

Starting Together, Diverging Later? Gender Differences in Universal Pre-K's Long-Term Effects

Assaf Kott*

March 11, 2025

Abstract

I estimate the causal effect of starting to attend universal public pre-K at age three versus age four on school progression and high school achievements. The analysis leverages a dramatic expansion of public pre-K in Arab-majority towns in Israel, which generated within-household variation in siblings' pre-K attendance, enabling a sibling-fixed-effect model. I find that starting pre-K at age three significantly improves school progression, though high school academic achievements remain unaffected. By jointly observing pre-K attendance and long-term outcomes, I document patterns of heterogeneity in both program take-up and treatment effects. The results reveal stark gender heterogeneity, with girls gaining substantially more than boys from early pre-K attendance. Investigation of potential mechanisms yields results consistent with the hypothesis that this gender heterogeneity results from differences in home environments between girls and boys. The findings indicate that in communities with prevalent gender norms, public pre-K can serve as an efficient equalizing force.

JEL Codes: I21, I28, I26, J13, J16, J18

Keywords: Children, Preschool, UPK, Gender Gap

*Department of Economics, Ben Gurion University of the Negev, kotta@bgu.ac.il. This paper benefited from discussions with Anna Aizer, Chien-Tzu Cheng, Danny Cohen-Zada, Naomi Gershoni, Ada González-Torres, and Marcella Mello, as well as participants at the Workshop on the Economics of Education at Valle Nevado and seminars at Ben-Gurion University and Brown. I also thank Israel's Ministry of Education for granting access to the data. All errors are my own.

1 Introduction

A broad consensus across multiple disciplines holds that shocks and interventions during infancy and early childhood can have long-lasting effects into adulthood and can significantly influence life trajectories (Knudsen et al., 2006; Almond, Currie, and Duque, 2018). This recognition has fueled increased interest in the public provision of preschool.¹ However, designing a pre-K program entails making choices about key parameters—such as the appropriate starting age (3 or 4) and curriculum—and evidence on the relative importance of different factors remains limited. Although isolating the causal impact of each parameter is empirically challenging, gaining a clearer understanding of the mechanisms through which pre-K attendance improves long-term outcomes can help researchers and policymakers learn about the optimal design of such programs.

Empirical evaluation of public pre-K is challenging because different types of parents choose to send their children to preschool (or meet the conditions for sending them). Estimating the long-run consequences is even harder, as data on pre-K attendance and long-term human capital outcomes are rarely observed jointly. Consequently, many studies use variation in the availability of pre-K across time and space due to a pre-K reform to identify an intention-to-treat (ITT) parameter.² While this approach is useful for studying pre-K expansion as a whole, it is less helpful for conducting a treatment effect heterogeneity analysis—often used to infer underlying mechanisms—because heterogeneity in the ITT estimate could be a result of both heterogeneity in the treatment effect and heterogeneity in take-up.³

In this study, I provide treatment estimates of the long-term effects of starting universal

¹Additional incentives include promoting female labor force participation and increasing fertility rates.

²To derive estimates of the treatment effect, these studies employ the rationale of the Wald estimator, scaling the ITT by the reform’s first stage (or uptake rate). Importantly, information on the first stage usually comes from an auxiliary source.

³There are two additional limitations of ITT parameters. First, this approach estimates both the direct effect of attending pre-K and indirect spillover effects to nonparticipating peers in the same cohort and geographical unit. Second, making inference about this treatment estimate is problematic, as the first-stage information is obtained from outside sources.

public pre-K (UPK) at age three instead of age four. Using comprehensive administrative data from Israel, I observe both children’s attendance in public preschool at age three and their subsequent school progression and educational achievements upon high school completion. My main identification strategy employs a sibling-fixed-effect model, which addresses the selection of different types of families into pre-K by comparing siblings within the same household.

Sibling-fixed-effect estimates may be biased when unobserved factors systematically influence parents’ decisions to enroll some of their children in pre-K but not others. As noted by [Deming \(2009\)](#), “*Something* is driving differences in participation among siblings.” To address this identification challenge, I leverage a quasi-natural experiment: the expansion of public pre-K in Arab-majority municipalities in Israel during the early 2000s. While Israel’s public pre-K system has long served children starting at age three, many municipalities—particularly those with Arab majorities—historically lacked access to the system. In 2000, Israel launched a national initiative to implement universal free public pre-K. This reform disproportionately affected Arab communities, driving a dramatic increase in pre-K enrollment of over 30 percentage points within just six years. I demonstrate that this expansion is the primary source of within-family variation in preschool attendance, mitigating concerns about systematic within-family selection that could violate the underlying identification assumption.

My setting enables more precise estimates than typical in the literature—and thus greater statistical power to detect subgroup effect differences—as I exploit the full variation in actual pre-K attendance rather than relying on predicted treatment variation. While sibling fixed effects have been previously used in studies of the US’s Head Start program ([Currie and Thomas, 1995](#); [Deming, 2009](#)), this is, to my knowledge, the first study to apply this strategy to UPK using administrative data. These features—a UPK context and the use of administrative records—yield a substantially larger sample size and consequently more precise estimates.

I find that starting pre-K at age three rather than four significantly improves markers of school progression. Children who attended pre-K are 1.9 percentage points more likely to start first grade on time and 2.4 percentage points more likely to graduate from high school on time. However, I find limited evidence that pre-K improves educational achievements: The effects on obtaining a high school diploma (Bagrut) and performance in key subjects such as mathematics and English are small and statistically nonsignificant.

The analysis of treatment heterogeneity yields a striking finding: Girls benefit substantially and statistically significantly more than boys across all progression outcomes. For example, pre-K attendance significantly increases girls' likelihood of ever completing high school, an effect not observed in the pooled sample. For other dimensions of heterogeneity, I document important divergences between selection patterns and treatment gains. For relative age (measured by birth month relative to the school-entry cutoff),⁴ I find evidence of reverse selection on gains: Younger children benefit more from the program but are less likely to attend. In contrast, maternal education strongly predicts attendance, but children of more and less educated mothers experience similar gains, suggesting that ITT heterogeneity estimates along this dimension would primarily reflect differences in take-up rather than treatment effects.

In the second part of the analysis I take a deeper look into the treatment heterogeneity across genders and test mechanisms that could generate the heterogeneity. First, given that girls develop faster than boys (especially in their social-emotional skills) (Crockenberg, 2003), I test the hypothesis that boys are not as ready as girls to benefit from pre-K. However, contrary to this hypothesis, I find that relatively younger boys (those born later in a given year) benefit significantly more than relatively older boys. Next, since I study a population in which traditional gender norms are prevalent (Yassine-Hamdan and Strate, 2020), I test the hypothesis that boys and girls have different home environments and thus different counterfactuals to pre-K. My results are consistent with the hypothesis, as the heterogeneous

⁴The eligibility cutoff date for age-three pre-K (and other school-entry rules) was set according to the Hebrew calendar, typically falling in December (Attar and Cohen-Zada, 2018).

treatment effect across genders is stronger in (1) families with low maternal education, in which we expect gender norms to be stronger, and (2) families that have both boys and girls (rather than siblings of the same sex), in which we expect within-household resources to be skewed toward the preferred sex.

This paper makes three key contributions to the literature. First, it provides treatment estimates of the long-term impacts of attending a UPK program. While numerous studies document the short- and medium-term effects of public pre-K (Gormley and Gayer, 2005; Cascio and Schanzenbach, 2013; Lipsey, Farran, and Durkin, 2018; Weiland et al., 2020; Humphries et al., 2024), examining long-term treatment effects has proven much harder.⁵ Existing research has largely estimated the long-run effects of *exposure* to pre-K programs, rather than the effects of actual attendance (Havnes and Mogstad, 2011; Felfe, Nollenberger, and Rodríguez-Planas, 2015; Baker, Gruber, and Milligan, 2019).⁶ Most closely related to this paper is DeMalach and Schlosser (2024) (henceforth DS), which also looks at expansion of public pre-K in Arab towns in Israel and estimates the effects of exposure to this expansion; it finds improvements in academic outcomes and no significant effects on maternal employment or fertility. This study extends DS’s analysis, most notably by estimating direct treatment effects and by studying the effect of starting pre-K at age three rather than the combined effect of attending at ages three and four.⁷ Gray-Lobe, Pathak, and Walters (2022) provide the only other estimates of long-term *treatment* effects of UPK I am aware of, finding significant gains in high school completion and college enrollment using Boston’s

⁵Examining long-term outcomes is essential for two reasons: First, research on early-childhood programs frequently shows a fade-out of effects in the medium term followed by a re-emergence in adulthood (Currie and Thomas, 1995; Heckman, Pinto, and Savelyev, 2013; Bruhn and Emick, 2023). Second, standardized test scores alone might not capture the full spectrum of skills that matter for adult human capital (Jackson, 2018).

⁶The literature on Head Start’s long-term consequences is more advanced in terms of econometric methods and data usage (for example, Currie and Thomas (1995); Garces, Thomas, and Currie (2002); Deming (2009); Bailey, Sun, and Timpe (2021); Johnson and Jackson (2019); Carneiro and Ginja (2014); Ludwig and Miller (2007)). However, Head Start is inherently different from UPK, as it is a targeted program that provides a more comprehensive set of services.

⁷This study differs from DS in several other key ways: It employs an alternative identification strategy; it examines effects on progression through the school system; the geographical coverages differ in the sets of towns analyzed; and the study periods only partially overlap.

pre-K lottery data. This paper differs from theirs by examining a context with a different counterfactual (home care versus other preschool options) and focusing on public pre-K for age three specifically.

Second, this paper contributes by estimating the effect of attending universal preschool at an earlier age than typically studied in the literature. The majority of papers in this field focus on attending either pre-K at age four (Gormley and Gayer, 2005; Gray-Lobe, Pathak, and Walters, 2022; Cascio and Schanzenbach, 2013) or pre-K at ages three and four (or older) (Havnes and Mogstad, 2011; Baker, Gruber, and Milligan, 2008; Deming, 2009; DeMalach and Schlosser, 2024). Exceptions are Cornelissen et al. (2018), who study the expansion of the German pre-K system that shifted children from two to three years of attendance and estimate a treatment effect for short-term outcomes, and Felfe, Nollenberger, and Rodríguez-Planas (2015), who estimate how exposure to the introduction of public childcare at age three in Spain affected long-run outcomes. This study differs from theirs by estimating a treatment effect of attending pre-K at age three for long-run outcomes.

My third contribution is to characterize the mechanisms driving heterogeneous treatment effects between boys and girls. While many studies of preschool report heterogeneous effects by gender—typically favoring girls (Havnes and Mogstad, 2011; Cornelissen et al., 2018; Felfe, Nollenberger, and Rodríguez-Planas, 2015), though some find the opposite (Gray-Lobe, Pathak, and Walters, 2022)—the sources of these differences remain unclear. I test two potential mechanisms. The first is that girls and boys differ in their human capital production function and thus respond differently to interventions such as UPK. A large literature has examined the early origins of the educational gender gap, suggesting that boys develop social-emotional skills more slowly (Bertrand, Kamenica, and Pan, 2015; Crockerberg, 2003; Deming and Dynarski, 2008; Reeves, 2022). This may result in boys benefiting less from teacher interactions and unstructured curricula, which are common in preschool programs (Magnuson et al., 2016; Fidjeland et al., 2023).

The gender differences may also emerge when boys and girls face different counterfactuals

to public pre-K attendance (Kline and Walters, 2016). Thus, I test, as a second potential mechanism, whether gender differences in the home environment—which is the typical form of childcare, in the absence of pre-K, for the Arab population in Israel—drive treatment heterogeneity. While son preference has been documented across various contexts, it tends to be stronger in societies with greater gender inequalities (Dahl and Moretti, 2008; Blau et al., 2020). Similarly, gender differences in parental time and resource investments vary by context: Investments favor boys in developing countries but girls in developed countries (Baker and Milligan, 2016; Barcellos, Carvalho, and Lleras-Muney, 2014; Jayachandran and Pande, 2017). Beyond parental preferences and investments, gender norms may independently shape children’s experiences in the home environment (Boxberger and Reimers, 2019; Tandon, Zhou, and Christakis, 2012). These norms are particularly salient in my setting, where girls are less likely to play outdoors and more likely to participate in household chores (UNDP, 2006; Feki et al., 2017). Building on this literature, I test the hypothesis that public pre-K is especially beneficial for girls by compensating for lower investment at home and providing opportunities—such as outdoor play with peers—that are otherwise limited.

The remainder of this article is organized as follows. Section 2 provides background information on the UPK expansion and the Israeli context. Section 3 presents the data and sample construction. Section 4 outlines the empirical framework and validates its underlying assumptions. Section 5 presents the main analysis, examining how pre-K attendance at age three (versus age four) affects various educational outcomes. Section 6 explores heterogeneity in the results. Section 7 investigates potential mechanisms behind the observed gender differences in public pre-K returns. Section 8 concludes.

2 Background

Provision and expansion of public pre-K in Israel. Prior to 1999, local municipalities in Israel had the option to choose whether to provide pre-K education. These local programs

were subject to regulation by the central government, which determined tuition fees, teacher requirements, and other regulations. Subsidies were available based on household income and were jointly funded by the central government and local municipalities, with subsidy rules established at the national level. This setup created an incentive for poorer municipalities, particularly those with a high proportion of low-income households—most of which were majority Arab—to refrain from opening pre-K programs (Kimhi, 2012).⁸

Israel launched its UPK program in 1999 by amending the Free and Compulsory Education Law to cover children as young as 3 years old, expanding beyond its original K–12 scope. The law outlined a 10-year rollout by locality, empowering the minister of education to determine which areas would implement the reform. Using this authority, the minister prioritized specific municipalities and neighborhoods based on three criteria: municipalities in the bottom two deciles of an economic deprivation index, areas designated as confrontation lines, and regions classified as national priority areas. Since most Jewish localities already offered public pre-K before 1999, the pre-K expansion primarily benefited Arab children. The reform was implemented through two ministerial orders issued in 1999 and 2001, which specified the localities where the Free and Compulsory Education Law would apply to 3- and 4-year-olds. Two orders were necessary because many eligible towns were not initially equipped with the infrastructure needed to provide pre-K services to all children as required by law. Although the original plan was meant to be implemented nationwide, political and economic constraints delayed nationwide UPK coverage until 2015.

The institution of UPK led to a significant increase in pre-K attendance among Arab children. Panel A of Figure 1 illustrates that from 2000 to 2005, the proportion of Arab students attending pre-K at age three doubled, from approximately 30% to about 60%. During this period, pre-K attendance among the Jewish population remained stable and high. The analysis in this paper is focused on children who resided in the Arab-majority

⁸Poor Jewish-majority municipalities were more likely to provide pre-K, as they received assistance from the central government through the Priority Localities program. The Supreme Court ruled in 2006 that the criteria for receiving this funding were discriminatory.

localities included in the 2001 order since my data on student-level pre-K attendance start only in 2000. Panel B of Figure 1 shows that, similarly to the entire Arab population, Arab towns included in the 2001 order experienced a dramatic 20 percentage point increase in public pre-K attendance. Importantly, the baseline pre-K attendance is quite high (70%), as some municipalities included in the order were quicker in rolling out pre-K classrooms after the 1999 amendment.

Quality of Israeli public pre-K. In Israel, public pre-K programs operate for six hours a day, six days a week, 10 months a year. The Ministry of Education mandates that each class have a teacher and an assistant teacher. The teacher must hold a teaching certificate and, in many cases, also a bachelor's degree (Kimhi, 2012). Additionally, the ministry sets the maximum classroom size, which, at the time of the study, was 35 students. The average class size in my sample was 32, resulting in a relatively low adult-to-child ratio of 1:16. For comparison, Norway, Germany, and Spain have adult-to-child ratios of 1:8, 1:12.5, and 1:13, respectively (Cornelissen et al., 2018; Havnes and Mogstad, 2011). While I lack data on expenditure per student for the period under study, in 2019, the estimated expenditure per student was \$4,300 (PPP), significantly lower than that of most European countries and Head Start.⁹

The Arab population in Israel. Israel's population includes a significant Arab minority that has been part of the state since its establishment in 1948. Arab citizens of Israel, also known as Palestinian citizens of Israel,¹⁰ represent about 20% of all 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, 2005b). The majority of the Arab population (80%) live in majority-Arab towns or villages (Haddad Haj-Yahya et al., 2021). There are substantial economic disparities between the Arab and

⁹In 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 11/27/2024)

¹⁰This refers to Arab citizens within Israel's pre-1967 borders and does not include Palestinians residing in the West Bank or Gaza Strip.

Jewish populations in Israel. In 2004, income of Arab households was 36% lower than that of Jewish households (CBS, 2004). Part of this gap is driven by the lower participation of Arab women in the labor force, which in 2004 was 24% for women aged 25–64, compared to 79% for Jewish non-Ultraorthodox women (CBS, 2005a).

The low labor force participation rate of Arab women in Israel at that time was closely linked to childcare arrangements for young children. According to a 2004 survey by the Israeli Central Bureau of Statistics (CBS), 81% of Arab children aged 0–5 were cared for at home by a parent or unpaid relative.¹¹ This rate was significantly higher than in the Jewish population, in which only 30% of children in the same age group received such care. The disparity extended to private early-childhood programs as well: During the 2004/5 school year, just 3.1% of Arab children aged 0–6 attended private programs, compared to 9.5% of Jewish children.¹² Given these patterns, the counterfactual to public pre-K attendance was most likely home care by mothers or relatives.

High school diploma (Bagrut). In Israel, high school students are examined in a series of centrally administered matriculation exams. 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 collectively worth at least 20 credits. Some subjects (such as English and math) have a minimum credit requirement, and all students have to take at least the lowest-level exam in these subjects. I therefore define an advanced subject as one in which a student earns more than the minimum required credits.

3 Data and Sample Construction

The data used in this study are sourced from administrative records of the Israeli Ministry of Education. I linked two types of files. First are enrollment records, which cover all students

¹¹Source: https://www.cbs.gov.il/he/publications/doclib/2004/social_survey/pdf/ty01.pdf (accessed 11/27/2024)

¹²Source: <https://www.cbs.gov.il/he/publications/doclib/2005/children/pdf/t07.pdf> (accessed 11/27/2024)

in the public education system. These records include details on each student’s grade, school identifier, and demographic characteristics such as date of birth, gender, ethnicity, immigration status, and parental education. Additionally, these records contain unique identifiers for parents, enabling sibling identification. The coverage extends from the 1995/96 school year to 2018/19 for students in 1st through 12th grades, and from 1999/2000 to 2017/18 for public pre-K and kindergarten participants. The second type of file contains data on achievements on high school matriculation tests. These outcomes include an indicator for whether a diploma was awarded, total number of credits awarded, and credits awarded in English, math, and Hebrew.

I begin by constructing a census of all children born between 1996 and 2001.¹³¹⁴ These children were aged three during school years 1999/2000 through 2004/5, coinciding with the pre-K expansion in Arab municipalities. I then restrict my sample to Arab-majority municipalities that were included in the special order issued in 2001.¹⁵ This sample restriction leaves 33 towns and an average of 9,105 students per cohort, compared to an average of 32,520 Arab students per cohort in the population. Column 1 of Table 1 presents summary statistics for the entire analysis sample, while column 2 focuses on children included in the sibling-fixed-effect analysis—that is, children from families with at least two children in the main sample. The sibling sample is characterized by slightly larger families, lower parental education, and slightly lower academic achievements compared to the main sample.

4 Empirical Strategy

A simple comparison of students who attended pre-K with those who did not is potentially biased, as the decision to enroll a child is correlated with other significant determinants of

¹³I can observe nearly all children in these birth cohorts, as almost all children in Israel attend publicly funded schools and thus appear in my data when they enroll in first grade.

¹⁴The 1996 birth cohort is the first with available data on pre-K attendance at age three, while 2001 is the last cohort for which I can observe high school graduation outcomes.

¹⁵In the main analysis, I do not use towns included in the 1999 special order since my pre-K enrollment data only begin in 2000.

human capital. Table 2 illustrates this. The table displays coefficients from regressing pre-K attendance on student and family-background characteristics. Column 1 presents results from univariate models, while columns 2 and 3 show results from a multivariate model that includes all the variables (column 2) and town-cohort fixed effects (column 3). The results reveal significant selection in who enrolls their children in public pre-K: Children of parents with higher education levels and children born earlier in the year are more likely to attend pre-K. As shown in Appendix Table A1, this selection into treatment biases OLS estimates of public pre-K’s effect on educational outcomes: The pre-K coefficient decreases substantially when controls are added, suggesting substantial selection on both observable and potentially unobservable characteristics (Oster, 2019).

To overcome this selection bias, I compare siblings within the same family who attended and did not attend public pre-K by estimating a family-fixed-effect model. The family-fixed-effect approach controls for family unobserved characteristics that are time invariant. This method has been used in the preschool literature, mainly in the context of Head Start (Currie and Thomas, 1995; Garces, Thomas, and Currie, 2002; Deming, 2009). Formally, I estimate the following model:

$$y_i = \alpha + \beta PreK_i + X_i\gamma + \delta_{j(i)} + \tau_i + \epsilon_i \quad (1)$$

Here, y_i is child i ’s educational outcome, $PreK_i$ is an indicator for whether child i attended public pre-K at age three, X_i is a vector of child i ’s background characteristics (which include sex, month of birth, and birth order), $\delta_{j(i)}$ is a family fixed effect, and τ_i is a town-by-birth-cohort fixed effect.

The primary focus of this paper is β . For β to have a causal interpretation as the effect of attending pre-K, the identifying assumption requires that within a family, attendance in a public pre-K program is not correlated with a child’s unobserved characteristics. For instance, if parents choose to enroll children with lower endowments in public pre-K as a compensatory measure, this could lead to an underestimate of the true effect. Conversely, if mothers who lose their jobs decide against sending their children to public pre-K, this could

result in an overestimate of the true effect, as pre-K attendance will be positively correlated with household income at age 3. Considering that the program is free and demands minimal investment from parents, both scenarios are unlikely. However, as noted by Deming (2009), “*Something* is driving differences in participation among siblings.” In this setting, the most probable cause of within-family variation is the substantial expansion of public pre-K in Arab municipalities during the sample period. This expansion is likely to induce within-household variations in attendance that are uncorrelated with unobserved differences among siblings.

To further investigate variation in public pre-K attendance within families, I adapt concepts from the local average treatment effect framework (Imbens and Angrist, 1994), with a slight abuse of terms. I categorize families into four types based on the pre-K enrollment status of the oldest and youngest child¹⁶: “never takers” (who sent none of their children to public pre-K), “always takers” (who sent all their children), “expansion compliers” (who sent their youngest but not eldest children), and “expansion defiers” (who sent their eldest but not youngest child). Following Miller, Shenhav, and Grosz (2023), I term the expansion compliers and defiers collectively as “switchers.” Importantly, only switchers help identify β (Miller, Shenhav, and Grosz, 2023).

The pattern that emerges in Table 3 is consistent with the idea that within-family differences in pre-K attendance are due to the expansion: Expansion compliers—families who, consistent with increased availability of pre-K, send only younger siblings—make up about 80% of switchers. The table also provides suggestive evidence that for a large portion of families that are expansion defiers—a group that is small—the reason for sending the oldest child but not younger children is that the younger children were born late in the year.¹⁷

The switcher analysis suggests that there are two important variables to control for. First, since I mostly compare older siblings to younger ones, a cohort fixed effect is necessary; otherwise, the estimates might conflate time trends with the treatment effect. Second, since

¹⁶The classification is based on the oldest and youngest siblings present in the sibling sample.

¹⁷Lower enrollment rates among children born late in the year may stem from two factors: parents’ concerns about readiness, as these children would be the youngest in their class; or towns’ capacity constraints, which lead them to prioritize older children.

within expansion-defier families treatment status among siblings is correlated with month of birth, I add a month-of-birth fixed effect. Importantly, both sets of parameters are identified off variation among siblings in always-taker and never-taker families.

Finally, the switcher analysis reveals that the most common counterfactual to attending pre-K at age three is attending it at age four. This can be seen in the last two rows of Table 3, which display pre-K attendance rates at ages three and four, respectively, for the oldest sibling across the four family types. By definition, the share of oldest children attending pre-K at age three is zero for never-taker and expansion-complier families, and one for always-taker and expansion-defier families. Among complier families, 73% of oldest siblings attended pre-K at age four. Therefore, my identification strategy primarily compares older children who did not attend pre-K at age three but did attend at age four with their younger siblings who started attending at age three. Thus, I interpret the results as the effect of beginning pre-K one year earlier—at age three rather than four.

5 Results

Table 4 presents results from estimating equation 1 for various outcomes. Panel A focuses on school-progression outcomes. The estimates suggest that pre-K attendance significantly enhances students' progression through the school system. Columns 1–3 show that children who attended pre-K are 1.9 percentage points more likely to start primary school on time, 1.3 percentage points less likely to repeat a grade, and 0.9 percentage points more likely to advance to 11th grade. While the second and third estimates are not statistically significant, column 4 shows that all three factors contribute to a statistically significant improvement in on-time high school graduation by 2.4 percentage points. However, this effect primarily reflects better progression rather than improved educational attainment, as the impact on overall high school graduation is smaller and not statistically significant (column 5).

Panel B further indicates that pre-K plays a limited role in improving educational achieve-

ments. Column 1 shows a small and statistically nonsignificant effect on obtaining a Bagrut diploma. While column 2 suggests some improvement in the quality of the diploma, as measured by an increase in diploma credits, this increase does not stem from taking additional credits in key subjects such as mathematics, English, and Hebrew (columns 3–5).

In the main analysis in Table 4 I include in the sample expansion defiers for two reasons. First, as discussed above, the decision not to send the youngest children is plausibly exogenous, as the majority of these children are born late in the year (and I control for month of birth). Second, the division into four family types is somewhat simplistic since families may have more than two children in the sample. However, Table 5 demonstrates that my results are robust to excluding expansion defiers. The findings are also robust to several additional specification checks. First, Table 5 shows that the results hold when imposing common cohort fixed effects across all towns, rather than town-specific cohort fixed effects. Second, while my main analysis focuses on Arab-majority towns included in the 2001 special order, expanding the sample to include Arab-majority towns in the 1999 order reveals similar patterns. Finally, to address concerns about families with more than two children receiving higher weight (Miller, Shenhav, and Grosz, 2023), I estimate models using only the oldest and youngest child from each family in the sibling-fixed-effect sample. These estimates remain consistent with the main analysis.

To contextualize my results, I compare my estimates to existing findings. I begin by examining estimates from large-scale pre-K programs that report treatment effects on similar outcomes and provide standard errors or confidence intervals. Panel A of Figure 2 shows that my estimates are smaller than other studies', though they generally fall within these studies' confidence intervals. The figure also demonstrates that my estimates are more precise than most existing estimates. The weaker effects of the Israeli expansion may stem from the program's lower quality as reflected in its lower per-student expenditure. Notably, despite the smaller effects, my results demonstrate that even lower-quality programs can improve school-progression outcomes, leading to more on-time high school graduation and potentially

earlier labor force entry.

Next, I compare my findings to estimates from DS, which also studies the pre-K expansion in Arab municipalities in Israel.¹⁸ Panel B of Figure 2 shows that my estimates are smaller than those in DS for overlapping outcomes. The most plausible explanation is that I estimate the effect of starting pre-K at age three rather than four, while DS studies an expansion that increased enrollment at both ages three and four (Figure 2 in DS). This gradient in effect sizes, in which impacts increase with child age, aligns with evidence from studies of public childcare for children aged zero to three that document negative impacts on child development (Baker, Gruber, and Milligan, 2008; Fort, Ichino, and Zanella, 2020). While a few studies examine pre-K at age three (Cornelissen et al., 2018; Felfe, Nollenberger, and Rodríguez-Planas, 2015), the majority of the literature focuses on age four (or ages three and four combined), and little is known about how returns to pre-K at age three compare to returns at age four in the same context. Thus, taken together, this study and DS suggest that expansion of pre-K at age four yields higher returns.

Additional important factors that can explain differences between my findings and those in DS are first, that in the presence of sibling spillovers, I may underestimate the true effect of attending pre-K; and second, that DS focuses on earlier years, and its estimates represent the effects on families who were quick to take up pre-K, while I examine families who enrolled later and thus may represent a more resistant population. If selection on gains into pre-K exists, we would expect smaller estimates in my setting.¹⁹

6 Heterogeneous Effects

I next analyze how effects vary across student subgroups. While prior research often reveals heterogeneous impacts of public pre-K, most long-term studies can only estimate ITT effects

¹⁸To enable appropriate comparison, I derive treatment estimates by dividing the ITT estimates by the first stage reported in DS. I construct confidence intervals by similarly scaling the implied confidence intervals from their study using the ITTs' standard errors.

¹⁹Another difference is that DS restricts its analysis to towns in northern Israel, while my study includes Arab municipalities across all regions.

because data on program participation are incomplete. By contrast, I am able to estimate treatment effects directly, providing new insights into how the impact of public pre-K varies across populations.

Table 6 reports estimates of the fixed-effects model for the school-progression outcomes for different subgroups.²⁰ Since this exercise involves cutting the sample multiple times into smaller subsamples, I report for each sample cut the p -value from a test that all differences are zero for all outcomes.

Columns 1 and 2 present results for boys and girls, respectively. The effects are consistently larger for girls: Girls who attend pre-K are more likely to start 1st grade on time than boys, are less likely to repeat a grade, and show a more pronounced increase in the probability of reaching 11th grade. These differences translate to a larger effect on girls' on-time high school graduation rates compared to boys. Notably, I find a significant effect on high school completion for girls—a finding not observed in the pooled sample. I can reject the null hypothesis of no differences across all outcomes with high statistical significance ($p = 0.000013$). In the next section, I explore what may explain this striking difference in treatment effects between boys and girls.

Columns 3 and 4 examine heterogeneity by relative age, comparing students born early versus late in the year.²¹ This analysis is particularly relevant given that month of birth significantly predicts enrollment: Children born later in the year are less likely to attend public pre-K (Table 2). While the point estimates suggest larger benefits for younger students, these differences are only marginally significant for on-time high school graduation, and a joint test cannot reject zero differences. These patterns suggest possible reverse selection on gains, where the group that potentially benefits more (younger children) is less likely to receive treatment.

²⁰Table A2 shows heterogeneity analysis for educational achievements, for which, as in the main analysis, I do not find an effect.

²¹I define older students as those born from January through August, and younger students as those born from September to December. This choice aligns with the school-entry-age cutoff date and the Ministry of Education policy that considers the latter group less ready to start first grade on time, thus creating more lenient conditions for redshirting younger children.

I also examine heterogeneity by maternal education and family size. While maternal education strongly predicts pre-K attendance (Table 2), columns 5 and 6 of Table 6 show no significant differences in treatment effects across maternal education levels, despite point estimates suggesting slightly larger benefits for children of more educated mothers. Similarly, columns 7 and 8 reveal no statistically significant differences by family size, though the point estimates indicate potentially larger benefits for children from larger households.

The divergence between selection patterns and treatment effects found in this section underscores the limitations of ITT heterogeneity analyses. For instance, while younger children appear to benefit more from pre-K, they are less likely to participate—a pattern that ITT estimates would systematically understate. Conversely, children of more educated mothers are more likely to attend pre-K despite not showing significantly larger benefits, potentially leading ITT analyses to overstate effect heterogeneity along this dimension. Most notably, the analysis reveals dramatically larger benefits for girls across all educational-progression outcomes, with effects that are both statistically and economically significant.

7 Mechanisms Behind Gender Heterogeneity

What mechanisms might explain the striking gender differences in returns to UPK? A hypothesis can be formed based on the voluminous literature on the early origins of gender gaps in educational achievements (Reardon et al., 2019; Autor et al., 2016; Lavy and Sand, 2018). This literature finds that girls have early advantages in social and behavioral skills, which may enhance their ability to benefit from pre-K environments. Specifically, boys are more prone to experiencing temperament and self-regulation challenges (Bertrand and Pan, 2013; DiPrete and Jennings, 2012), whereas girls exhibit a greater propensity for engaging in constructive interactions with both peers and adults (Magnuson et al., 2016). Furthermore, girls tend to derive more benefit from unstructured curricula (Fidjeland et al., 2023). These developmental disparities are reflected in school-readiness assessments, where boys

are more frequently perceived as less prepared for first grade and consequently experience higher rates of redshirting (Cook and Kang, 2020; Deming and Dynarski, 2008). This apparent lag in boys’ noncognitive skill development has led some experts to propose extending their kindergarten experience by an additional year (Reeves, 2022).

Thus, it is possible that the pre-K classroom environment is more conducive to girls’ learning because of their better interactions with teachers, greater benefits from unstructured play, and overall higher level of readiness. While I cannot provide direct evidence for the first two factors, I can test whether my results are consistent with the readiness hypothesis. If differential school readiness were driving the heterogeneity in returns to UPK, we would expect young boys, presumed to be the least ready, to benefit the least from attending UPK. In Panel A of Figure 3, I test whether gender differences in readiness explain the gender gap in UPK returns by estimating heterogeneous effects of UPK attendance on graduating from high school on time—which, as shown in the previous section, serves as a good summary index for overall school progression.²² The pattern that emerges in the figure contradicts the readiness hypothesis: The point estimate for young boys is positive and statistically larger than the estimate for older boys.

What else can explain heterogeneity in the returns to UPK? The literature on early childhood education highlights the importance of the quality of the counterfactual care environment in generating treatment heterogeneity (Kline and Walters, 2016). Since most children in this setting would remain in home care in the absence of UPK, sex differences in the home environment emerge as a potential explanation for treatment heterogeneity. In this setting—Arab towns with a Muslim majority—gender norms and parental preferences for boys may create different home environments for young children: Girls typically stay home with a parent or other adult relatives and participate in household chores, while boys are more likely to spend time outdoors and play with peers (UNDP, 2006; Feki et al., 2017). Consequently, the introduction of UPK may have represented a more dramatic change for

²²In Appendix Figures A1 to A4, I show that these patterns are largely consistent across other school-progression outcomes.

girls by providing substantial exposure to peer play and nonrelative-adult interactions. For boys, these experiences were already more accessible even in the absence of UPK. Moreover, parental investments could play a crucial role. Given that favoritism toward boys is well documented in Arab and Muslim communities (Yassine-Hamdan and Strate, 2020), the introduction of UPK might have provided girls with access to quality time with adults, while for boys, the change in adult interaction was less substantial.

I provide two tests for the home-care-quality hypothesis. First, I examine whether the heterogeneity by sex is stronger in more traditional households, using maternal education as a proxy. Since gender norms and son preference are expected to be more pronounced in more traditional households, under the home-care-quality hypothesis the heterogeneous effect across sexes should be stronger among this group. Panel B of Figure 3 confirms this prediction: Girls from more traditional households benefit significantly more than boys from pre-K attendance, while in less traditional households there is no statistically significant difference between the sexes.

Second, I examine whether heterogeneity by sex is stronger among families with both sons and daughters compared to families with children of only one sex. Since there should be less scope for son preference in investments in families with only daughters, the home-care-quality hypothesis predicts that the heterogeneous effect should be driven by girls who have brothers. Panel C of Figure 3 corroborates this prediction: Girls from mixed-sex families gain significantly more from attending pre-K than boys, while in single-sex families there is no statistical difference between girls and boys.

8 Conclusions

This paper provides treatment estimates of the long-term effects of starting pre-K at age 3 rather than 4. By leveraging comprehensive administrative data from Israel and exploiting a natural experiment that generated within-family variation in pre-K attendance, I obtained

precise treatment-effect estimates of attending pre-K at age 3 on school progression and long-term educational-achievement outcomes. I find that attending pre-K at age 3 significantly improves school progression but has limited effects on academic achievements. These findings suggest that while earlier pre-K entry can help children progress through school more smoothly, it might not substantially enhance their academic performance. While, in theory, smoother school progression can result in significant increases in public funds—as children will enter the labor force earlier and thus start paying income taxes earlier—in practice, under reasonable assumptions, these savings are small and cannot justify pre-K at age 3 by themselves.²³

The analysis reveals striking heterogeneity in treatment effects across gender, with girls benefiting significantly more than boys across all progression outcomes. I explore two potential mechanisms for this pattern: differential school readiness and differences in the home environment. The results align with the latter hypothesis, as the gender-treatment heterogeneity is stronger for children from families that are more likely to adhere to gender norms. This finding highlights the potential for UPK to serve as an equalizing force by providing opportunities that some children, particularly girls in traditional settings, might not otherwise receive at home. While this is a desired outcome, it will contribute to widening the gender educational gap, as girls outperform boys across all educational outcomes in my setting.

The findings of this paper raise several questions for future research. First, are the smaller effects I find (compared with other studies' findings) the result of estimating the impact of attending pre-K at age three rather than four? Or are they the result of a lower-quality pre-K program, as reflected by the lower expenditure per child, larger classrooms, and lower teacher qualifications? In addition, as I stop the analysis at the end of high school, it remains

²³Assuming that children enter the labor force right after graduating from high school and earn NIS 7,000 in the first year with a 4% annual increase, children who graduate on time start earning at age 19 (16 years after pre-K) and children who graduate 1 year later start earning at age 20 (17 years after pre-K). The difference in discounted lifetime income between these two cases is NIS 2,721. Where $t=0$ is the year of pre-K for age 3 and the discount factor is 3%. Assuming a 20% income tax and an effect of 0.024 on graduating on time, I calculate that the government saves NIS 65 for every child attending pre-K. This is significantly lower than the expenditure per child of about NIS 16,000 in 2019.

unknown whether later life outcomes will follow the pattern of school-progression outcomes (which show significant improvements) or educational-achievement outcomes (which show limited effects). Finally, a crucial question is whether it is possible to modify the preschool program in ways that maintain the benefits for girls while also generating positive effects for boys.

References

- Almond, Douglas, Janet Currie, and Valentina Duque. 2018. “Childhood Circumstances and Adult Outcomes: Act II.” *Journal of Economic Literature* 56 (4): 1360–1446.
- Attar, Itay, and Danny Cohen-Zada. 2018. “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.
- Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2016. “School Quality and the Gender Gap in Educational Achievement.” *American Economic Review* 106 (5): 289–295.
- Bailey, Martha J., Shuqiao Sun, and Brenden Timpe. 2021. “Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency.” *American Economic Review* 111 (12): 3963–4001.
- Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. “Universal Child Care, Maternal Labor Supply, and Family Well-Being.” *Journal of Political Economy* 116 (4): 709–745.
- Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2019. “The Long-Run Impacts of a Universal Child Care Program.” *American Economic Journal: Economic Policy* 11 (3): 1–26.
- Baker, Michael, and Kevin Milligan. 2016. “Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries.” *Journal of Human Capital* 10 (4): 399–441.
- Barcellos, Silvia Helena, Leandro S. Carvalho, and Adriana Lleras-Muney. 2014. “Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?” *American Economic Journal: Applied Economics* 6 (1): 157–189.
- Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. “Gender Identity and Relative Income within Households.” *The Quarterly Journal of Economics* 130 (2): 571–614.

- Bertrand, Marianne, and Jessica Pan. 2013. “The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior.” *American Economic Journal: Applied Economics* 5 (1): 32–64.
- Blau, Francine D., Lawrence M. Kahn, Peter Brummund, Jason Cook, and Miriam Larson-Koester. 2020. “Is there still son preference in the United States?” *Journal of Population Economics* 33 (3): 709–750.
- Boxberger, Karolina, and Anne Kerstin Reimers. 2019. “Parental Correlates of Outdoor Play in Boys and Girls Aged 0 to 12—A Systematic Review.” *International Journal of Environmental Research and Public Health* 16 (2): 190.
- Bruhn, Jesse, and Emily Emick. 2023. “Lottery evidence on the impact of preschool in the United States: A review and meta-analysis.” Discussion Paper 2023.20, Blueprints Labs.
- Carneiro, Pedro, and Rita Ginja. 2014. “Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start.” *American Economic Journal: Economic Policy* 6 (4): 135–173.
- Cascio, E.U., and D.W. Schanzenbach. 2013. “The Impacts of Expanding Access to High-quality Preschool Education.” *Brookings Papers on Economic Activity* (FALL 2013): 127–178.
- CBS. 2004. “Income Survey.”
- CBS. 2005a. “population aged 15 and over and population aged 25-54 (at main working ages), by civilian labour force characteristics, population group and sex.” Statistical Report 12.1.
- CBS. 2005b. “Population Estimates: Population, by Religion and Population Group.” Statistical Report 2.1. Jerusalem.

- Cook, Philip J., and Songman Kang. 2020. "Girls to the Front: How Redshirting and Test-score Gaps are Affected by a Change in the School-entry Cut Date." *Economics of Education Review* 76: 101968.
- Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy* 126 (6): 2356–2409.
- Crockenberg, Susan C. 2003. "Rescuing the Baby From the Bathwater: How Gender and Temperament (May) Influence How Child Care Affects Child Development." *Child Development* 74 (4): 1034–1038.
- Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *The American Economic Review* 85 (3): 341–364.
- Dahl, Gordon B., and Enrico Moretti. 2008. "The Demand for Sons." *Review of Economic Studies* 75 (4): 1085–1120.
- DeMalach, Elad, and Analia Schlosser. 2024. "Short- and Long-Term Effects of Universal Preschool: Evidence from the Arab Population in Israel." *SSRN Electronic Journal*.
- Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–134.
- Deming, David, and Susan Dynarski. 2008. "The Lengthening of Childhood." *Journal of Economic Perspectives* 22 (3): 71–92.
- DiPrete, Thomas A., and Jennifer L. Jennings. 2012. "Social and Behavioral Skills and the Gender Gap in Early Educational Achievement." *Social Science Research* 41 (1): 1–15.
- Feki, Shereen el, Brian Heilman, Gary Baker, United Nations Entity for Gender Equality and the Empowerment of Women, and Promundo. edited 2017. *Understanding masculinities:*

results from the International Men and Gender Equality Survey (IMAGES) - Middle East and North Africa ; Egypt, Lebanon, Morocco, and Palestine. New York/N.Y.

Felfe, Christina, Natalia Nollenberger, and Núria Rodríguez-Planas. 2015. “Can’t Buy Mommy’s Love? Universal Childcare and Children’s Long-term Cognitive Development.” *Journal of Population Economics* 28 (2): 393–422.

Fidjeland, Andreas, Mari Rege, Ingeborg F. Solli, and Ingunn Størksen. 2023. “Reducing the Gender Gap in Early Learning: Evidence From a Field Experiment in Norwegian Preschools.” *European Economic Review* 154: 104413.

Fort, Margherita, Andrea Ichino, and Giulio Zanella. 2020. “Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families.” *Journal of Political Economy* 128 (1): 158–205.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. “Longer-Term Effects of Head Start.” *The American Economic Review* 92 (4): 999–1012.

Gormley, William T., and Ted Gayer. 2005. “Promoting School Readiness in Oklahoma: An Evaluation of Tulsa’s Pre-K Program.” *The Journal of Human Resources* 40 (3): 533–558.

Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters. 2022. “The Long-Term Effects of Universal Preschool in Boston.” *The Quarterly Journal of Economics* 138 (1): 363–411.

Haddad Haj-Yahya, Nasreen, Muhammad Khlaily, Arik Rudnitzky, and Ben Fargeon. 2021. “The Arab Society in Israel Yearbook: 2021.” Annual Report, The Israel Democracy Institute. Jerusalem.

Havnes, Tarjei, and Magne Mogstad. 2011. “No Child Left Behind: Subsidized Child Care and Children’s Long-Run Outcomes.” *American Economic Journal: Economic Policy* 3 (2): 97–129.

- Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. “Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review* 103 (6): 2052–2086.
- Humphries, John Eric, Christopher Neilson, Xiaoyang Ye, and Seth Zimmerman. 2024. “Parents’ Earnings and the Returns to Universal Pre-Kindergarten.” w33038, National Bureau of Economic Research. Cambridge, MA.
- Imbens, Guido W., and Joshua D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62 (2): 467–475.
- Jackson, C Kirabo. 2018. “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes.” *Journal of Political Economy* 126 (5): 2072–2107.
- Jayachandran, Seema, and Rohini Pande. 2017. “Why Are Indian Children So Short? The Role of Birth Order and Son Preference.” *American Economic Review* 107 (9): 2600–2629.
- Johnson, Rucker C., and C. Kirabo Jackson. 2019. “Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending.” *American Economic Journal: Economic Policy* 11 (4): 310–349.
- Kimhi, Ayal. 2012. “Pre-primary Education in Israel: Organizational and Demographic Aspects.” policy Report, Taub Center for Social Policy Studies in Israel. Jerusalem.
- Kline, Patrick, and Christopher R. Walters. 2016. “Evaluating Public Programs with Close Substitutes: The Case of HeadStart.” *The Quarterly Journal of Economics* 131 (4): 1795–1848.
- Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff. 2006. “Economic, Neurobiological, and Behavioral Perspectives on Building America’s Future Workforce.” *Proceedings of the National Academy of Sciences* 103 (27): 10155–10162.

- Lavy, Victor, and Edith Sand. 2018. "On the Origins of Gender Gaps in Human Capital: Short- and Long-term Consequences of Teachers' Biases." *Journal of Public Economics* 167: 263–279.
- Lipsey, Mark W., Dale C. Farran, and Kelley Durkin. 2018. "Effects of the Tennessee Prekindergarten Program on Children's Achievement and Behavior Through Third Grade." *Early Childhood Research Quarterly* 45: 155–176.
- Ludwig, J., and D. L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *The Quarterly Journal of Economics* 122 (1): 159–208.
- Magnuson, Katherine A., Robert Kelchen, Greg J. Duncan, Holly S. Schindler, Hilary Shager, and Hirokazu Yoshikawa. 2016. "Do the Effects of Early Childhood Education Programs Differ by Gender? A Meta-analysis." *Early Childhood Research Quarterly* 36: 521–536.
- Miller, Douglas L., Na'ama Shenhav, and Michel Grosz. 2023. "Selection into Identification in Fixed Effects Models, with Application to Head Start." *Journal of Human Resources* 58 (5): 1523–1566.
- Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.
- Reardon, Sean F., Erin M. Fahle, Demetra Kalogrides, Anne Podolsky, and Rosalía C. Zárate. 2019. "Gender Achievement Gaps in U.S. School Districts." *American Educational Research Journal* 56 (6): 2474–2508.
- Reeves, Richard V. 2022. *Of Boys and Men: Why the Modern Male is Struggling, Why it Matters, and What to Do About it*. Washington, D.C: Brookings Institution Press.
- Tandon, Pooja S., Chuan Zhou, and Dimitri A. Christakis. 2012. "Frequency of Parent-

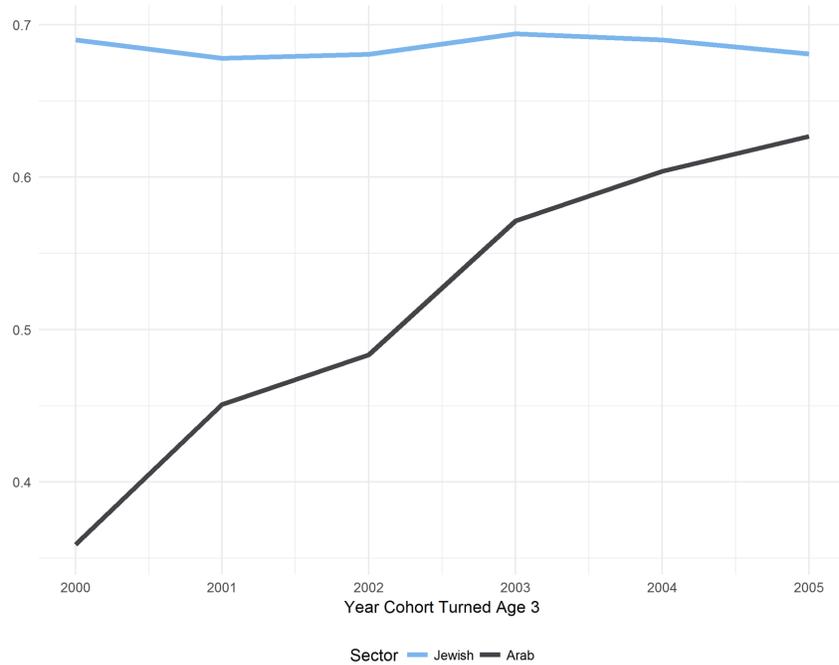
Supervised Outdoor Play of US Preschool-Aged Children.” *Archives of Pediatrics & Adolescent Medicine* 166 (8).

UNDP. 2006. “The Arab Human Development Report: Towards the Rise of Women in the Arab World.” UN Report, UNDP, New York.

Weiland, Christina, Rebecca Unterman, Anna Shapiro, Sara Staszak, Shana Rochester, and Eleanor Martin. 2020. “The Effects of Enrolling in Oversubscribed Prekindergarten Programs Through Third Grade.” *Child Development* 91 (5): 1401–1422.

Yassine-Hamdan, Nahla, and John Strate. 2020. “Gender Inequality in the Arab World: A Comparative Perspective.” *Contemporary Arab Affairs* 13 (3): 25–50.

Figures



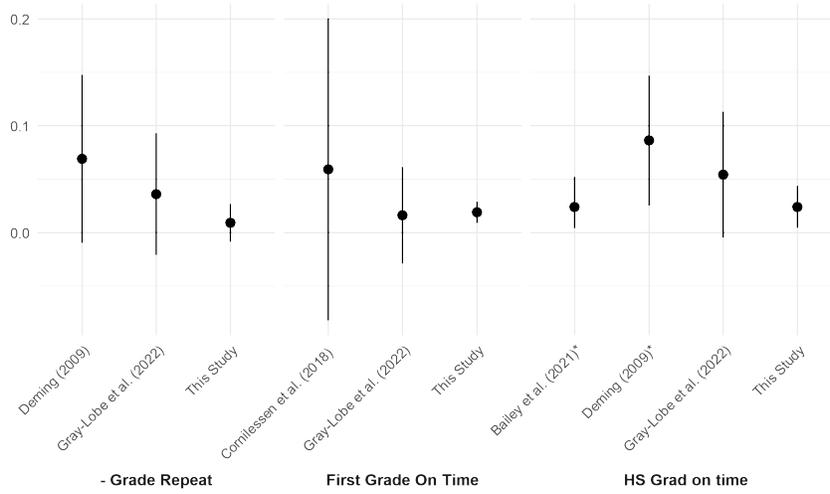
(a) Enrollment Rate in Pre-K at Age 3 by Year and Sector



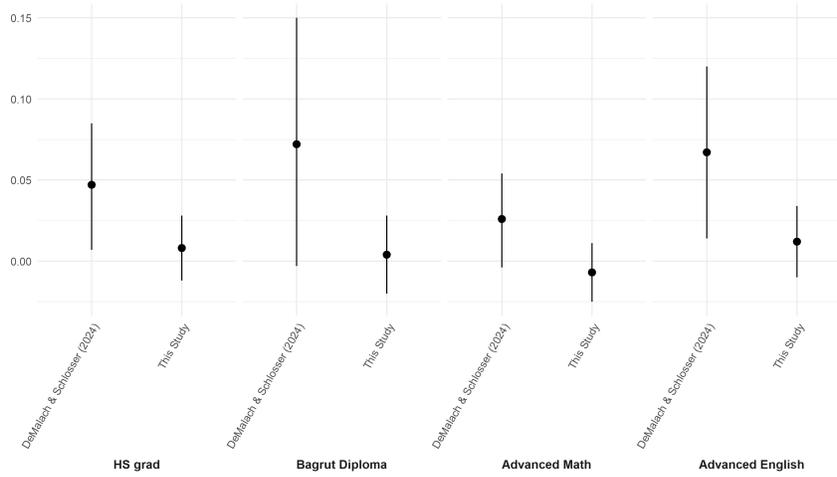
(b) Enrollment Rate in Pre-K at Age 3 in Sample Municipalities

Figure 1: Enrollment Rate in Pre-K at Age 3

Notes. The figure depicts enrollment rates in public pre-K for age three over time. Panel A shows rates for the universe of Arab (black line) and Jewish students. Panel B shows enrollment rate for the Arab localities included in the 2001 special order.



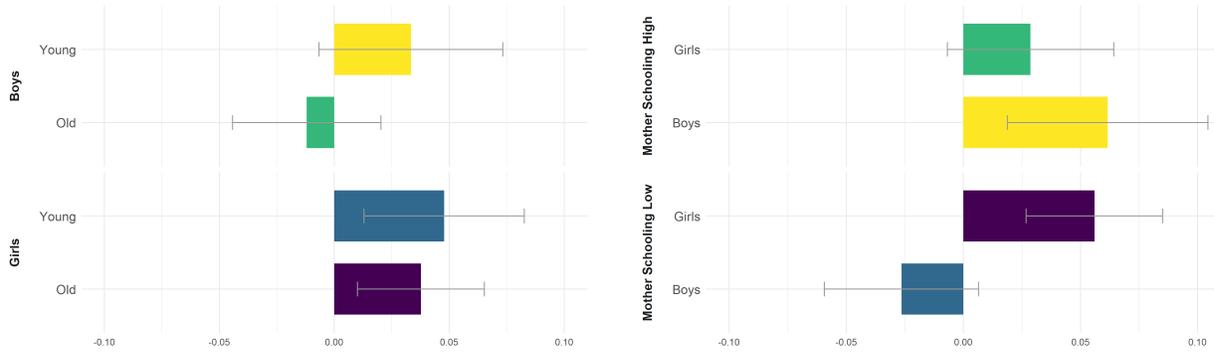
Panel A: Comparison to Other Pre-K Treatment Estimates



Panel B: Comparison to Estimates in DeMalach and Schlosser (2024)

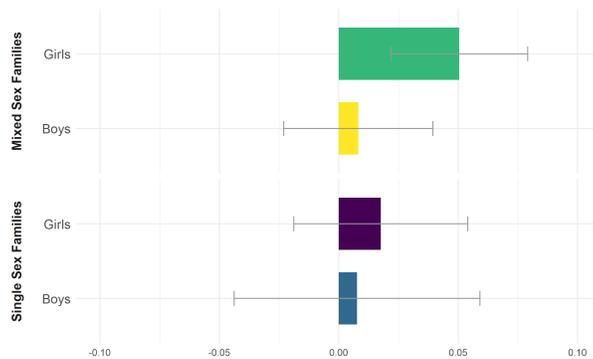
Figure 2: Effects of Large-Scale Preschool Programs Across Studies

Notes. Panel A plots point estimates and 95% confidence intervals for treatment effects on three outcomes from different studies. Effects on grade repetition are multiplied by negative one for ease of presentation. For on-time high school graduation, an asterisk indicates that an on-time estimate was not available and an ever-graduated-high-school outcome is used instead. Panel B shows treatment estimates and 95% confidence intervals from this study and DeMalach and Schlosser (2024). To calculate treatment point estimates, I scale the ITT estimate by the first-stage estimate reported in DeMalach and Schlosser (2024). To construct confidence intervals, I first calculate the implied 95% confidence interval for the ITT from the reported ITT standard error and then scale it by the first stage.



Panel A: Sex and Relative-Age Treatment Heterogeneity

Panel B: Sex and Maternal-Schooling Treatment Heterogeneity



Panel C: Sex and Sibling-Sex-Composition Treatment Heterogeneity

Figure 3: Estimates of Heterogeneous Effect on Graduating from High School on Time

Notes. The figure plots estimates of β in equation 1 for different subgroups. In each panel, estimates are obtained by interacting pre-K with a full set of dummy variables for each subgroup.

Tables

Table 1: Summary Statistics

Statistic	Entire Sample	Family FE Sample
Attended Pre-K	0.815 (0.388)	0.806 (0.395)
Girl	0.489 (0.500)	0.499 (0.500)
Father Schooling	10.296 (2.778)	10.241 (2.723)
Mother Schooling	10.092 (2.611)	10.008 (2.555)
Family Size	3.784 (1.954)	3.915 (1.973)
Month of Birth	6.621 (3.418)	6.626 (3.423)
Birth Order	3.144 (2.001)	3.161 (1.995)
First Grade on Time	0.980 (0.141)	0.979 (0.143)
Graduated High School on Time	0.826 (0.379)	0.820 (0.385)
Graduated High School	0.858 (0.349)	0.852 (0.355)
Attended 11th Grade	0.884 (0.321)	0.879 (0.326)
Ever Repeated a Grade	0.106 (0.308)	0.108 (0.326)
Diploma (Bagrut)	0.540 (0.498)	0.527 (0.499)
Diploma Credits	19.745 (12.840)	19.403 (12.872)
Advanced Math	0.223 (0.417)	0.213 (0.409)
Advanced English	0.483 (0.500)	0.466 (0.499)
Advanced Hebrew	0.316 (0.465)	0.312 (0.463)
Number of Students	54,264	40,669
Number of Families	30,466	16,747

Notes. The table shows summary statistics for key background and outcome variables. Column 1 shows statistics for students in birth cohorts 1996 to 2001 who resided in sample localities. Panel B shows results for children who have a sibling in the sample.

Table 2: Predictors of Public Pre-K Enrollment

	<i>Dependent variable: Pre-K Enrollment</i>		
	Univariate Models	Multivariate Model	
	(1)	(2)	(3)
Girl	-0.001 (0.003)	0.001 (0.003)	-0.002 (0.003)
Father Schooling > 10	0.014*** (0.003)	0.009*** (0.003)	0.025*** (0.005)
Mother Schooling > 10	0.118*** (0.003)	0.068*** (0.004)	0.024*** (0.008)
Family Size	-0.033*** (0.001)	-0.039*** (0.003)	-0.001 (0.004)
Month of Birth	-0.007*** (0.0005)	-0.007*** (0.0005)	-0.007*** (0.001)
Birth Order	-0.025*** (0.001)	0.024*** (0.002)	-0.005* (0.003)
Town-Cohort Fixed Effects	X	X	✓

Notes. The table presents OLS estimates from a regression of public pre-K enrollment on various characteristics. Column 1 shows coefficients from separate univariate regressions in which each characteristic is the sole independent variable. Columns 2 and 3 show results from two multivariate regression specifications: one without town-cohort fixed effects (column 2) and one with them (column 3). The sample includes all children residing in sample municipalities from birth cohorts 1996–2001. Robust standard errors are reported in parentheses below each coefficient. *p<0.1; **p<0.05; ***p<0.01.

Table 3: Characteristics of Switchers and Non-switchers

	Non-switchers		Switchers	
	Never Takers (1)	Always Takers (2)	Expansion Compliers (3)	Expansion Defiers (4)
Number of Families	1,430	11,982	2,608	727
Children in Sample per Family	2.58	2.37	2.63	2.47
Father Years of Schooling	9.87	10.44	9.92	10.08
Mother Years of Schooling	9.09	10.37	9.53	9.61
Young Child Born Sep-Dec (%)	44.8%	38.8%	35.3%	50.6%
Birth Order	3.99	2.83	3.69	3.50
Family Size	4.82	3.49	4.45	4.19
Oldest Child Attended				
Pre-K at Age 3	0	1	0	1
Pre-K at Age 4	0.40	0.97	0.73	0.85

Notes. The table reports means of selected family characteristics by four mutually exclusive groups of families based on pre-K enrollment status of siblings that are included in the sample: neither the oldest nor youngest sampled sibling attended (column 1), both the oldest and youngest sampled siblings attended (column 2), the youngest sampled sibling attended but the oldest did not (column 3), and the oldest sampled sibling attended but the youngest did not (column 4).

Table 4: Effects of Public Pre-K On School Progression and High School Achievements

Panel A: School-Progression Outcomes					
	First Grade on time (1)	Grade Repeat (2)	11th Grade (3)	Grad HS on Time (4)	Grad HS (5)
Estimate	0.019*** (0.005)	-0.013 (0.009)	0.009 (0.009)	0.024** (0.010)	0.008 (0.010)
Mean	0.979	0.108	0.879	0.820	0.852
Observations	40,669	40,669	40,669	40,669	40,669
Adjusted R^2	0.091	0.240	0.293	0.306	0.309
Panel B: HS Diploma Outcomes					
	HS Diploma (Bagrut) (1)	Diploma Credits (2)	Advanced Math (3)	Advanced English (4)	Advanced Hebrew (5)
Estimate	0.004 (0.012)	0.487* (0.279)	-0.007 (0.009)	0.012 (0.011)	0.011 (0.011)
Mean	0.527	19.403	0.213	0.466	0.312
Observations	40,669	40,669	40,669	40,669	40,669
Adjusted R^2	0.396	0.548	0.436	0.476	0.438

Notes. The table shows estimates of β in equation 1. Panel A shows estimates of the effects on school-progression outcomes. Panel B shows estimates of the effects on educational outcomes measured at the end of high school. All models control for gender, month of birth, birth order, and town-specific cohort fixed effects. Standard errors, clustered at the family level, are displayed in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: Robustness Tests

Robustness Test	HS Grad On Time	HS Diploma (Bagrut)
No Expansion Defiers	0.020 (0.012)	-0.005 (0.014)
Number of Observations		38,846
Only Cohort Fixed Effect	0.025*** (0.010)	-0.005 (0.011)
Number of Observations		40,669
Including 1999 municipalities	0.017** (0.008)	0.0002 (0.008)
Number of Observations		61,379
Only Youngest and Oldest Child	0.025* (0.014)	0.003 (0.016)
Number of Observations		33,026

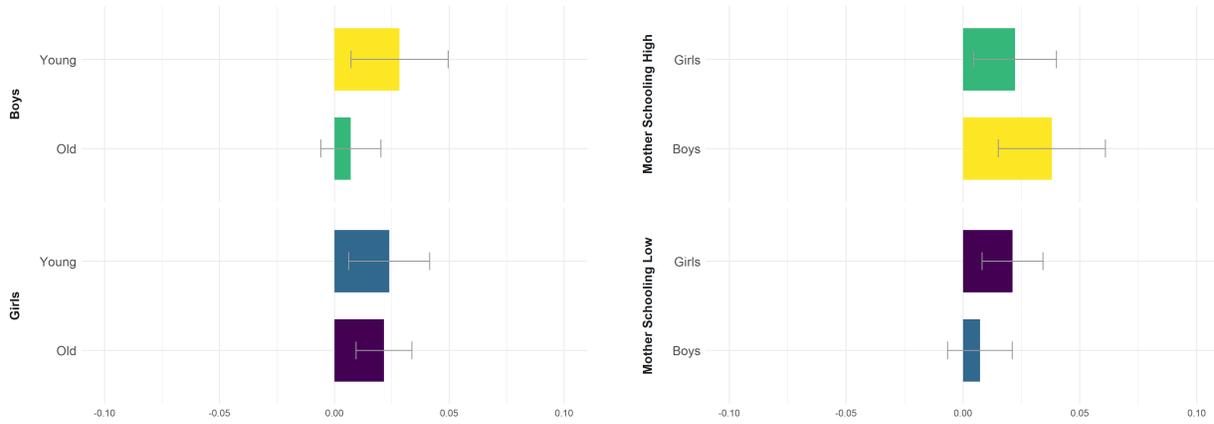
Notes. The table shows estimates of β for high school graduation on time (column 1) and obtaining a high school diploma (Bagrut) (column 2) across four robustness tests: excluding expansion defiers from the sample; replacing town-specific cohort fixed effects with common cohort fixed effects; including municipalities from the 1999 special order; and keeping in the sample only the oldest and youngest children in the sample. Standard errors, clustered at the family level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 6: Effects of Public Pre-K on School Progression for Subgroups

	By Sex		By Relative Age		By Mother Education		By Family Size	
	Boys (1)	Girls (2)	Young (3)	Old (4)	High (5)	Low (6)	Small (7)	Large (8)
First Grade on Time	0.016*** (0.006)	0.022*** (0.006)	0.026*** (0.007)	0.014*** (0.005)	0.039*** (0.010)	0.014** (0.007)	0.016** (0.007)	0.021*** (0.006)
<i>p</i> -value for equal effects	0.370		0.122		0.029		0.522	
Grade Repeat	-0.007 (0.011)	-0.018* (0.010)	-0.018 (0.012)	-0.009 (0.010)	-0.021 (0.016)	-0.017 (0.015)	-0.008 (0.012)	-0.016 (0.011)
<i>p</i> -value for equal effects	0.370		0.493		0.859		0.583	
11 th Grade	-0.017 (0.012)	0.036*** (0.010)	0.016 (0.012)	0.005 (0.011)	0.016 (0.015)	0.015 (0.014)	-.011 (0.011)	0.023* (0.012)
<i>p</i> -value for equal effects	0.0002		0.393		0.985		0.002	
Grad HS on Time	0.006 (0.014)	0.042*** (0.012)	0.040*** (0.014)	0.013 (0.012)	0.050*** (0.019)	0.016 (0.017)	0.020 (0.014)	0.027** (0.013)
<i>p</i> -value for equal effects	0.021		0.089		0.157		0.686	
Grad HS	-0.014 (0.013)	0.030*** (0.011)	0.011 (0.013)	0.005 (0.011)	0.024 (0.016)	0.001 (0.015)	0.003 (0.013)	0.011 (0.012)
<i>p</i> -value for equal effects	0.003		0.711		0.262		0.609	
Joint <i>p</i> -value across outcomes	0.000013		0.25		0.42		0.27	

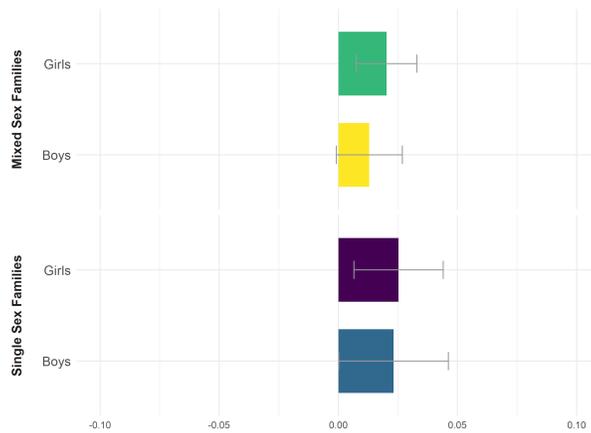
Notes. This table reports estimates of β in equation 1 on school-progression outcomes for subgroups. Columns 1 and 2 compare estimates for boys and girls; columns 3 and 4 display estimates by relative age; columns 5 and 6 show estimates by maternal education (more or less than 10 years); and columns 7 and 8 show estimates by family size (larger than three versus smaller than or equal to three). *p*-values for equal effects come from tests of the null hypothesis that effects are equal across subgroups. All models control for gender, month of birth, birth order, and town-specific cohort fixed effects. The last row shows joint *p*-values from tests of the null hypothesis that subgroups have equal effects across all outcomes in the table. **p*<0.1; ***p*<0.05; ****p*<0.01.

Appendix



Panel A: Sex and Relative Age Treatment Heterogeneity by

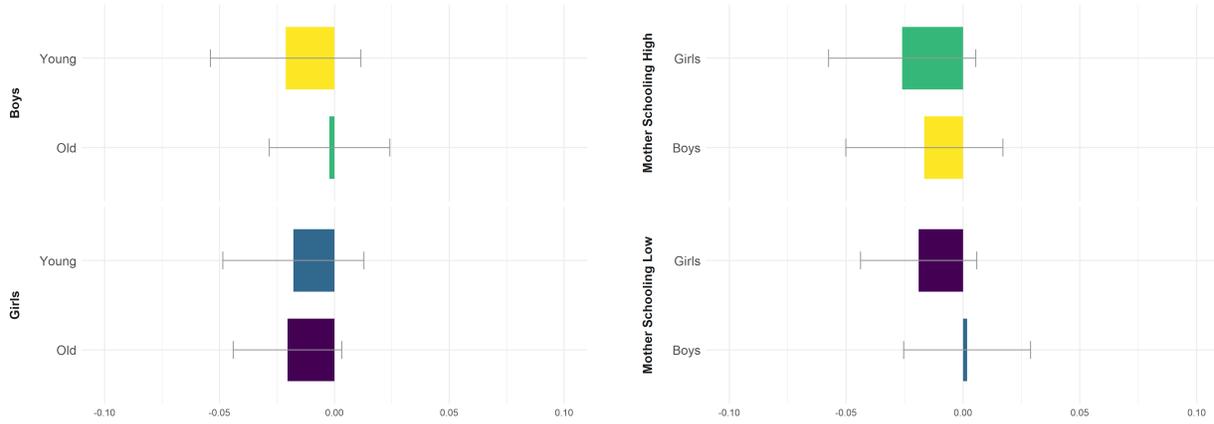
Panel B: Sex and Maternal Schooling Treatment Heterogeneity



Panel C: Sex and Sibling Sex Composition Treatment Heterogeneity

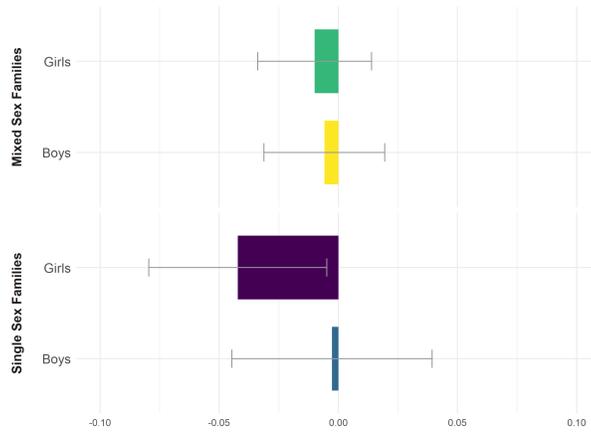
Figure A1: Estimates of Heterogeneous Effect on First Grade on Time

Notes. The figure plots estimates of β in equation 1 for different subgroups. In each panel estimates are obtained by interacting pre-K with a full set of dummies variables for each subgroup.



Panel A: Sex and Relative Age Treatment Heterogeneity by

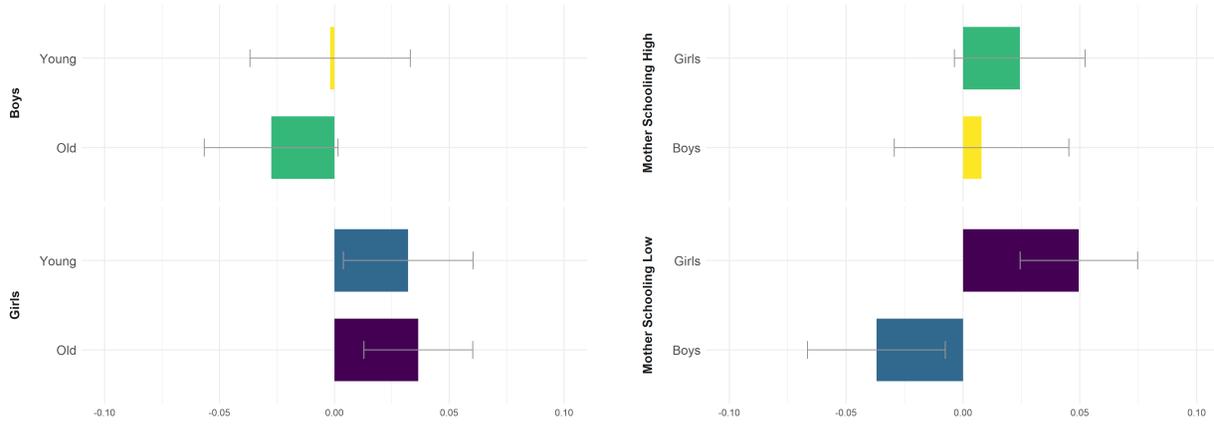
Panel B: Sex and Maternal Schooling Treatment Heterogeneity



Panel C: Sex and Sibling Sex Composition Treatment Heterogeneity

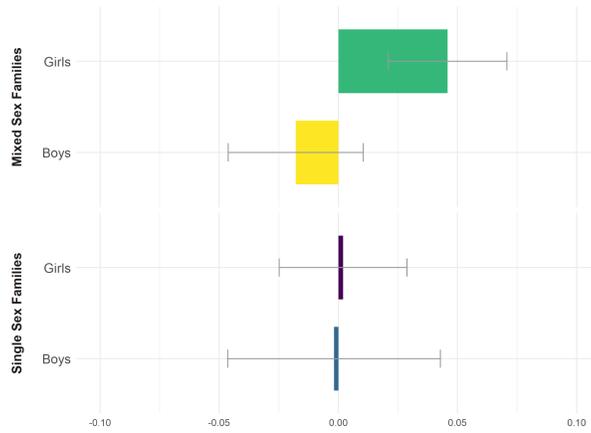
Figure A2: Estimates of Heterogeneous Effect on Ever Repeating a Grade

Notes. The figure plots estimates of β in equation 1 for different subgroups. In each panel estimates are obtained by interacting pre-K with a full set of dummies variables for each subgroup.



Panel A: Sex and Relative Age Treatment Heterogeneity by

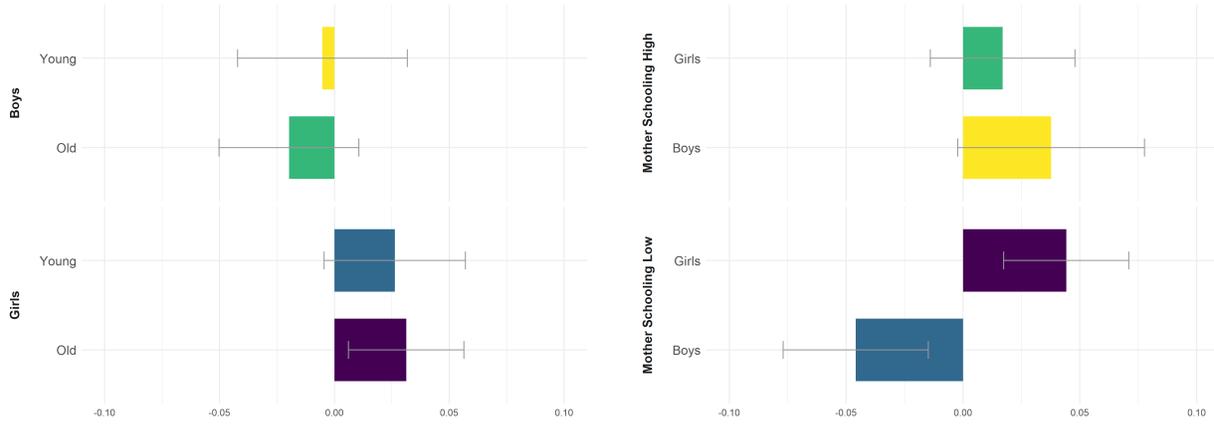
Panel B: Sex and Maternal Schooling Treatment Heterogeneity



Panel C: Sex and Sibling Sex Composition Treatment Heterogeneity

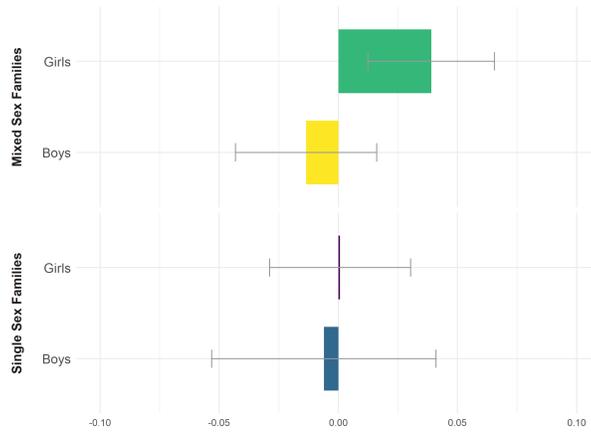
Figure A3: Estimates of Heterogeneous Effect on Attending 11th Grade

Notes.— The figure plots estimates of β in equation 1 for different subgroups. In each panel estimates are obtained by interacting pre-K with a full set of dummies variables for each subgroup.



Panel A: Sex and Relative Age Treatment Heterogeneity by

Panel B: Sex and Maternal Schooling Treatment Heterogeneity



Panel C: Sex and Sibling Sex Composition Treatment Heterogeneity

Figure A4: Estimates of Heterogeneous Effect on Ever Graduating from High School

Notes.— The figure plots estimates of β in equation 1 for different subgroups. In each panel estimates are obtained by interacting pre-K with a full set of dummies variables for each subgroup.

Table A1: Estimates of the Impact of Public pre-K attendance on Obtaining a High School Diploma (*Bagrut*)

	<i>Dependent Variable: High School Diploma (Bagrut)</i>				
	(1)	(2)	(3)	(4)	(5)
PreK-3	0.124*** (0.006)	0.089*** (0.007)	0.056*** (0.006)	0.050*** (0.007)	0.004 (0.012)
Background Controls	✗	✓	✓	✓	✓
Town-Cohort FE	✗	✗	✓	✓	✓
Siblings Sample	✗	✗	✗	✓	✓
Sibling FE	✗	✗	✗	✗	✓
Observations	54,264	54,264	54,264	40,669	40,669
Adjusted R ²	0.009	0.058	0.195	0.194	0.396

Notes. This table reports coefficients from regressing high school diploma on public pre-K attendance at age three. Column 1 presents the estimated coefficient from a univariate model. Column 2 adds controls for sex, month of birth, birth order, and parental education. Columns 3 and 4 also incorporate town-cohort fixed effects. Column 5 further adds family fixed effects. The sample in columns 1-3 includes all children residing in sample municipalities from birth cohorts 1996-2001. Columns 4-5 restrict the sample to children with at least one sibling in the sample. Robust standard errors are reported in parentheses below each coefficient. *p<0.1; **p<0.05; ***p<0.01.

Table A2: Effects of Public Pre-K on High School Achievements for Subgroups

	By Sex		By Relative Age		By Mother Education		By Family Size	
	Boys (1)	Girls (2)	Young (3)	Old (4)	High (5)	Low (6)	Small (7)	Large (8)
HS Diploma (<i>Bagrut</i>)	0.003 (0.015)	0.002 (0.015)	0.009 (0.016)	-0.001 (0.014)	-0.013 (0.024)	-0.009 (0.018)	0.002 (0.017)	0.003 (0.014)
Diploma Credits	0.277 (0.356)	0.608* (0.341)	0.499 (0.380)	0.405 (0.322)	0.764 (0.551)	0.307 (0.424)	0.432 (0.394)	0.446 (0.342)
Advanced Math	-0.013 (0.010)	-0.000 (0.012)	-0.005 (0.012)	-0.008 (0.011)	-0.013 (0.022)	-0.005 (0.013)	-0.003 (0.014)	-0.010 (0.010)
Advanced English	0.010 (0.013)	0.014 (0.014)	0.016 (0.015)	0.010 (0.013)	0.020 (0.023)	0.013 (0.017)	0.014 (0.016)	0.011 (0.013)
Advanced Hebrew	0.019 (0.013)	0.001 (0.014)	0.023 (0.015)	0.001 (0.013)	-0.020 (0.024)	0.029 (0.017)	0.010 (0.016)	0.010 (0.013)

Notes. This table reports estimates of β in equation 1 on high school educational achievements for subgroups. Columns (1) and (2) compare estimates for boys and girls; columns (3) and (4) display estimates by relative age; columns (5) and (6) show estimates by maternal education (more or less than 10 years); and columns (7) and (8) show estimates by family size (larger than three vs. smaller or equal to three). All models control for gender, month of birth, birth order, and town-specific cohort fixed effects. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.