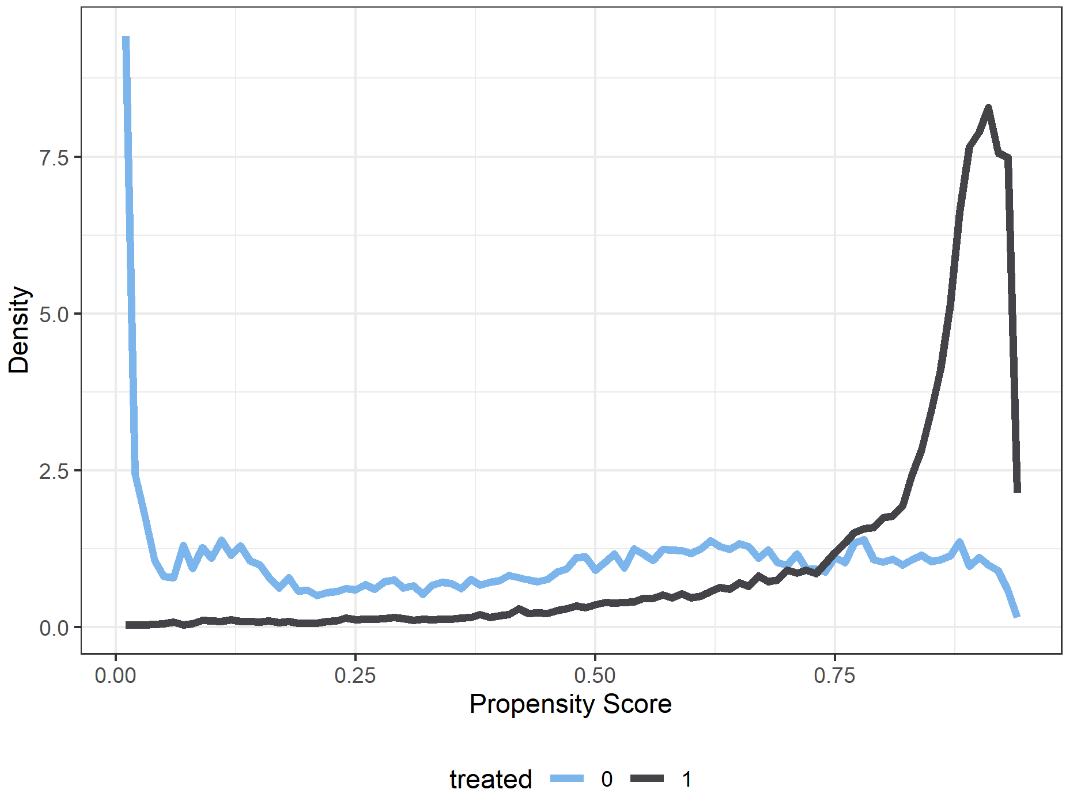
Notes: Each box-plot presents the distribution of public pre-K coverage rate across the 28 sample municipalities for birth cohorts 1996 to 2000. Birth cohort 1996 was eligible for three-year-olds public pre-K in school year 1999-2000, birth cohort 1997 was eligible in 2000-2001, and so on.

Figure 4: Logit Selection Model



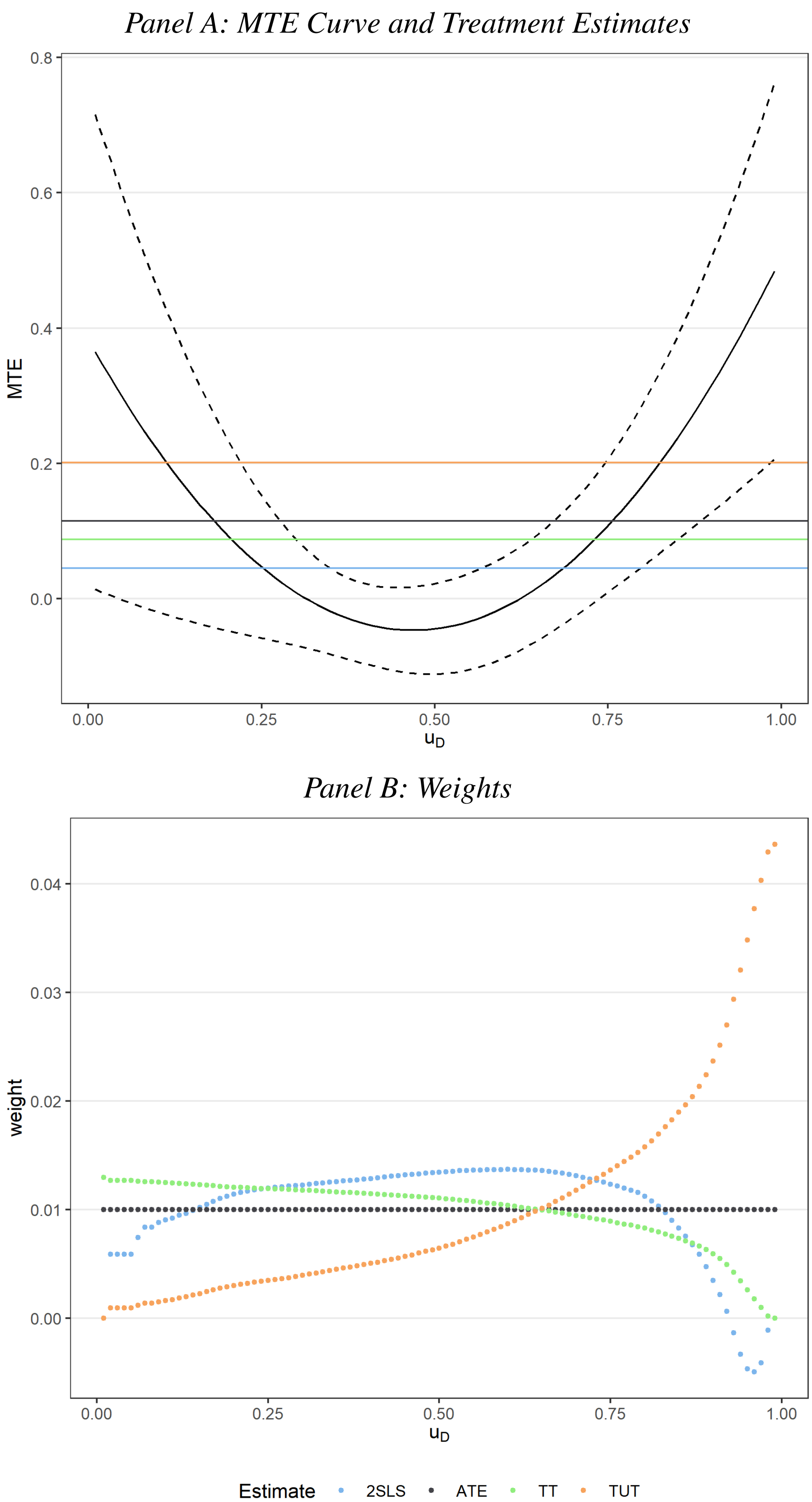
Notes: The figure shows the estimated propensity score as a function of pre-K coverage rate separately for girls with lower-educated mothers, boys with lower-educated mothers, girls with higher-educated mothers, and boys with higher-educated mothers. The propensity score is predicted with a logit regression. The regression includes the background controls in the notes table 3, municipality and cohort fixed effect and the coverage rate and its square interacted with a dummy for boys and a dummy for higher-educated mother. The lines show LOESS smoothing of student–level predicted probabilities and shaded area mark the 90% confidence interval.

Figure 5: Common Support



Notes: The figure shows the density of the estimated propensity score by treatment status. Propensity score is estimated as explained in the notes for figure 4.

Figure 6: MTE Estimates for Graduating with a Diploma



Notes: the figure shows results for the MTE estimation described in section 4.2. The bold line in Panel A depicts the MTE curve for graduating with a high school diploma, where the MTE is specified as described the notes for figure 4 and $K(p)$ specified as a third-order polynomial. The dashed lines in Panel A show the 90% confidence interval. The horizontal lines in Panel A show the point estimate 2SLS, ATE, TT and TUT effects on graduating with a diploma. Panel B shows the weights associates with each treatment parameter.